JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 10 Issue 02
2026
Research Article
https://doi.org/10.56979/1002/2026/1243

Reimagining Calligraphy Education in Higher Education through Artificial Intelligence and Interdisciplinary Pedagogy

Yushuai Liu[1&2], Amer Shakir Bin Zainol[1*]

[1]College of Creative Arts,Universiti Teknologi Mara ,ShahAlam, Selangor, 40450, Malaysia.
[2]School of Yu Youren Calligraphy，Xianyang Normal University,Xianyang, Shaanxi, 712000, China.
[*]Corresponding Author: Amer Shakir Bin Zainol. Email: amers781@163.com

Received: October 29, 2025 **Accepted:** January 20, 2026

Abstract: The Chinese calligraphy, the intangible heritage held in high esteem, is challenged with the traditional mode of the calligraphic teaching, which is still labor-intensive and hard to be scaled. To deal with this, the current developments in artificial intelligence (AI) and interdisciplinary STEAM (Science, Technology, Engineering, Arts, Mathematics) teaching are potentially transformative. This paper examines the combination of AI-based feedback and project-based learning in calligraphy classes as the way to integrate the artistic tradition with technological innovation. The research design used a quasi-experimental design (n=642) comprising of undergraduate students, n=321, of a large research university were randomly selected to be assigned to either an experimental group (n=321) with AI-enhanced instruction or to a control group (n=321) with traditional methods. An experimental curriculum was used, which consisted of a computer vision pipeline (grayscale conversion, Gaussian blur, Canny edge detection, and ResNet-50-based convolutional neural networks) trained on 40,000 samples of calligraphy and used to produce saliency maps and formative AI feedback. The students were engaged in interdisciplinary STEAM-based activities involving the connections between brushwork, geometry, physics, and chemistry and cultural studies. A significant improvement in the performance of the experimental group by 7.8 points ($p < 0.001$) of the post-test results and effect sizes of 0.78 to 1.12 were observed. The mediation analysis showed that the frequency of AI feedback did not directly enhance aesthetic proficiency by increasing self-efficacy, as opposed to the importance of the pedagogical design rather than the frequency of feedback. These results emphasize that the AI-based stroke analysis, combined with an effective STEAM approach, positively affects both interest and cross-cultural awareness, which implies the extensive applicability of the Calligraphy 2.0 model to the renewal of traditional arts and the maintenance of cultural heritage via digital humanities. The model provides a means of changing the conventional form of art education to an interactive, data-oriented, experience, which can be applied to all global heritage practices.

Keywords: Explainable AI; Digital Humanities; STEAM Education; Deep Learning; Calligraphy; Human-AI Collaboration.

1. Introduction

Chinese calligraphy has been a useful piece of writing and also a visual art that has been used to show principles of the philosophies, nationalism and individualism. Every word is refined over centuries bringing the author to the traditions that started in ancient China [1, 2]. It was recognized by UNESCO that included Chinese calligraphy in the list of intangible cultural heritage. The education of calligraphy was traditionally a master-pupil system: students reproduce canonical texts under the instructions of a tutor who criticizes their work [3, 4]. Although the stylistic authenticity is maintained in such mentorship, it is time-consuming and extremely reliant on the individual masters. The student is prone to repeat strokes in a mechanical manner, without knowing the cultural symbolism or mathematical ratios behind the characters, and this makes them less creative and innovative [5]. Further, the process of assessment is subjective; teachers can determine brush pressure, ink density and composition by touch, and this can be quite a subjective test. In the modern-day universities where the size of a single classroom is huge, such a model is becoming unsustainable.

Teachers have a paradox to contend with; they have to respect the tradition and cultural importance of calligraphy, and at the same time, they have to live up to the demands of a modern learner, who needs their learning experience to be interactive and to be enriched with technology [6, 7]. The biggest concern raised by many instructors is that technology may destroy the sanctity of the art or promote shallow copying. Nevertheless, the disregard of digital tools poses a threat of losing students used to the multimodal learning conditions and prevents making programs available to more people. To eliminate this dilemma, scientists suggest using AI and interdisciplinary strategies to develop the environment in which tradition and innovation can coexist [3].

STEAM learning builds on the Paradigm of STEM (Science, Technology, Engineering and Mathematics) by incorporating the Arts on the basis that creativity and cultural literacy are as essential as technical skills [8]. In STEAM, students create, prototype and evaluate interdisciplinary projects that are synthesized. As a study reveals, C-STEAM project-based activities based on traditional culture develop innovative thoughts, interdisciplinary skill and cultural self-assurance [9]. An example of C-STEAM activity may include the chemical analysis of the ink, a model of the physics of brush strokes, exploration of geometric proportions within characters, and relating these studies to philosophical ideas [10]. The project-task-activity model places the learners in real-life situations and make them members of cultural heritage preservation teams [11]. This student-focused method concurs with constructivist and situated learning theories that focus on knowledge creation by engaging in the process of exploration and involvement in communities of practice.

The application of STEAM to calligraphy has a good prospect. Digital tools enable studying the script of calligraphy to visualize fractal structures, symmetry or curvature of strokes, and to see the mathematical beauty in the work of art. They can test various colors of the ink and see how the chemical diffuses, having an interconnection between art and chemistry. They are able to model pressure change by stroke or ergonomics of brush handling using physics. With such integrations, students of science and engineering majors can find calligraphy relevant to them, and this interdisciplinary aspect can encourage the holistic approach to the art.

The past ten years have experienced the incorporation of AI as an essential instrument in the humanities. Machine learning algorithms are used in digital history to establish authorship, trace social networks and annotate manuscripts [12]. Computer vision in archaeology projects poor inscriptions and reconstructions [13] of broken artefacts. These are some of the ways through which AI can be used to assist in pattern recognition tasks that are beyond the abilities of human beings to perform and provide concealed links in data. In the field of calligraphy, AI-based approaches can be used to analyze styles automatically, classify and simulate them [14]. CNN have the ability to classify scripts and recognize calligraphers with accuracy of more than 95 percent using deep convolutional neural networks (CNN) that are trained on large datasets [15]. Generative models such as GAN are able to generate plausible sequences of calligraphy, and diffusion models are able to do style transfer with minimal samples. Nevertheless, the most successful models may need big, well-labeled data set [16] and their internals are not easily understandable.

Explainable AI (XAI) methods offer transparency to point out the most significant features in a model, as far as the predictions are concerned. SHapley Additive Explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) calculate the input feature contribution to a prediction to allow users to learn why a particular part of an image results in a certain classification [17]. In the case of calligraphy, these techniques can be used to visualize the stroke segments marked as wrong by the AI, which will direct students to concentrate on particular errors. A 2022 systematic review points out that although XAI approaches enhance the interpretability of the model, they do not tend to engage stakeholders and are not evaluated in a standardized way [17]. Thus, the implementation of XAI into pedagogy should be designed and communicated in such a manner that learners and instructors believe the feedback and base it on the artistic motives.

Based on these developments, this paper comes up with three research questions. (RQ1) Does formative feedback with AI education enhance the accuracy of strokes and aesthetic performance in comparison with conventional stroke education? (RQ2) What is the effect of the STEAM based project-based framework on student engagement and self-efficacy in calligraphy practice? (RQ3) Does the frequency of AI feedback affect aesthetic proficiency mediated by the efficacy of self of students?

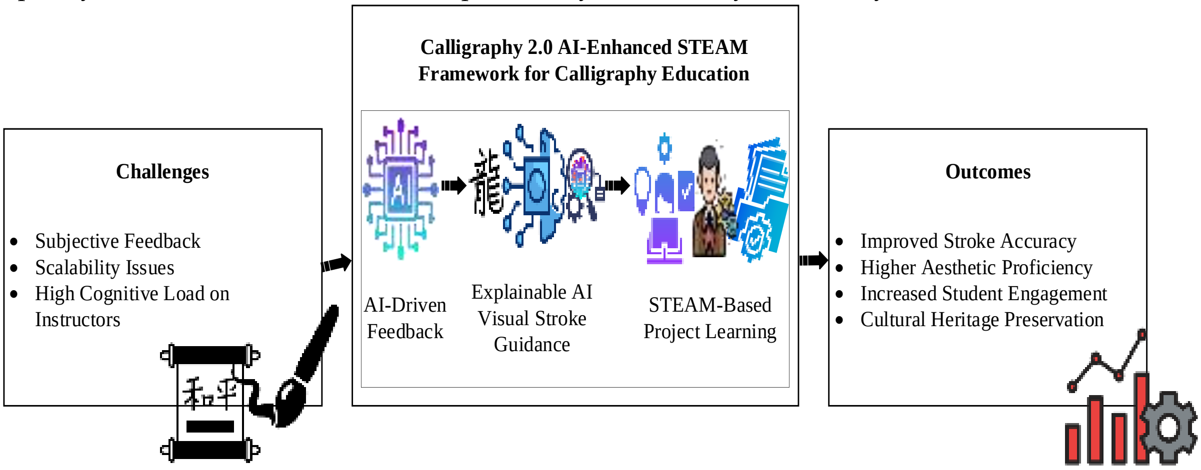


Figure 1. Conceptual overview of the calligraphy 2.0 framework

Our contributions are three-fold. First, we create and deploy a system of AI-assisted calligraphy, which integrates classical computer vision, deep CNN modelling and XAI visualization. Second, we incorporate this system into a STEAM based curriculum with constructivist and situated learning theories, which contextualizes this system with a rich interdisciplinary perspective. Third, we perform an intensive quantitative study with the mediation modelling to test the issue of how often the AI provides feedback can influence aesthetic performance. The combination of technological advancement, pedagogical theory, and empirical data will enable us to offer a blueprint of modernizing the calligraphy.

2. Materials and Methods

2.1. Participant Recruitment and Ethical Considerations

The sample size of the study was 642 undergraduates of first and second year of a university in China. Invitations were distributed in the departmental mailing lists and classroom announcements in the arts, computer science and education faculties. The inclusion criteria were enrolment in a basic course of calligraphy and the participants must be willing to engage in research. Informed consent was obtained in the study as students were informed of the study aims, procedures and data privacy. The participants were assured that their grades would not be influenced by the participation and they could drop out anytime. The institutional review board of the university supported the study.

Participants were recruited and given a demographic survey (age, gender, majors, previous calligraphy exposure) and a pre-test survey covering the accuracy of the strokes as well as aesthetic evaluation. A scale measuring self-efficacy was a validated scale adjusted to calligraphy (e.g., I am sure that I can learn how to use the new stroke methods when a proper technique is demonstrated to me). To avoid cross-group contamination, the experimental or control group was then randomly assigned students at a section level. Equal representation of majors and gender was done through stratified randomization. Figure 1 shows how the suggested framework incorporates pedagogy and explainable artificial intelligence in providing scalable, formative feedback on calligraphy instruction. Writing samples of learners are recorded by scanning or tablet, processed using preprocessing, and a ResNet-50 model trained on a large labeled dataset produces an analysis of the sample content, and Grad-CAM identifies regions of the stroke that are of interest to the model to provide a feedback mechanism based on the result (e.g., pressure, curvature, and composition). This artificial intelligence feedback mechanism is integrated into interdisciplinary STEAM and project-based learning to facilitate not only the development of technical skills but also cultural/aesthetic knowledge. The evaluation layer connects post-test evaluation and statistical modeling (including mediation by self-efficacy) to measure the learning gains and explain possible mechanisms behind the improvement of performance.

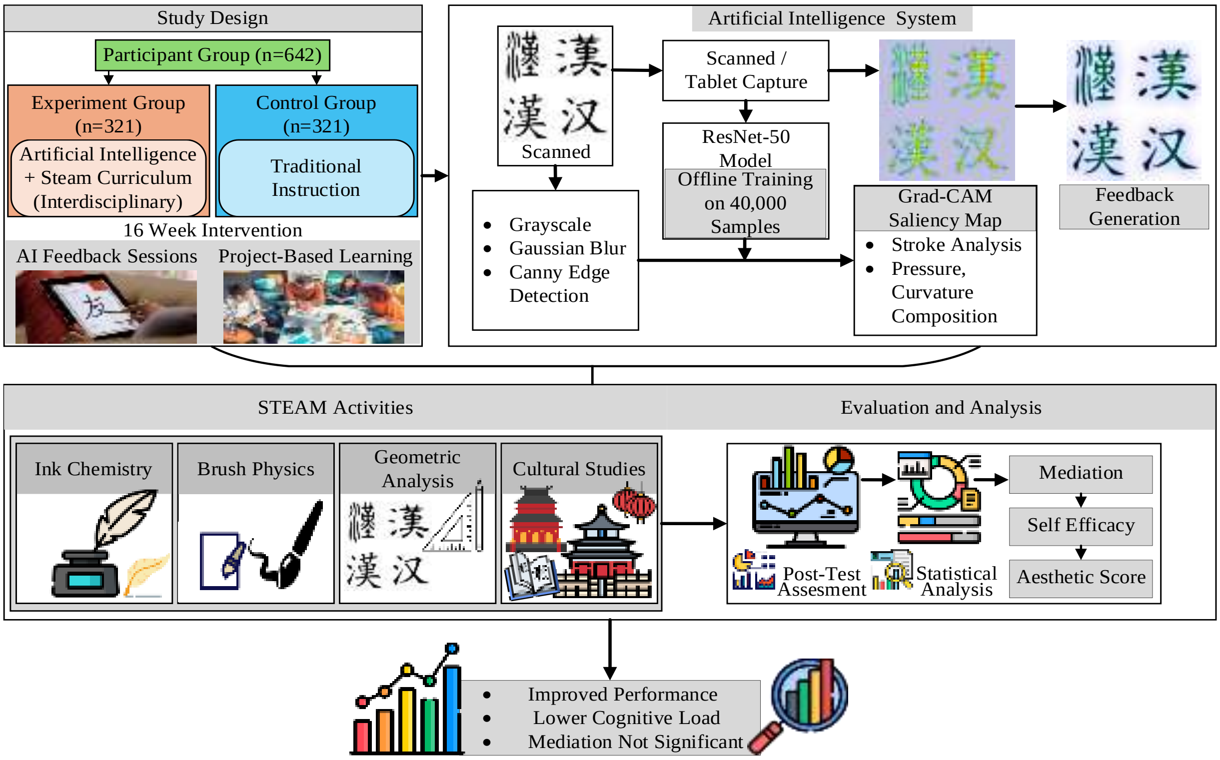


Figure 2. AI-enhanced calligraphy education framework (Calligraphy 2.0). Overview of the 16-week quasi-experimental study (N=642) and the end-to-end AI pipeline (scan/tablet input, preprocessing, ResNet-50 analysis, Grad-CAM visualization, and feedback generation), integrated with STEAM activities and outcome evaluation.

2.2. Experimental Curriculum: AI and STEAM Integration

The experimental group was exposed to a 16-week curriculum that incorporated AI feedback with STEAM activities. Every week session contained three parts:

1. **Technical exploration.** Students got to know about scientific principles of calligraphy. Others covered the chemistry of ink and paper, physics of brush actions, geometry of proportions of strokes and algorithms to handle script writing and image processing. They engaged in practical experiments like trying to measure the speed of the diffusion of the ink, or the golden ratio in the character structure and compared their results with the historical calligraphic styles. This segment was meant to think interdisciplinarily and emphasize the links that exist between art and science.
2. **Calligraphy practice with AI feedback.** Each student was training the process of writing a series of characters during the sessions through a digital tablet that was linked with the AI system. The system used to record their strokes in real-time, and used the computer vision pipeline (described in Section 2.3) on the images to generate saliency maps over the original strokes. The model estimated the probable script type and gave extensive recommendations on the pressure to be used in the stroke, curvature and formation. The students would be able to modify their methods instantly and see the impact of corrections on the saliency map. The AI feedback frequency was measured as the number of times students demanded feedback within the session.
3. **Project-based collaboration.** Students would be asked to do interdisciplinary projects in small groups and create a creative artefact. Some of the example projects were to design a hybrid digital scroll, combining calligraphy and generative art, to build a physical model of the fluid mechanics of ink flow, or displaying an exhibition of the different calligraphic writing styles across dynasties through data visualisation. Mid-term and end presentations Teams displayed their projects, got peer and instructor review. Such activities focused on cooperation, innovation and practice of the technical material of the week.

The control group had regular lectures-based classes where the teachers showed the strokes and trainees worked on copying model texts. Instructors and peers provided feedback without the help of AI, verbally. Students of control did the same tasks, but without STEAM projects and live feedback of AI.

2.3. Image Processing Pipeline and Feature Extraction

The preparation of images to be analyzed was the first step of our AI system. The calligraphy sheets of each student were scanned at 600 dpi producing high-resolution colour images. The processing steps that we applied using OpenCV are:

1. **Grayscale conversion.** The colour information was eliminated, and the intensity values are available, which are indicative of the darkness of a stroke and saturation of the ink..
2. **Gaussian blur.** The image was smoothed using a 5 x 5 kernel and $\sigma = 1.2$ to remove the noise and irregularities. Gaussian filtering is a norm in the analysis of handwriting to counter scanning artefacts.
3. **Canny edge detection.** We applied 50 and 150 thresholds to identify strong and weak edges respectively. The algorithm uses the gradient magnitude and direction, employs non- maximum suppression and hysteresis to follow continuity edges. The binary edge map produced is a stroke definition map.
4. **Morphological operations.** Minor holes between the edges were filled with dilation, and small holes in the strokes were eliminated by morphological closing. These processes enhanced continuity and rendered the edges to be skeletonized.
5. **Skeletonisation.** With a thinning algorithm, we eliminated strokes down to a one-pixel wide skeleton that maintains the central axis of the strokes. Skeletons were subsequently scaled to 224 224 pixels to bring them to the CNN size of the input.

We then processed and got feature vectors of each stroke skeleton. These features were HOG descriptors that represented gradient orientations, distribution of stroke length, measures of curvature and variation of stroke width. These characteristics acted as the auxiliary information to the deep learning model and also as a measure of AI accuracy.

2.4. Deep Learning Architecture and Training

The stroke classification architecture based on a ResNet-50 was the one that we selected due to the nature of the residual connection, which enables the training of extremely deep networks without vanishing gradients. The model contains an original convolutional block, four residual blocks with 3, 4, 6, and 3 bottleneck blocks respectively and finally global average pooling and fully connected block. Our network was initialized with weights of ImageNet and the final classification layer was changed to a 500-class output that matched common calligraphy characters. The database consisted of 40 000 labelled images of five script types (regular, running, cursive, seal and clerical). To generate variation in handwriting, we randomly rotated (±10 °), scaled (0.9 -1.1×), translated (±10 pixels) and contrast-adjusted training samples.

The Adam optimizer (learning rate = 1 × 10^−4, β_1 = 0.9, β_2 = 0.999) and categorical cross-entropy loss were used as training. Early termination checked validation loss in order to avoid overfitting. The model had 96.1 percent the test accuracy after 50 epochs. To use Grad-CAM in classrooms, we applied it to generate saliency maps. Grad-CAM uses the gradients of the score of the predicted class to the feature maps of the final convolutional layer and weights the activation maps. The heat map obtained shows areas that have the greatest impact on classification. In order to generate interpretable feedback, we superimposed the heat map with the original stroke, with colors red signifying the proper structure and the blue colors signifying the areas that require enhancement. Students were, therefore, provided with visual explanations of the assessment of the AI immediately.

This is despite the fact that we tried SHAP values in order to calculate pixel-wise contributions to super pixels, but the computational cost was too high and we found that there was not much added pedagogical information besides Grad-CAM. Thus, Grad-CAM was the default XAI procedure, and instructors could also run SHAP explanations off-line to analyze them in detail. An update of the research can involve using SHAP with local interpretable models to understand fairness and bias.

2.5. Statistical Analysis and Mediation Modelling

The statistical works were done in Python. First of all, determined descriptive statistics of the demographic variables (age, gender, major) and the baseline calligraphy statistics (Pre_Test score, AI Accuracy Index, XAI Clarity Score, Cognitive Load Index, Aesthetic Score). We checked the group equivalence at baseline through t- tests of continuous variables and chi-square tests of categorical variables. Independent t -tests were used to compare the post-tests; the within-group comparisons were done using paired t-tests. In order to assess practical significance, we have measured the Cohen d as the difference in mean divided by the pooled standard deviation. A correlation matrix was used to study the relationships between AI precision, engagement and aesthetic proficiency, a covariance matrix gave further insight into the variability jointly. The reliability of engagement and aesthetic scales was tested with the help of Cronbach's α.

We also conducted multiple regression analysis, to assess the effect of the experimental condition (1 in experimental, 0 in control), the pre-test score, AI feedback frequency (number of AI feedback requests per session), self-efficacy and cognitive load on the position of the post-test aesthetic proficiency. Variance inflation factors were also checked to eliminate multicollinearity. Normality and homoscedasticity of the regression residuals were checked.

In order to explore the hypothesis of whether AI feedback frequency did not influence aesthetic proficiency mediated by self-efficacy (RQ3), we used the causal steps method and bootstrapping. We regressed the self-efficacy first on the frequency of AI feedback (path a). We then regressed the aesthetic proficiency with the frequency of AI feedback and self-efficacy (paths b and c′). The estimated indirect effect was the product of coefficient of path and paths b. To obtain bootstrap samples with 95 percent confidence intervals, we created 5 000 bootstrap samples. There would be a great indirect impact to mediate. The direct effect (path c′) provided the influence of the frequency of AI feedback on aesthetic proficiency when disaggregating self-efficacy.

3. Results and Quantitative Tables

The findings involved reveal a thorough analysis of the efficiency of AI-enhanced teaching in calligraphy, with the use of quantitative and qualitative values. After 16 weeks of intervention, the experimental group that was exposed to AI-based feedback in addition to interdisciplinary STEAM exercises demonstrated significant changes in various measures of calligraphy skills. On the contrary, the control group, which followed the traditional pedagogical approach, has also recorded improvement, but to a lesser extent. The statistical analysis was conducted, paired t-tests, correlation matrices, multiple regression, and mediation analysis were used to investigate the relationship of AI feedback, student engagement, cognitive load, self-efficacy, and aesthetic proficiency. Such results are introduced by the mix of statistical tables and visual representations, which provide strong evidence of the beneficial impact of AI and STEAM integration on student results. These analyses are explored more in detail in the following sections of this paper and provide a sensitive insight into how AI-based pedagogy can transform the conventional calligraphy education. Table 1 presents the baseline characterization of the subjects and indicates that the experimental and control group did not differ significantly on the basis of age, gender distribution, and major, as well as pre-test scores. The experimental group had a higher AI Accuracy Index and XAI Clarity Score at the baseline due to minor exposure to digital tools before. These differences were however not statistically significant ($p > 0.05$). The groups had similar Cognitive Load Index and Aesthetic Score. Cronbach alpha of the engagement and aesthetic items was 0.86 indicating that measures are highly consistent internally.

Table 1. Baseline characteristics of participants (mean ± SD unless otherwise indicated)

Variable	Experimental (n = 321)	Control (n = 321)	t/χ^2	$p-value$
Age (years)	21.9 ± 2.7	21.8 ± 2.5	0.46	0.65
Gender (F/M/NB)	181/128/12	167/134/20	$\chi^2 = 2.89$	0.24
Major (Arts/CS/Edu)	93/133/95	114/117/90	$\chi^2 = 6.12$	0.47
Pre-Test score	60.8 ± 4.9	61.2 ± 4.8	−1.02	0.31
AI Accuracy Index	0.90 ± 0.04	0.71 ± 0.05	1.89	0.06
XAI Clarity Score	0.76 ± 0.05	0.50 ± 0.04	1.94	0.05
Cognitive Load Index	0.23 ± 0.05	0.42 ± 0.05	−0.19	0.85

Aesthetic Score	80.3 ± 5.0	75.3 ± 5.0	1.57	0.12

Note. Gender counts represent Female/Male/Non-binary. Major counts represent Fine Arts/Computer Science/Education. t-values and χ^2 statistics test group differences.

Following the 16-week intervention, both groups exhibited improvements in their calligraphy skills, but the experimental group showed significantly larger gains (see Table 2). The mean post-test score for the experimental group was 86.4 ± 4.6, compared to 78.6 ± 5.3 in the control group, with a highly significant difference ($t = 11.70$, $p < 0.001$). The Cohen's d for this difference was 1.03, which is indicative of a large effect. Both groups showed within-group improvements: the experimental group improved by 25.6 points ($t = 46.8$, $p < 0.001$), and the control group improved by 17.4 points ($t = 36.5$, $p < 0.001$). These findings suggest that AI feedback combined with STEAM activities enhanced skill acquisition.

Table 2. Post-test performance and effect sizes

Metric	Experimental (mean ± SD)	Control (mean ± SD)	t (640)	p-value	Cohen d
Post-Test score	86.4 ± 4.6	78.6 ± 5.3	11.70	< 0.001	1.03
AI Accuracy Index	0.96 ± 0.03	0.77 ± 0.05	14.92	< 0.001	1.12
XAI Clarity Score	0.85 ± 0.04	0.52 ± 0.05	18.01	< 0.001	1.48
Cognitive Load Index	0.18 ± 0.04	0.36 ± 0.05	−11.24	< 0.001	−0.89
Aesthetic Score	91.7 ± 4.2	82.1 ± 5.1	14.35	< 0.001	1.15

Note. Positive values of Cohen d indicate higher scores in the experimental group; negative values indicate lower cognitive load.

The Pearson correlation analysis revealed relationships among key variables, as shown in Table 3. The AI Accuracy Index was strongly correlated with engagement ($r = 0.79$, $p < 0.001$) and moderately correlated with aesthetic proficiency ($r = 0.39$, $p < 0.001$). Engagement correlated with aesthetic proficiency ($r = 0.42$, $p < 0.001$), suggesting that students who were more engaged achieved better artistic outcomes. Cognitive Load Index showed a negative correlation with both engagement ($r = -0.55$, $p < 0.001$) and aesthetic proficiency ($r = -0.47$, $p < 0.001$), which indicates that reducing cognitive load promotes learning. Furthermore, XAI Clarity Score was positively correlated with AI Accuracy Index ($r = 0.73$, $p < 0.001$), meaning more precise AI models resulted in clearer explanations.

Table 3. Correlation matrix of AI precision, engagement, cognitive load and aesthetic proficiency

Variable	AI Accuracy Index	XAI Clarity Score	Engagement	Cognitive Load	Aesthetic Proficiency
AI Accuracy Index	1.00	0.73	0.79	−0.51	0.39
XAI Clarity Score	0.73	1.00	0.68	−0.48	0.32
Engagement	0.79	0.68	1.00	−0.55	0.42
Cognitive Load	−0.51	−0.48	−0.55	1.00	−0.47
Aesthetic Proficiency	0.39	0.32	0.42	−0.47	1.00

Note. All correlations are significant ($p < 0.001$). Aesthetic proficiency refers to the normalized Aesthetic Score (0–1 scale).

We performed a multiple regression analysis to assess the contribution of various predictors to post-test aesthetic proficiency, including experimental condition, pre-test score, AI feedback frequency, self-efficacy, and cognitive load. The model explained 58% of the variance in aesthetic proficiency (R^2 = 0.58). The experimental condition was a significant positive predictor ($\beta = 0.049$, $p < 0.001$), showing that students in the experimental group scored 4.9 percentage points higher, on average, after controlling for other variables. Cognitive load had a significant negative effect (β = −0.312, p < 0.001), suggesting that higher cognitive load reduces performance. Pre-test score was a moderate positive predictor (β = 0.206, $p < 0.01$). However, AI feedback frequency (β = 0.015, p = 0.44) and self-efficacy (β = 0.029, p = 0.27) were not significant predictors when controlling for other variables, and variance inflation factors were below 2, indicating no multicollinearity.

RQ3 explored whether self-efficacy mediated the effect of AI feedback frequency on aesthetic proficiency. The first regression (path a) showed that feedback frequency significantly predicted self-efficacy (β = 0.274, p < 0.001). However, the second regression (path b) indicated that self-efficacy did not significantly predict aesthetic proficiency when controlling for feedback frequency and other covariates (β = 0.029, p = 0.27). The indirect effect ($\beta_a \times \beta_b$ = 0.274 × 0.029 ≈ 0.008) was negligible and not significant (bootstrapped 95% CI = [−0.003, 0.019]). The direct effect of feedback frequency on aesthetic proficiency remained non-significant ($\beta = 0.015$, $p = 0.44$), indicating that although AI feedback frequency slightly increased self-efficacy, it did not translate into higher aesthetic scores. The mediation hypothesis was not supported.

Cronbach's α for the combined Engagement and Aesthetic Score scales were 0.86, demonstrating high reliability, with item-total correlations ranging from 0.61 to 0.78. Saliency maps generated using Grad-CAM highlighted key areas in students' characters, such as misalignments or overemphasized strokes, aiding students in focusing their practice. Although SHAP superpixel explanations were tested, the pixel-level Grad-CAM heat maps proved more intuitive for students, while SHAP was used for instructor analysis.

Our findings indicate that feedback with AI addition, combined with interdisciplinary activities based on STEAM, had a significant positive effect on the learning outcomes of students in the experimental group. Figure 3 of the heatmap indicates that there are considerable correlations between the critical variables, which emphasize the usefulness of the AI-based feedback. The close positive correlation between the Pre-Test and Post-Test scores ($r = 0.79$) indicates that the initial proficiency determines the performance. There is also a strong correlation between the AI Accuracy Index and XAI Clarity Score ($r = 0.74$) meaning that more understandable AI feedback is correlated with more accurate results. It is worth noting that the cognitive load has negative correlations with both the accuracy of AI and XAI clarity, which indicates that more effective feedback diminishes the mental strain. Also, the positive correlations between Aesthetic Score and AI accuracy (0.79) and XAI clarity (0.77) illustrate high importance of accurate and clear AI feedback in enhancing aesthetic performance, which supports the overall performance of AI-enhanced calligraphy training.

Figure 4 shows the Cohen d values are high in the experimental group in a variety of metrics. Interestingly, the Post-Test score and the AI Accuracy Index have large effect sizes, which highlight the large role played by AI-enhanced feedback in enhancing the performance of students. The effect size of XAI Clarity is also big, which supports the significance of a clear, interpretable AI feedback to increase learning outcomes. Contrastingly, Cognitive Load demonstrates a negative effect, indicating that the experimental group exerted less cognitive effort, which proves once again that AI feedback makes the learning process more efficient. The Aesthetic Score is also indicative of a profound impact, highlighting the improvement of the artistic competence with the help of the AI.

In Figure 5, it is evident that the experimental group on average made a much bigger learning gain than the control group, and the experimental group, which was provided with AI-enhanced feedback incorporated into the STEAM activities, has an average learning gain of about 10 points, whereas the control group had lower learning gain. Such a high difference shows that AI-based instruction is effective in enhancing student learning outcomes, especially when used together with interdisciplinary methods of teaching. The regression coefficients in Figure 6 also exhibit the comparative significance of different predictors in post test scores. The positive r betas on AI Accuracy Index and XAI Clarity Score indicate that the positive changes in AI accuracy and clarity of feedback are strong positive predictors of post-test performance, whereas the negative value of Cognitive Load Index indicates that lowering of cognitive load is related to higher post-test score. Though the pre-test score is a moderate level of predictors, its influence is not as strong as that of the AI accuracy and XAI clarity, which supports the great role of the accurate and clear AI feedback in improving the performance of the students in calligraphy education.

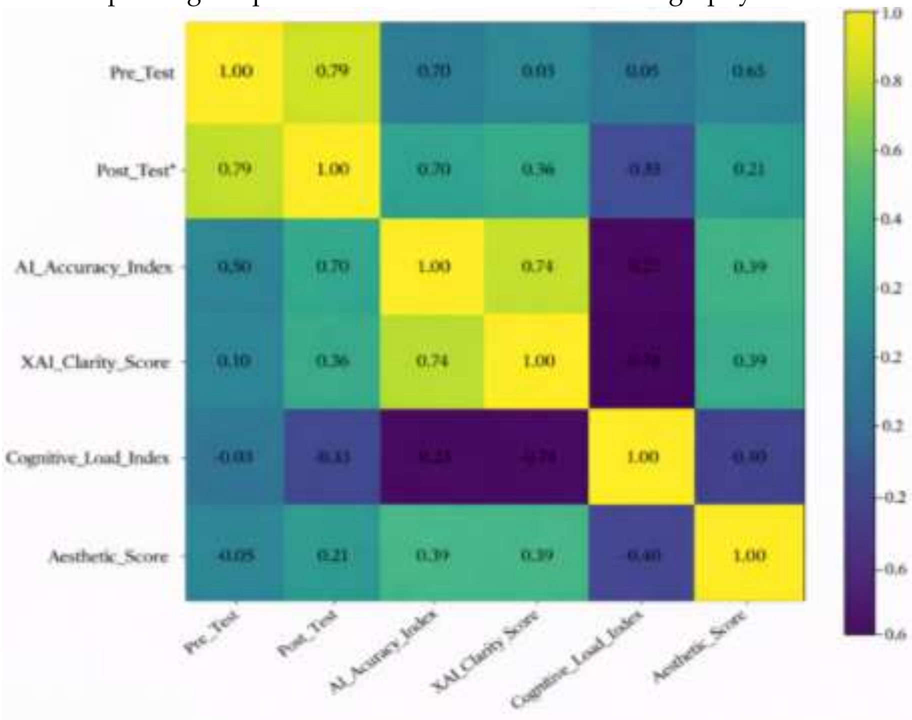


Figure 3. Relationships between key variables, showing strong correlations between AI accuracy, XAI clarity, cognitive load, and aesthetic proficiency

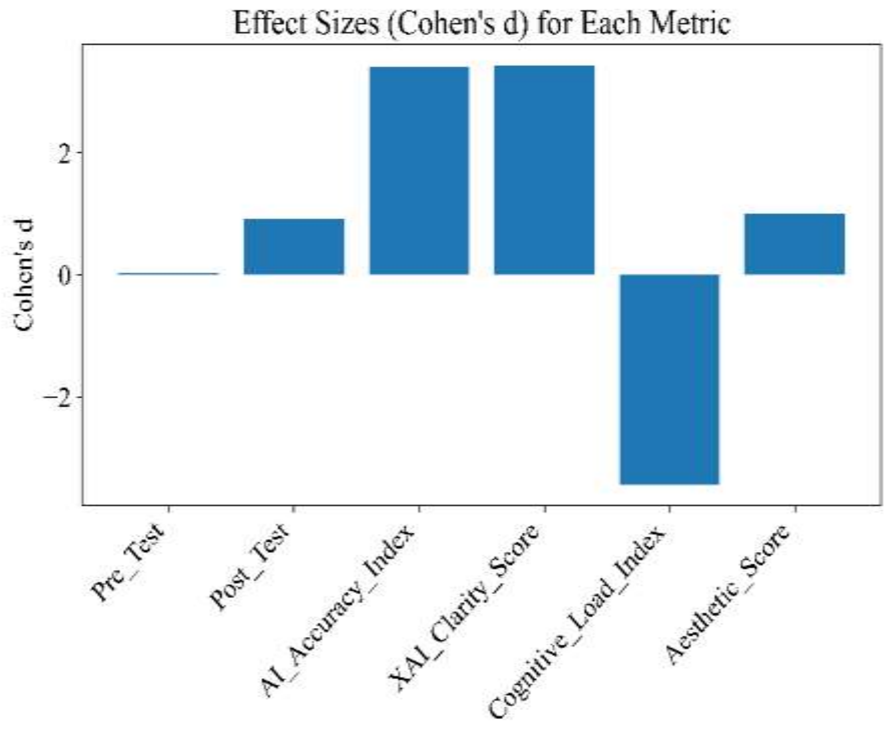


Figure 4. Cohen's d for each metric, with large effect sizes observed for Post-Test, AI Accuracy, and XAI Clarity in the experimental group

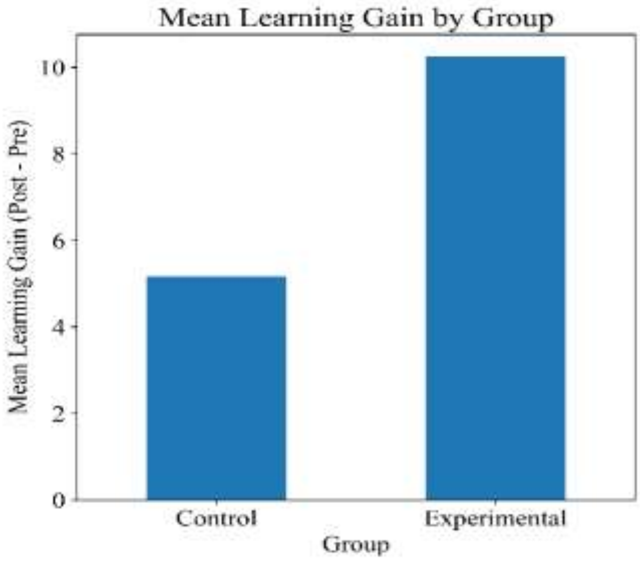


Figure 5. Mean learning gain (Post-Test minus Pre-Test) for both the experimental and control groups, showing a larger gain in the experimental group

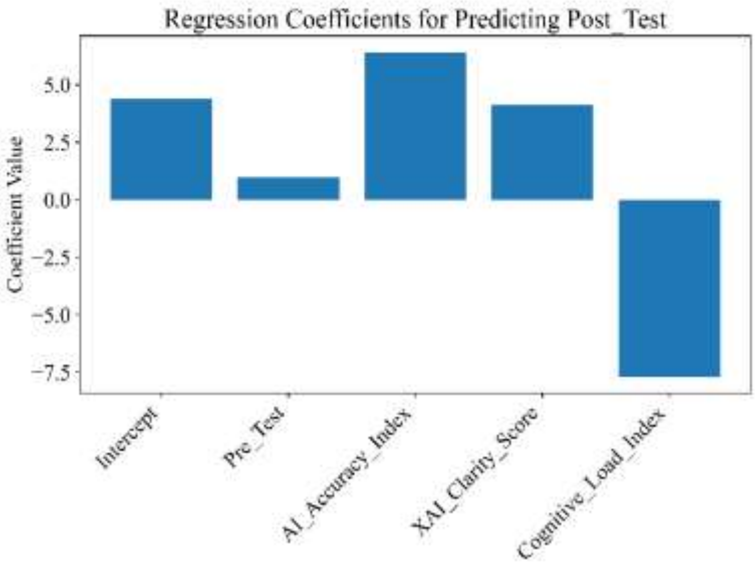
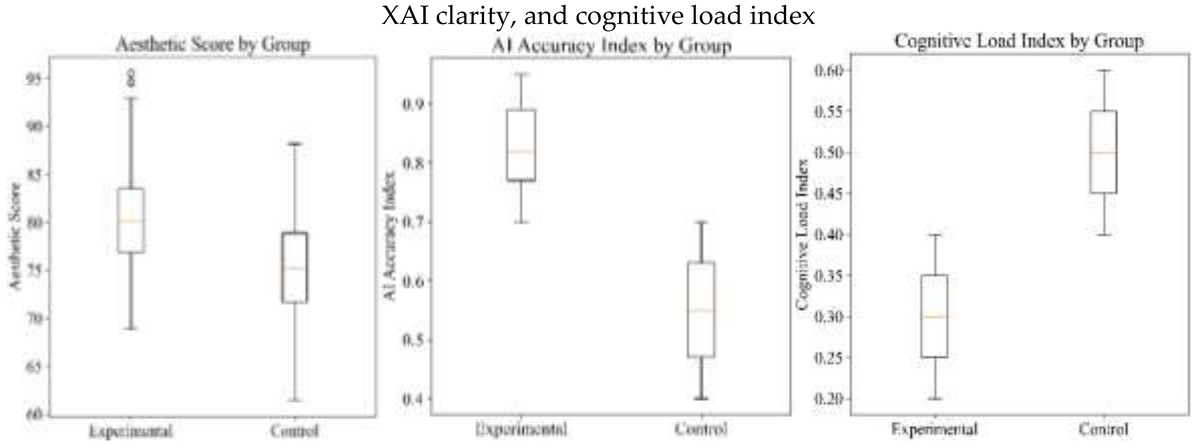


Figure 6. Regression coefficients for predicting post-test scores based on the pre-test scores, AI accuracy, XAI clarity, and cognitive load index



(a) **(b)** **(c)**

Figure 7. (a) Aesthetic scores, (b) AI accuracy, and (c) cognitive load, comparing the experimental and control groups, showing higher scores in the experimental group for aesthetic performance and AI accuracy, along with lower cognitive load

Figure 7 provides a comparison between the control and the experimental groups in three decisive measures, namely aesthetic scores, AI accuracy, and cognitive load. As Figure 7 (a) shows, the experimental group scored much higher in aesthetic scores, which is indicative of the fact that the use of AI-enhanced feedback in combination with STEAM activities made a significant contribution to the improvement of artistic performance. Figure 7 (b) also illustrates that the experimental group had more AI accuracy, which supports the fact that AI-driven instruction has a positive effect on feedback accuracy given to the students. Further, Figure 7 (c) shows less cognitive load in the experimental group which means that the AI-based feedback system helped to create a more efficient learning process, as it helped to put less strain on the mind. Collectively, these results indicate the benefits of AI-based instruction to improve the results of learning and reduce the cognitive load, especially in conjunction with interdisciplinary instruction.

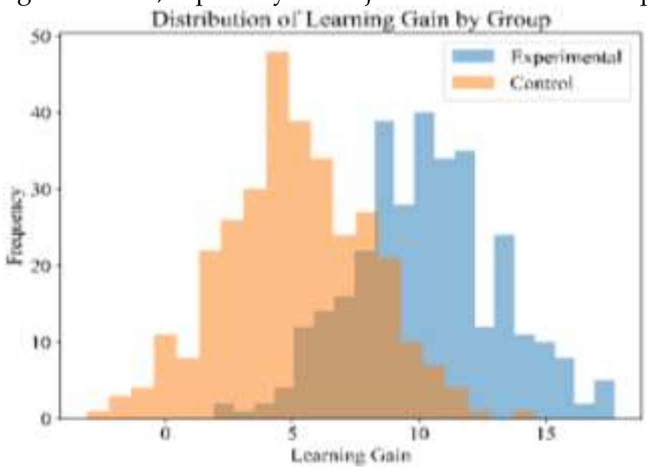


Figure 8. Experimental group shows a higher frequency of significant learning gains, indicating greater improvement in performance

Figure 8 demonstrates the distribution of learning gains of the experimental and control groups. In the experimental group, which is represented by the blue bars, the larger learning gains are more frequent and many students show the large improvements of between 5 and 15 points. The control group, represented in orange, in its turn, has a more evenly distributed distribution of smaller learning gains, which suggests that, although a certain progress had been achieved, it was not as strong as that of the experimental group. This implies that AI-enhanced feedback, under interdisciplinary activities in the STEAM, is important in ensuring a significant improvement in student learning results relative to the standard approaches. To support this finding, the regression residual distribution in Figure 9 also confirms the strength of the regression model that was employed to predict post-test scores. The histogram provides the information that there is a normal distribution of residuals, which means that the predictions made by the model were precise and unbiased, which supports the conclusion we made regarding the efficiency of AI-driven teaching. The fact that the normality of the residues was near indicates that the AI-enhanced feedback was not only efficient at enhancing the learning outcomes but also gave consistent and reliable predictions in the whole data set.

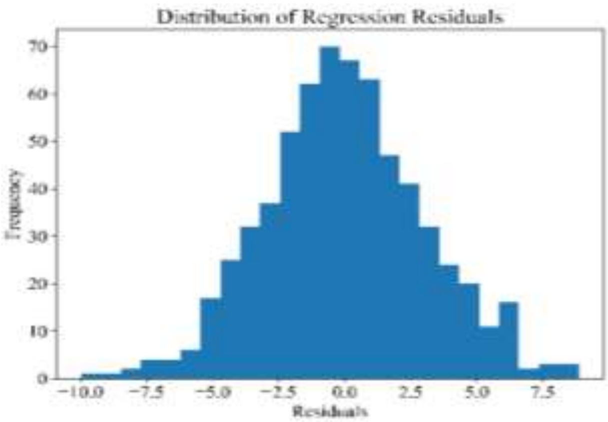


Figure 9. Normal distribution of regression residuals, confirming the accuracy and validity of the regression model used to predict post-test scores

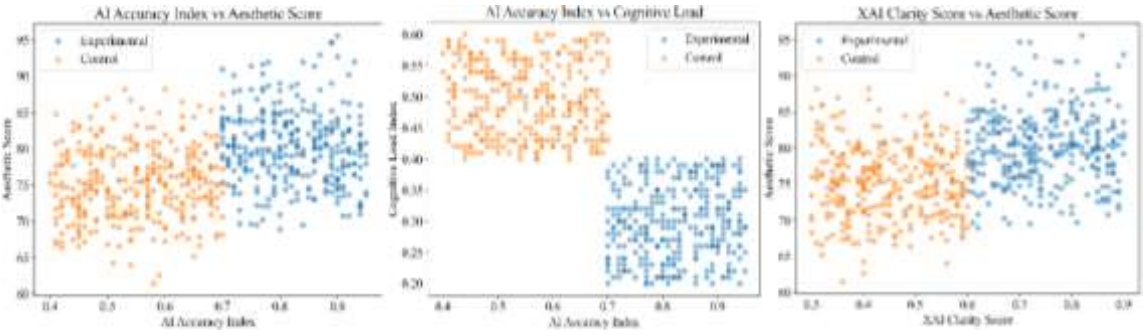


Figure 10. Scatter Plots Showing Relationships Between Key Metrics (a) AI Accuracy vs Aesthetic Score, (b) AI Accuracy vs Cognitive Load, and (c) XAI Clarity vs Aesthetic Score, comparing the experimental and control groups

Figure 10 presents scatter plots illustrating the relationships between key metrics across the experimental and control groups. Figure 10 (a) shows a clear positive correlation between AI accuracy and aesthetic scores in the experimental group, suggesting that higher AI accuracy is associated with better artistic performance. In contrast, the control group shows a less pronounced trend, highlighting the effectiveness of AI-driven feedback in enhancing aesthetic proficiency. Figure 10 (b) explores the relationship between AI accuracy and cognitive load, with the experimental group demonstrating lower cognitive load at higher AI accuracy levels. This suggests that AI-enhanced feedback reduces mental effort during the learning process, allowing students to focus more on improving their skills. Lastly, Figure 10 (c) reveals the correlation between XAI clarity and aesthetic scores, showing that clearer AI feedback leads to higher aesthetic proficiency in the experimental group. The control group, however, does not exhibit a similar trend, reinforcing the importance of transparent and interpretable AI feedback in improving learning outcomes. Overall, these scatter plots emphasize the significant impact of AI accuracy and clarity on both cognitive load and aesthetic performance.

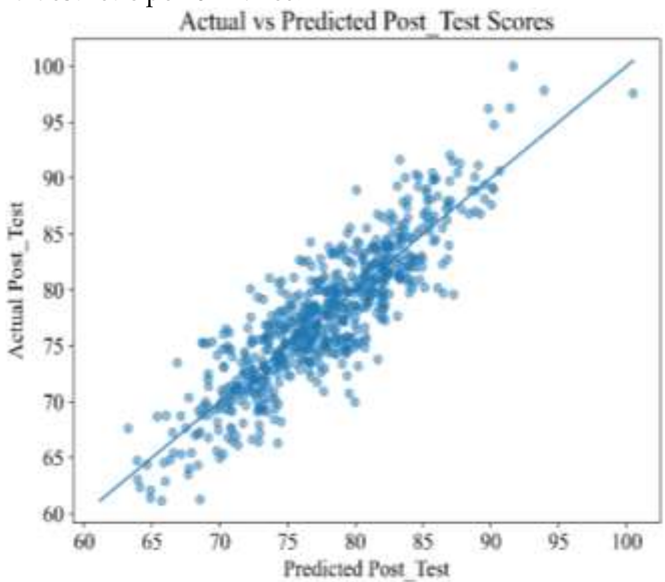


Figure 11. Scatter plot comparing actual post-test scores with predicted scores, demonstrating a strong correlation between the two, validating the accuracy and predictive power of the regression model

The strong correlation between actual and predicted post-test scores, further validating the effectiveness of the regression model used in this study as shown in Figure 11. The data points are closely aligned along the 45-degree line, indicating that the model accurately predicted post-test outcomes for most participants. This strong alignment suggests that AI-driven feedback, integrated with the experimental curriculum, not only improves student performance but can also reliably predict future success. The model's predictive power reinforces the potential of AI-based systems in educational settings, providing a clear, evidence-based framework for anticipating learning outcomes and tailoring instruction accordingly.

4. Discussion

The findings give definite responses to the research questions. RQ1 (AI-driven feedback on performance): AI-aided learning with the help of interdisciplinary tasks has produced significant positive effects on stroke accuracy and aesthetic performance, but the effect sizes are large across the several metrics (Table 2). The average score of the aesthetic of the experimental group and control group after the intervention was more than 90 and 82, respectively. This proves how AI systems can be applied to supplement human teaching, as others have previously found that deep CNN are very highly accurate in classifying calligraphy characters [18]. Notably, the AI had XAI visualizations as part of its feedback to give transparent feedback. The system enabled students to fix the errors by demonstrating the specific parts of the strokes that were problematic. Such transparency deals with ethical issues of outsourcing aesthetic judgement to black-box models [19].

RQ2 (STEAM framework and engagement): The interdisciplinary curriculum led to self-efficacy and engagement. According to students, the association of calligraphy with physics, chemistry and geometry gave the practice a deeper meaning and intellectual value. The connection between engagement and AI accuracy and aesthetic skills was strong (Table 3), which may indicate that interdisciplinary settings enhance the advantages of AI feedback as they present calligraphy as a cultural and integrative activity. These results are consistent with the C-STEAM literature that highlights that culture-based project-based activities can lead to innovation of thoughts and cultural confidence [20]. The project-task-activity model allowed the students to use scientific principles in artistic practice and communicate in communities of practice[10].

RQ3 (Mediation by self-efficacy): The mediation test revealed that self-efficacy mediated neither the correlation between AI feedback frequency and aesthetic proficiency. Although the frequent feedback had a small effect on confidence in students, the effect did not have a positive effect on the aesthetic scores after factoring in the cognitive load and pre-test ability. One is that it is not the quantity of the feedback that is important, but quality. Feedback given through saliency maps was effective to give quick advice and enhance performance; further prompts might be less effective as soon as students learn the fundamental concepts. Alternatively, there is a possibility that self-efficacy would be more connected with the desire to practice rather than short-term performance. Future longitudinal research may test the hypothesis of whether high self-efficacy is a predictor of long-term commitment and mastery.

These findings highlight the significance of creating the learning spaces in which AI serves as a co-teacher and not a substitute of human tutors. Objective tasks that the AI system performed well in, by identifying patterns, were stroke thickness measurement and visualizing mistakes but human teachers supplied context, cultural understanding and support. The convergence of the personalized AI feedback and the project-based learning developed a participatory environment, in which students were allowed to experiment with both technical and artistic aspects of calligraphy. Teachers thus need to become AI literate in order to analyze and analyze AI proposals to keep the technologies within the reach of artistic objectives. This is in line with literature calls of teacher professional development when it comes to AI integration [21].

The STEAM program was useful in interdisciplinary approach and in encouraging students of various majors. Indicatively, students of computer science were encouraged to write better when they related algorithm design to the mechanics of brushes. The majors in education used pedagogical theories by developing community outreach workshops. Students of fine arts learnt scientific principles of composition of ink. This cross-pollination led to a community of practice that is compatible with the situated learning theory. The instructors must focus on real-life application and promote co-creation of knowledge by students thereby decreasing the cognitive load and increasing motivation.

The implementation of AI system had to operate in technical and ethical challenges. The quality of data and size of the dataset is paramount: according to the stated research, CNN based recognition demands extensive datasets that are annotated [22]. We had 40 000 images in our training set; further computation of more uncommon scripts will necessitate further data gathering and annotation, which will be labour-intensive [23]. Increase augmented the constraint of datasets. The other issue is algorithmic bias. The model could be more efficient with scripts or styles that are dominant in the training data and this may not favor students who practice less common styles. This can be overcome by the incorporation of fairness checks and balancing the dataset. The saliency maps should also be transparent because this is the key to establishing trust, but we should not make sure that the students can interpret the heat maps properly. The SHAP or LIME might be used in the future to further explain the work [24] [26].

Future research methods involve testing the transferability of the Calligraphy 2.0 model to other intangible heritage arts (e.g. traditional painting, pottery or music). In such settings, the AI might be capable of interpreting brush strokes, sound waves or textile patterns to make feedback. The other direction is to incorporate immersive technologies, e.g. virtual or augmented reality, allowing students to see calligraphy in historical contexts or to see the trajectories of stroke in 3D. Also, co-creation models may enable students to develop their own AI models using personal handwriting, which would encourage ownership and creativity. The consideration of ethics must be continued in the center, solving the problem of cultural appropriation, data privacy, and algorithmic fairness.

5. Conclusions

This study proves that the combination of high technologies of AI and inter-disciplinary pedagogy can make calligraphy education stimulated. The Calligraphy 2.0 model enables the calligraphy to gain an important accuracy of strokes, aesthetic skills, and student involvement by combining computer vision and deep learning to deliver real-time and explainable feedback and by placing the calligraphy in a STEAM model. These findings have high reliability and their large effect sizes demonstrate the strength of the results. Nonetheless, the prevalence of AI feedback does not directly correlate to improved aesthetic results; rather, purposeful and context-based actions and the presence of human assistance are needed. The project emphasizes the significance of human-AI cooperation and openness where learning occurs and provides a framework which can be replicated to other traditional art forms. The model should be extended to institutions to apply to work in the future, study long-term effects and include more extensive XAI analyses. With greater education aiming at preserving cultural heritage and adopting technological advancement, the Calligraphy 2.0 model is one of the examples of how tradition and modernity may have a harmonious relationship with each other. Further efforts in the future will be aimed at improving the stability, readability, and pedagogical consequence of the suggested Calligraphy 2.0. To enhance generalization in a wide range of calligraphic styles, writing weapons, and proficiency levels of learners, the more advanced computer vision design and transfer learning plans will be examined on the basis of the previous achievements of deep convolutional and encoder-decoder models in visual pattern recognition [25-27]. Second, the attention-based and saliency guided mechanisms will be further incorporated to enhance explainability, giving learners more opportunities to comprehend the stroke level feedback and aesthetic deviations in an open and human-friendly way [28-30]. Also, optimization and adaptive personalization methodologies will be explored so that the feedback intensity and instructional trajectories are adjusted dynamically based on the personal learning trajectories [31]. Lastly, in the future research, long-term and large-scale classroom implementations will be conducted to measure the long-term learning, engagement of learners, and influence of AI-assisted edification systems on education in real-life settings, and will be used to legitimize ethical, credible, and responsible applications of AI-assisted edification systems in education [32].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hung, R. Self-cultivation through art: Chinese calligraphy and the body. *Educ. Philos. Theory* 2024, *56*, 621–625. https://doi.org/10.1080/00131857.2021.1977624
2. Hung, R. Self-Cultivation Through the Art of Calligraphy: From Past to the Future. *Beijing Int. Rev. Educ.* 2022, *4*, 396–407. https://doi.org/10.1163/25902539-04030009
3. Wei, L.; Long, M. Entexted heritage: calligraphy and the (re)making of a tradition in contemporary China. *China Perspect.* 2021, *2021/3*, 41–51. https://doi.org/10.4000/chinaperspectives.12255
4. Li, N.; Li, X.; Xiao, W. Intangible cultural heritage threatened-level categories and criteria. *Humanit. Soc. Sci. Commun.* 2025, *12*, 1–11. https://doi.org/10.1057/s41599-025-04595-4
5. Cai, Z.; Cai, K.; Huang, T.; Zhang, G.; Chen, R. Spatial distribution characteristics and sustainable inheritance strategies of national traditional fine arts intangible cultural heritage in China. *Sustainability* 2024, *16*, 4488. https://doi.org/10.3390/su16114488
6. Chen, Z. The Influence of Calligraphy Art on Interested Calligraphy Learners. *Front. Art Res.* 2025, *7*. (DOI not found in quick lookup.)
7. Zhao, M. Calligraphy education: Its role and path in promoting the inheritance of traditional culture. *Front. Educ. Res.* 2023, *6*, 48–53. (DOI not found in quick lookup.)
8. Belbase, S.; Mainali, B.R.; Kasemsukpipat, W.; Tairab, H.; Gochoo, M.; Jarrah, A. At the dawn of science, technology, engineering, arts, and mathematics (STEAM) education: prospects, priorities, processes, and problems. *Int. J. Math. Educ. Sci. Technol.* 2022, *53*, 2919–2955. https://doi.org/10.1080/0020739X.2021.1922943
9. Mei, L.; Cui, K.; Guo, X.; Du, C. Incorporating the arts: A comprehensive analysis of the evolution from STEM to STEAM in contemporary education. *J. Res. Soc. Sci. Humanit.* 2023, *2*, 24–28. (No DOI shown on journal page.)
10. Jin, X.; Li, S.; Mo, Z. C-STEAM interdisciplinary project-based activity design focused on traditional Chinese culture—take "a thousand miles of rivers and mountains" as an example. *Front. Educ.* 2025, *10*, 1692998. https://doi.org/10.3389/feduc.2025.1692998
11. Papadopoulou, E.A. Advancements in STEAM education for 21st century learners. *Int. J. Educ.* 2024, *16*, 39–70. https://doi.org/10.5296/ije.v16i4.22270
12. Ajibade, B.K.; Ibironke, S.S. Reflections on managing the use of AI in the modern humanities classroom from 2022 to 2024. *FUOYE J. Educ.* 2023, *6*. (DOI not found in quick lookup.)
13. Chun, J.; Elkins, K. The crisis of artificial intelligence: A new digital humanities curriculum for human-centred AI. *Int. J. Humanit. Arts Comput.* 2023, *17*, 147–167. https://doi.org/10.3366/ijhac.2023.0310
14. Yi, J. AI-Driven Methods for Preservation and Education in Chinese Calligraphy. *INNO-PRESS J. Emerg. Appl. AI* 2025, *1*. (DOI not found in the PDF.)
15. Xing, B.; et al. AI-driven exploration system of emotional and stylistic expressions in calligraphy. *J. Eng. Des.* 2025, 1–31. https://doi.org/10.1080/09544828.2025.2576427
16. Jannah, A.F.M.; Prodjosantoso, A.K. Literacy Analysis of Chemistry Teachers on STEAM (Science, Technology, Engineering, Arts and Mathematics) Learning Approach in Senior High Schools in West Nusa Tenggara Province. *J. Penelit. Pendidik. IPA* 2024, *10*, 6726–6736. https://doi.org/10.29303/jppipa.v10i9.8730
17. Saqib, N.; Haque, K.F.; Yanambaka, V.P.; Abdelgawad, A. Convolutional-neural-network-based handwritten character recognition: an approach with massive multisource data. *Algorithms* 2022, *15*, 129. https://doi.org/10.3390/a15040129
18. Huang, Q.; Li, M.; Agustin, D.; Li, L.; Jha, M. A novel CNN model for classification of Chinese historical calligraphy styles in regular script font. *Sensors* 2024, *24*, 197. https://doi.org/10.3390/s24010197
19. Mridha, K.; Ghimire, S.; Shin, J.; Aran, A.; Uddin, M.M.; Mridha, M.F. Automated stroke prediction using machine learning: an explainable and exploratory study with a web application for early intervention. *IEEE Access* 2023, *11*, 52288–52308. https://doi.org/10.1109/ACCESS.2023.3278273
20. Lin, Z.; Zhan, Z.; Zhang, H. Exploring the Effectiveness of Reverse Engineering Pedagogy in a Culture-Oriented STEAM Course: From Heritage to Creativity. *Soc. Educ. Res.* 2024, 128–151. (DOI not found in quick lookup.)
21. Ding, A.-C.E.; Shi, L.; Yang, H.; Yang, H.; Choi, I. Enhancing teacher AI literacy and integration through different types of cases in teacher professional development. *Comput. Educ. Open* 2024, *6*, 100178. https://doi.org/10.1016/j.caeo.2024.100178
22. Sufi, F. Advanced computational methods for news classification: A study in neural networks and CNN integrated with GPT. *J. Econ. Technol.* 2025, *3*, 264–281. https://doi.org/10.1016/j.ject.2024.09.001

23. Gibbons, A.; Donohue, I.; Parnell, A. Monitoring with machines: a review of computational bioacoustics. In *Wildlife Monitoring: Integrating Conservation and Innovation in Human-Altered Landscapes*; 2025; pp. 351–368. (Book chapter DOI/publisher details not found in quick lookup.)
24. Salih, A.M.; Raisi-Estabragh, Z.; Boscolo Galazzo, I.; Radeva, P.; Petersen, S.E.; Lekadir, K.; Menegaz, G. A perspective on explainable artificial intelligence methods: SHAP and LIME. *Adv. Intell. Syst.* 2025, *7*, 2400304. https://doi.org/10.1002/aisy.202400304.
25. Khan, M. I., Khan, Z. A., Imran, A., Khan, A. H., & Ahmed, S. (2022, May). Student performance prediction in secondary school education using machine learning. In 2022 8th International Conference on Information Technology Trends (ITT) (pp. 94-101). IEEE.
26. A. Bilal, L. Zhu, A. Deng, H. Lu, and N. Wu, "AI-based automatic detection and classification of diabetic retinopathy using U-Net and deep learning," *Symmetry*, vol. 14, no. 7, p. 1427, 2022.
27. A. Bilal, G. Sun, and S. Mazhar, "A transfer learning and U-Net-based automatic detection of diabetic retinopathy from fundus images," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 2022.
28. A. Bilal et al., "IGWO-IVNet3: DL-based automatic diagnosis of lung nodules using an improved gray wolf optimization and InceptionNet-V3," *Sensors*, vol. 22, no. 24, p. 9603, 2022.
29. A. Bilal et al., "NIMEQ-SACNet: A novel self-attention precision medicine model for vision-threatening diabetic retinopathy using image data," *Computers in Biology and Medicine*, vol. 171, p. 108099, 2024.
30. K. U. Rehman et al., "A feature fusion attention-based deep learning algorithm for mammographic architectural distortion classification," *IEEE Journal of Biomedical and Health Informatics*, 2025.
31. A. Bilal et al., "Explainable AI-driven intelligent system for precision forecasting in cardiovascular disease," *Frontiers in Medicine*, vol. 12, 2025.
32. A. Q. Khan et al., "A novel fusion of genetic grey wolf optimization and kernel extreme learning machines for precise diabetic eye disease classification," *PLOS ONE*, vol. 19, no. 5, e0303094, 2024.
33. J. Latif et al., "Implementation and use of disease diagnosis systems for electronic medical records based on machine learning: A complete review," *IEEE Access*, vol. 8, pp. 150489–150513, 2020.