JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 38, XXX-XXX (2022)
DOI: 10.6688/JISE.20220X_38(X).00XX

Towards Transparent Recognition of Calligraphic Fonts with Explainable AI

WEN-IAN CHEN[1‡], YING-LIN HSU[2‡] AND JIA-LANG XU[3*]
1 Department of Chinese Literature, National Chung Hsing University, Taichung City 402, Taiwan
2 Department of Applied Mathematics and Institute of Statistics, National Chung Hsing University, Taichung City 402, Taiwan
3 Department of Applied Statistics, National Taichung University of Science and Technology, Taichung, Taiwan
*E-mail: jlxu@nutc.edu.tw**

Chinese calligraphy has traditionally been taught via intensive practice with copybooks, requiring learners to begin at an early age and learn through repetitive imitation. Although this method has long been considered the most effective pathway for mastery, the digital learning era has introduced new opportunities to support the appreciation and instruction of calligraphy via technology. Recent educational innovations have emphasized the integration of art into the framework of science, technology, engineering, and mathematics (STEM), leading to STEAM education, promoting creativity, interdisciplinary learning, and expanded cognitive modes. This study develops a dataset featuring characters from four prominent calligraphers in Chinese history and applies deep learning models to identify calligraphic fonts. To enhance transparency, this work presents an explainable learning framework that highlights the critical image regions contributing to recognition. This framework enhances model interpretability and offers pedagogical insight, enabling learners to gain a better understanding of structural features of calligraphy and focus on critical elements during practice. The experimental results demonstrate that the EfficientNet model achieved an accuracy of 98.34% in font recognition. Moreover, explainable artificial intelligence identified the salient areas of character images, supporting teachers and students in appreciating the aesthetic principles of traditional calligraphy while fostering more informed, creative learning approaches.

Keywords: E-Learning, STEAM, Explainable AI, Deep Learning, Calligraphy

1. INTRODUCTION

Chinese calligraphy has a long history as font art. Chinese calligraphy was written on tortoise shells and animal bones with brushes in the oracle bone period, although the writing at that time was for the convenience of engraving, with less attention to the brushwork. With the evolution of fonts and the development of brushes and improved writing paper, calligraphers began to focus on penmanship and the presentation of lines. After oracle bone and bronze inscriptions, the continuous evolution of fonts, including seal script, official script, cursive script, regular script, and running script, became the principal characters circulating across various dynasties. Nearly every calligraphy font has a corresponding outstanding calligrapher style, producing calligraphy work with distinctive styles. Chinese calligraphy is extensive and pro-found. In addition to popularizing various calligraphy styles in different eras, calligraphers have also created schools and rich calligraphy theories

during the evolution of calligraphy, explaining why it can become a unique discipline. Individuals can reveal their personal characters through their writing, and it is easy to observe their accomplishments from their work. Calligraphy is a visual art and has calligraphy principles that support its theory. This profound Chinese culture necessitates a lifetime of study to comprehend. However, learning must start from imitating a calligrapher's calligraphy stroke by stroke. Students typically feel frustrated when they cannot learn calligraphy skills because they must repeatedly practice without observing obvious progress, easily becoming bored and giving up on learning.

With the gradual advancement of types of artificial intelligence (AI), researchers could be prompted to incorporate such technology into various fields to solve problems arising in those fields. Calligraphy applications have motivated many researchers to study and investigate numerous problems that can be applied in this field. For example, Lin et al. explored the potential of AI to drive innovation in calligraphy and emphasized the importance of AI in the development of traditional calligraphic art. They encouraged integrating art and technology to advance contemporary calligraphy practices [38]. In addition, Du argued that AI technology can offer diversified teaching approaches while addressing challenges in calligraphy education. The study suggested that universities should strengthen their collaboration with technology enterprises to enhance the development of calligraphy education [37]. Further, Jiang et al. investigated the application of deep learning to variations in the Chinese calligraphic landscape. They suggested that such research may inspire geographers to better understand the evolving role of AI in the modern era [36]. Xiao et al. proposed the invention of a content-style generative adversarial network (GAN) method that can transfer the characters of each calligraphic style in addition to the style transfer of the calligraphic structure [1]. Moreover, Lin et al. proposed a robot that can autonomously write Chinese calligraphy, automatically matching strokes and generating text [2]. Zhao et al. proposed a novel approach to feature extraction using a convolutional neural network (CNN) model followed by a calligraphy-clustering attention module and a me-ta-style matrix for embedding calligraphy. The results were input into a conditional gated recurrent unit model, where they generated highly similar handwriting [3]. Wu et al. proposed a model that combined a GAN and deep reinforcement learning to learn strokes in calligraphic images. They concluded that their method can be effective for robotic calligraphy [4].

Valeri et al. explored how the chat generative pretrained transformer (ChatGPT) model can be experienced and applied in science, technology, engineering, and mathematics (STEM) education using a mixed-methods approach. The study offered insight into student experiences with ChatGPT and suggested that such perspectives are crucial for future educational development [39]. Zucker et al. explored whether the parental expectations, values, and costs of STEM are related to parental involvement in science and math activities. They concluded that parents are attracted by the value of STEM rather than the expectations and costs [5]. Mutambara and Bayaga concluded that teachers and parents of learners in STEM at the high school level were less receptive to using cell phones for learning and proposed using a high school cell phone learning acceptance model, which was effective in ex-plaining and predicting the acceptance of learners, teachers, and parents in rural are-as [6]. Morris et al. explored whether using a model for learning rural-related knowledge could increase student engagement in STEM learning. They concluded that the model in-creased engagement and knowledge retention [7]. Sun et al. investigated whether

STEM learning attitudes affect elementary school students' computational thinking skills. They concluded that girls have better learning attitudes toward STEM and that computational thinking skills undergo critical development in the fourth grade of elementary school [8]. Zheng et al. argued that self-regulation is a crucial aspect in STEM tasks and assessed whether competent self-regulated learners, cognitively oriented learners, learners who think backward, and learners with the lowest self-regulation have different learning outcomes. They concluded that competent learners can evaluate information they have obtained. Cognitively oriented learners underestimate their self-adjustment. Learners who think backward can know the outcome of the task, and learners with the lowest self-adjustment overestimate their ability [9].

This study pioneers an interdisciplinary framework that integrates traditional Chinese calligraphy education with AI technology to transform the conventional experiential learning process into a data-driven and interpretable instructional model. Unlike prior research that has focused on character recognition accuracy, this study emphasizes the feasibility of CNNs for Chinese calligraphy character classification and the pedagogical interpretability and educational applicability of deep learning models. The novelty of this research lies in three critical aspects. First, this research systematically evaluates whether various CNN architectures can be trained to recognize stylistically diverse Chinese calligraphy characters, which often exhibit significant intraclass variation due to personal brushwork styles. This work addresses a more complex recognition task than standard printed character datasets. Second, beyond the performance comparison, this study integrates explainable AI (XAI) techniques to visualize and analyze the attention regions learned by each model. This approach enables the examination of whether the models capture essential stroke structures, brush pressure transitions, and spatial composition patterns that reflect each calligrapher's stylistic emphasis. Third, this study bridges technical model interpretation with educational application by translating model-discovered crucial features into pedagogically meaningful guidance for teachers and learners.

The remainder of this article is organized as follows. Section 2 reviews InceptionV3, MobileNet, EfficientNet, and XAI. Next, Section 3 introduces the research framework and method. Then, Section 4 details the results and performance evaluation. Section 5 discusses the findings, and Section 6 presents the conclusion.

2. LITERATURE REVIEW

The work reviewed in this research is divided into sections based on the model. The review covers the InceptionV3, MobileNet, EfficientNet, and XAI models.

2.1 InceptionV3

Szegedy et al. proposed the InceptionV3 model (Fig. 1) [10]. In addition, Castro-Zunti et al. designed a model for predicting rib fractures using a migratory learning approach for model training. They concluded that the first seven blocks of InceptionV3 could improve accuracy and training time [11]. Ishengoma et al. suggested a technique for image modification using corner-point detection and concluded that the accuracy of the InceptionV3 model could be improved from 96.75% to 100% using this method [12]. Further,

Liu et al. presented a prediction model for prostate cancer ultrasound maps, concluding that the combination of the S-Mask R-CNN and Inception-v3 models could effectively make predictions [13]. Moreover, Apostolopoulos et al. designed an effective predictive cardiovascular disease model using a combination of InceptionV3 and the random forest method [14].

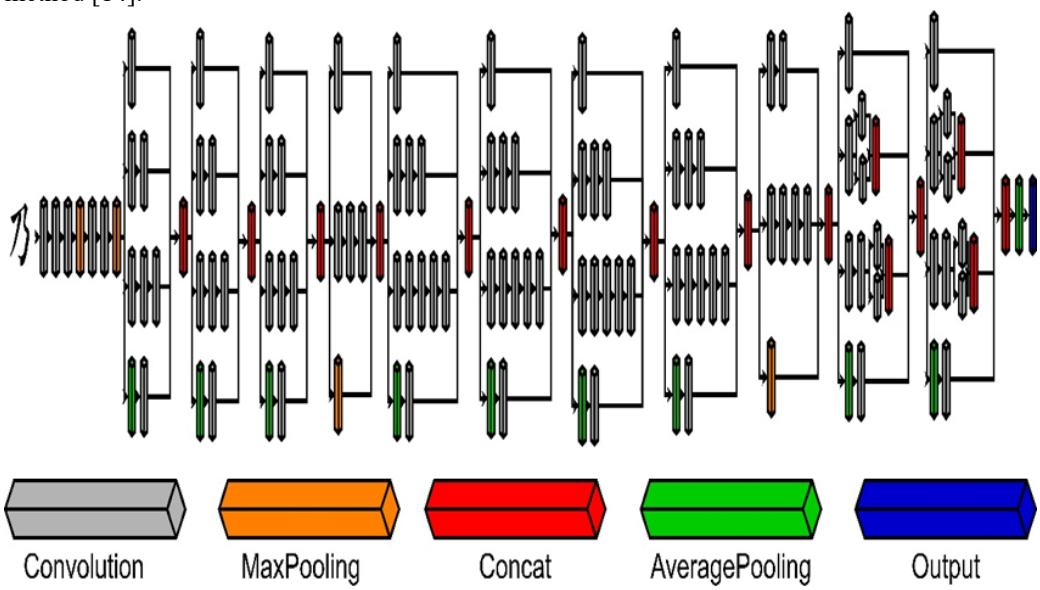


Fig. 1. InceptionV3 architecture

2.2 MobileNet

The MobileNet model applies depthwise separable convolution as the basic architecture (Fig. 2) [15]. This architecture can reduce computational time compared with the traditional convolutional layer. In addition, Hong et al. compared the performance of various models on knowledge distillation, finding that MobileNetv1 achieves better results in compression [16]. Chen et al. proposed a model that introduces location-wise soft attention into a pretrained MobileNet model to identify crop diseases [17]. Elwirehardja and Prayoga designed a model for palm-oil maturity prediction on mobile devices, and Mo-bileNetv1 produced good prediction results, albeit slightly inferior to those of Efficient-NetB0 [18]. Further, Goceri suggested a model for diagnosing patient skin diseases using a new hybrid loss function and a modified MobileNet architecture to reduce patient contact [19]. Nan et al. applied a prediction model for facial expressions by incorporating an attention mechanism into a MobileNet model [20]. Additionally, Wang et al. combined a MobileNet model with an attention mechanism to predict safety hazards at power stations early and reduce casualties [21]. Moreover, Cai et al. combined the YoLOv3 and MobileNetv1 models to detect fish, finding that this method improved the detection results [22].

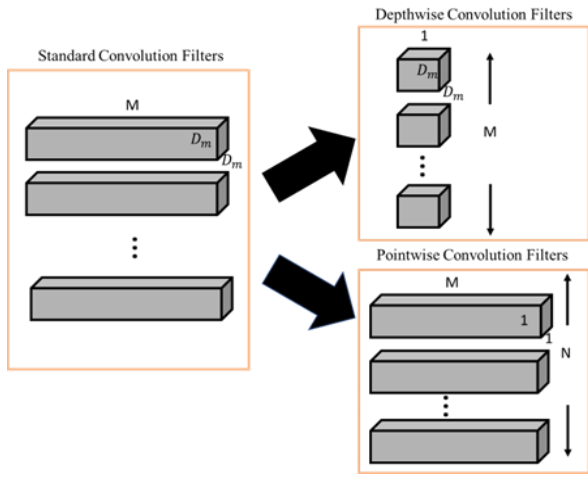


Fig. 2. Depthwise separable convolution Structure

2.3 EfficientNet

The EfficientNet model [23] compounds coefficients in the unified scaling of all dimensions simply and efficiently (Fig. 3). Atila et al. employed the EfficientNet model and migration learning techniques for model training and application in classifying plant leaf diseases [24]. Ho et al. combined the EfficientNet and long short-term memory models to automate the frequency of sow lactation [25]. Poudel and Lee presented automatic segmentation in medical images and employed EfficientNet for feature extraction and U-Net for fine-grained reconstruction [26]. Moreover, Yang et al. designed an EfficientNet model with a dual-attention mechanism to analyze fish feeding behavior [27]. Further, Marques et al. implemented a medical decision support system to automate the diagnosis of the coronavirus disease 2019 (COVID-19) using the EfficientNet model [28].

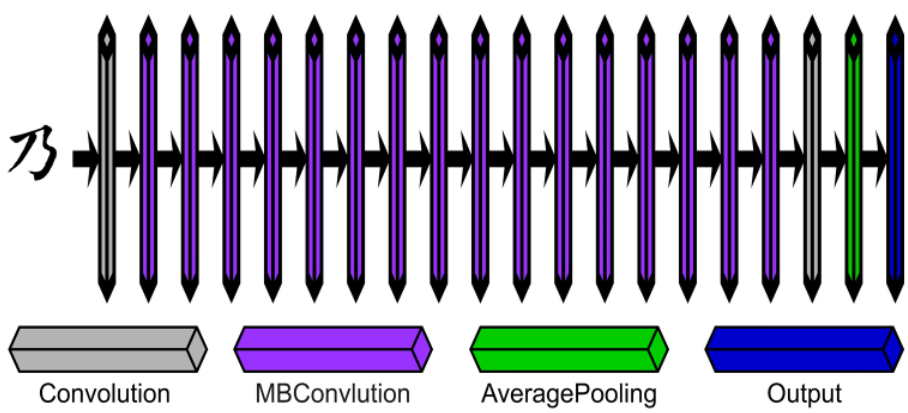


Fig. 3. EfficientNet architecture

2.4 Explainable AI

An XAI model can help users understand results created using machine or deep learning algorithms. For example, Samek et al. proposed several orientations regarding why AI requires explanation: verification of the soundness of a system, improvement of a system's algorithms, learning from a system, and compliance with legislation [29]. Bussmann et al. incorporated XAI into credit risk management, arguing that their approach groups risky or risk-free borrowers by financial characteristics [30]. Further, Dik-shit and Pradhan suggested explainable machine learning models for spatial drought prediction, employing Shapley additive explanations for result interpretability and finding the interpreted results consistent with the physical model interpretation [31]. In addition, Ma et al. designed an explainable learning framework for gait detection in patients with Parkinson's disease, concluding that excellent prediction results can be achieved using extreme gradient boosting or deep learning frameworks [32]. Moreover, Raza et al. presented a module to incorporate XAI to classification models, suggesting that it could facilitate faster and more reliable decision-making in clinical settings [33].

3. RESEARCH DESIGN AND PROCESS

The research design is structured into the following four sequential stages to ensure methodological clarity and reproducibility: data collection, data processing, model training and evaluation, and XAI analysis (Fig. 4). The purpose of each stage is described below.

(1) Data Collection

This stage constructs a representative, stylistically diverse dataset of Chinese calligraphy characters. A Python-based web crawler was developed to collect high-resolution calligraphy images systematically from publicly available online resources. This automated approach ensures sufficient data volume while preserving variations in brushwork, structure, and personal style across calligraphers. The collected dataset serves as the foundation for model training and comparative analysis.

(2) Data Processing

This stage standardizes image quality and enhances structural features to augment model learning efficiency. Image preprocessing encompasses image format conversion to ensure consistency, image thresholding to mitigate background noise and highlight stroke structures, and image resizing to unify input dimensions across deep learning models. These preprocessing steps diminish irrelevant visual variability and enable the models to focus on critical stroke patterns and spatial composition.

(3) Model Training and Performance Evaluation

This stage examines whether various CNN architectures recognize Chinese calligraphy characters and compares their learning capacities across stylistic variations. Multiple classical and lightweight CNN models were implemented, including LeNet-5, VGG-16, ResNet-101, InceptionV3, MobileNetV1, and EfficientNetB0. These models represent varying depths and architectural complexities, allowing a systematic comparison between traditional and modern CNN frameworks. Model performance was evaluated using

precision, recall, specificity, the F1-score, and accuracy to ensure a comprehensive assessment of the effectiveness of classifications rather than relying on a single metric.

(4) Image Hotspot Analysis Using Explainable Artificial Intelligence (XAI)

The final stage interprets model decision-making behavior and identifies the visual regions that most influence classification outcomes. The XAI techniques were applied to generate image hotspot visualizations, indicating which stroke areas and structural components each model focuses on during recognition. This stage bridges technical performance with educational interpretation by determining whether the models capture meaningful brushstroke characteristics that are relevant to the calligraphy instruction.

This four-stage research design ensures logical progression from data acquisition to interpretable model analysis, enhancing the methodological transparency and educational applicability. Table 1 presents the detailed specifications of the hardware and software configuration for the training and interpretation of the models using deep learning algorithms in the experimental environment: i7-9700k CPU, 64 GB RAM, RTX 2070 GPU, Windows 10, CUDA (v.11.0), Python (v.3.7.1), and TensorFlow (v.2.4.0).

Table 1. Specifications of the hardware and software

Type	Specification
Central processing unit	Intel® Core™ i7-9700K
Memory	64 GB
Operating system	Windows 10
Graphics processing unit	Nvidia GeForce RTX 2070
CUDA	11.0
Python	3.7.1
TensorFlow-GPU	2.4.0

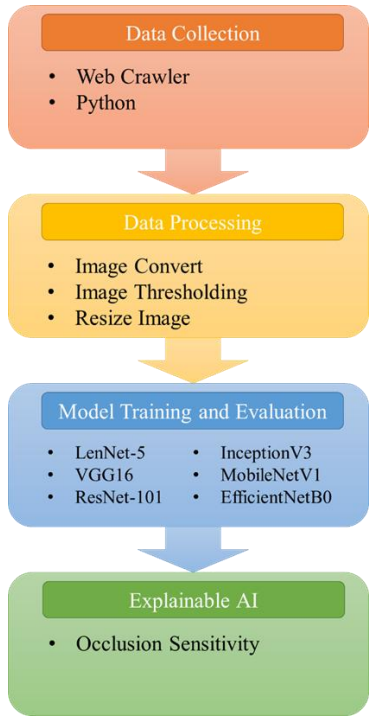


Fig. 4. Flowchart of this study

3.1 Data Collection

This study employs a Python program for data crawling to capture the regular calligraphy font images from the Handian calligraphy web page of four famous calligraphers: Gonquan Liu, Xun Ouyang, Mengfu Zhao, and Zhenqing Yan. The number of fonts captured in the regular calligraphic font images is 512 images for Gongquan, 459 images for Xun, 120 images for Mengfu, and 958 images for Zhenqing, as listed in Table 2.

Table 2. The number of font images

Calligrapher	Number
Gongquan Liu	512
Ouyang Xun	459
Mengfu Zhao	120
Zhenqing Yan	958

3.2 Data Processing

In the initial data processing, the captured calligraphic images are in graphics interchange format (GIF); therefore, they were converted into portable network graphics (PNG) files and then binarized. This study sets grayscale values higher than 127 to the maximum gray-scale value and those less than 127 to zero. Finally, this work cuts and shrinks the images of the calligraphic characters. The pixel value of the captured images is 378×416; thus,

this study crops the calligraphic character images to 378×378 (Fig. 5). Finally, this study converts the calligraphic character images to 256×256 as the input image size for this study model.

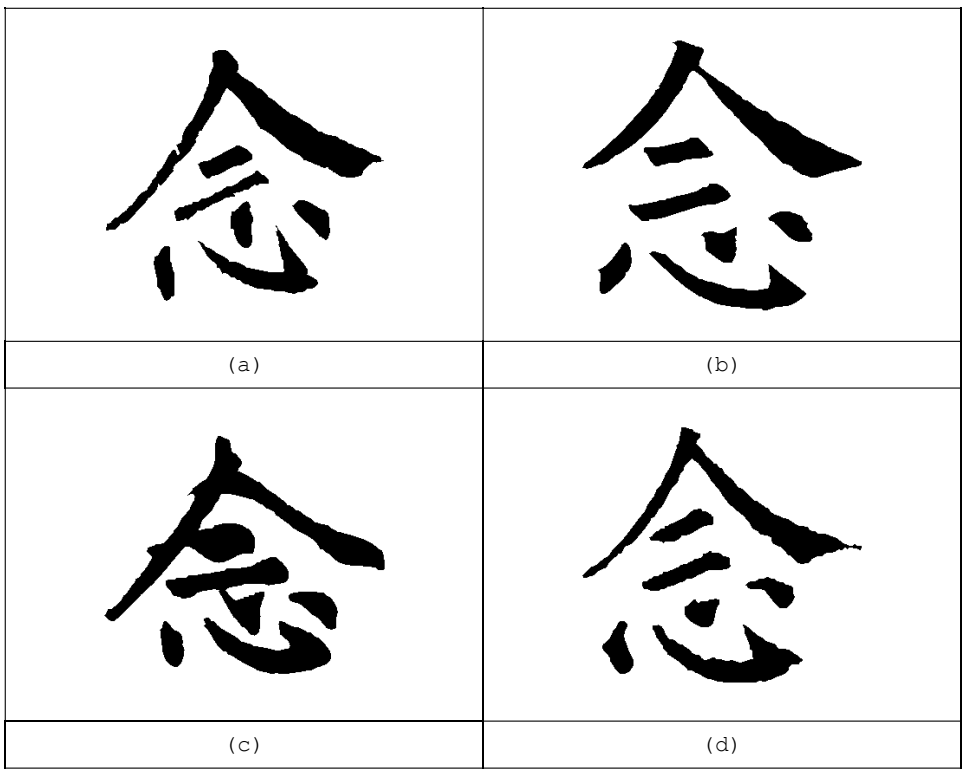


Fig. 5. Results of cropping images of 念 by (a) Liu, (b) Ouyang, (c) Zhao, and (d) Yan.

3.3 Model Training and Evaluation

This study compares the following six models in the training section: LeNet-5, VGG-16, ResNet-101, InceptionV3, MobileNetV1, and EfficientNetB0. The tf.keras module was employed for model generation, with the following parameter settings. The input shape was 256×256×1, and random initialization was employed for weighting. The Adam optimizer was employed for optimization (see Eqs. (1)), and the learning rate was 0.001. The number of training sessions was set to 50, and the number of training samples in a training session was 16. This study involves a classification problem; hence, the softmax activation function was used in the output. Prechelt (1998) stated that early stopping avoids overfitting during training and accelerates the learning process or can improve the generalizability [34]. Therefore, this work applies early stopping to monitor accuracy and prevent overfitting (Table 3). Finally, precision, recall, specificity, the F1-score, and accuracy were applied for model evaluation (see Eqs. (2) to (6)):

Table 3. Specifications of the hardware and software

Parameter Name	Parameter Value
Input shape	256×256×1
Weights	Random initialization
Optimizer	Adam
Learning rate	0.001
Output activation function	Softmax
Epochs	50
Batch size	16
Early stopping monitor	Accuracy
Early stopping patience	3

$$\theta_{t+1} = \theta_t - \alpha \frac{m_t}{\sqrt{v_t} + \varepsilon} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$Specif\ icity = \frac{TN}{TN + FP} \tag{4}$$

$$F1 - score = \frac{2 * Precision * Recall}{(Precision + Recall)} \tag{5}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

where α denotes the learning rate, ϵ ensures numerical stability, $\widehat{m_t}$ denotes the bias-corrected first moment estimate, $\widehat{v_t}$ denotes the bias-corrected second moment estimate, TP is a true positive, FP is a false positive, TN is a true negative, and FN is a false negative.

3.4 Explainable AI

This study employs the tf-explain framework to enhance the interpretability of the deep learning model applied for calligraphic character recognition. This work adopts the occlusion sensitivity method that Zeiler and Fergus [35] proposed to quantify pixel-level importance. Zeiler and Fergus introduced visualization techniques based on deconvolution and unpooling operations, which project hidden feature representations of a CNN back into the input space. These operations generate images with practical and inter-pretable meaning, enabling researchers to examine how internal feature activations correspond to spatial patterns in the original image. Building on this interpretability framework, occlusion sensitivity systematically masks localized regions of the input image and measures the change in prediction confidence. For each spatial location, a small region is occluded, and the variation in the predicted class probability is recorded. If the predicted probability decreases after occluding a re-gion, that region is considered to contribute positively to the model decision. The importance values are visualized using Viridis color maps, where higher intensity indicates a greater contribution to the classification outcome. The proposed model is trained under a multiclass classification framework, in which each calligrapher corresponds to one class in a shared global model. Under this setting, the generated XAI im-portance maps represent discriminative features learned via joint optimization across all calligraphers, rather than stylistic features derived in isolation.

The occlusion-based importance scores are computed with respect to the predicted class probability derived from the softmax output. Because the softmax probability is normalized across all classes, the importance value at each spatial location reflects how much that region contributes to distinguishing the predicted calligrapher from others in the dataset. Therefore, the resulting attribution is class-specific while being derived from a globally trained model. The reported high and low importance scores were computed relative to the predicted class in the multiclass setting, rather than independently in each calligrapher's image subset. The magnitude of the importance values reflects the region's contribution to interclass separability in the trained model. Thus, the heatmaps indicate spatial regions that most influence the model classification decision in a comparative context. These regions may correspond to the stroke curvature, structural balance, brush pressure distribution, or edge characteristics that differentiate one calligrapher from others. By emphasizing influential stroke structures and compositional elements, heatmaps offer a quantitative and visually interpretable means of analyzing stylistic distinctions between calligraphers (e.g., Liu Gongquan, Zhao Mengfu, Ouyang Xun, and Yan Zhenqing), while supporting model interpretability and an educational analysis in the global classification framework.

4. DISCUSSION OF RESULTS

Table 4 presents a comparative evaluation of multiple image recognition models, including LeNet-5, VGG-16, MobileNetV1, InceptionV3, ResNet-101, and EfficientNetB0. The results reveal substantial performance differences between the architectures. For example, LeNet-5 and VGG-16 did not meaningfully improve in accuracy during training; therefore, both models were stopped at five epochs, achieving only 46.75% accuracy. In addition, MobileNetV1 was trained for 16 epochs, but its performance was limited, reaching an accuracy of 51.00% with modest precision and recall values. In contrast,

InceptionV3, ResNet-101, and EfficientNetB0 demonstrated improved results, with EfficientNetB0 achieving the highest overall performance across all evaluation metrics. EfficientNetB0 achieved 98.34% in accuracy, 97.35% in precision, 98.80% in recall, 99.47% in specificity, and 98.05% in the F1-score. The consistently high values across these complementary metrics imply stable and balanced classification behavior rather than isolated metric inflation. The high recall and specificity indicate strong TP detection while maintaining minimal FP predictions. If overfitting were present, instabil-ity among metrics would typically be observed (e.g., a disproportionately high precision with a reduced recall or inflated accuracy with inconsistent F1-score values). However, the uniform strength across all indicators indicates reliable generalization in the evaluated dataset. From a training perspective, EfficientNetB0 was trained for 23 epochs, which is equivalent to the training for ResNet-101 at 23 epochs, and it required fewer epochs than InceptionV3 at 29. Despite similar training durations, EfficientNetB0 significantly outperformed ResNet-101, which achieved 88.04% accuracy and an F1-score of 75.24%.

The progressive performance enhancement from simpler architectures to more advanced models signifies that the gains are primarily attributable to the architectural de-sign and feature representation capacity rather than to extended training time alone. The superior performance of EfficientNetB0 can be explained by its compound scaling strategy, which balances network depth, width, and input resolution in a coordinated manner. This design enhances the ability of the model to extract fine-grained visual features (e.g., the stroke curvature, brush pressure variation, edge sharpness, and structural composition), all of which are critical in distinguishing subtle stylistic differences in calligraphy. Furthermore, EfficientNetB0 achieved strong representational capacity with efficient parameter utilization, reducing redundancy while maintaining expressive power. This balance is advantageous for moderately sized datasets because it facilitates generalizability without excessive model complexity. The shallow LeNet-5 architecture was initially developed for simple digit recognition tasks and lacks the depth necessary to capture complex spatial patterns in calligraphic strokes. Although deeper, VGG-16 contains numerous parameters and lacks mechanisms (e.g., residual connections or compound scaling), which may lower optimization efficiency and generalization when applied to a dataset of limited scale. In addition, MobileNetV1 prioritizes computational efficiency via depthwise separable convolutions, which may restrict feature richness for tasks re-quiring subtle stylistic discrimination. The InceptionV3 and ResNet-101 models improve performance via multiscale feature extraction and residual learning, respectively; however, they do not incorporate the balanced scaling principle that characterizes Efficient-Net, which may explain their comparatively poorer performance. Therefore, the observed differences in accuracy can be attributed to variations in architectural capacity, parameter efficiency, and suitability of the dataset size. Further, Effi-cientNetB0 demonstrates an effective balance between expressive power and generali-zability, achieving superior classification performance without evident signs of overfit-ting under the current experimental conditions. Future research should further validate these findings using cross-validation, larger datasets, or expanded calligrapher categories to examine performance stability under more complex classification scenarios.

Table 4. Comparison of model training results

Model Name	epoch	Precision	Recall	Specificity	F1-score	Accuracy
LeNet-5	5	25.00%	0%	0%	0%	46.75%
VGG16	5	25.00%	0%	0%	0%	46.75%
MobileNetV1	16	64.58%	68.48%	85.07%	53.72%	51.00%
InceptionV3	29	79.10%	79.62%	94.89%	73.42%	82.19%
ResNet-101	23	74.04%	90.85%	96.27%	75.24%	88.04%
EfficientNetB0	23	97.35%	98.80%	99.47%	98.05%	98.34%

This study employs XAI technology to analyze the calligraphic characteristics of the characters 念, 象, 刑, and 為 (illustrated in Figs. 6 to 9), representing the styles of Gonquan Liu, Ouyang Xun, Mengfu Zhao, and Zhenqing Yan, respectively. By applying occlusion-based attribution to the EfficientNetB0 model, the study identifies the spatial regions that most strongly shaped classification decisions, revealing discriminative stylistic features embedded in each calligrapher's work.

For Gongquan Liu, the importance maps indicate that the model strongly emphasizes the initial and terminal strokes of the characters. These regions often reflect Liu's distinctive angular brushwork, controlled stroke initiation, and sharply defined stroke endings. The concentration of importance at the beginning and end of strokes suggests that these structural features play a critical role in differentiating his style from others.

In the case of Ouyang Xun, the attribution maps display a more distributed pattern of importance across the entire character structure. Rather than focusing on isolated stroke segments, the model highlights crucial structural nodes throughout the composition, reflecting Ouyang's emphasis on balanced proportion, precise alignment, and structural stability. This approach suggests that global compositional coherence, rather than localized stroke variation, contributes significantly to stylistic identification.

For Mengfu Zhao, the heatmaps emphasize regions associated with stroke connections and transitional curves. The model assigns greater importance to areas where strokes merge or change direction, indicating sensitivity to Zhao's smooth, flowing transitions and rounded turning techniques. These findings align with the stylistic characteristics of Zhao's calligraphy, known for its fluid continuity and elegant modulation of brush movement.

The attribution results for Zhenqing Yan demonstrate a stronger concentration of importance on the left-side structural components of the characters. This pattern suggests that the model relies heavily on the spatial weight distribution and forceful left-side stroke construction that define Yan's bold and vigorous style. The emphasis on these structural regions indicates that asymmetry and dynamic stroke thickness contribute to distinguishing his work from that of other calligraphers.

The XAI visualizations demonstrate that EfficientNetB0 does not merely memorize superficial patterns but instead learns stylistically meaningful spatial features. The attribution maps offer insight into how structural and stroke-level characteristics contribute to model decisions, providing computational validation and pedagogical value for under-standing stylistic distinctions between major calligraphers.

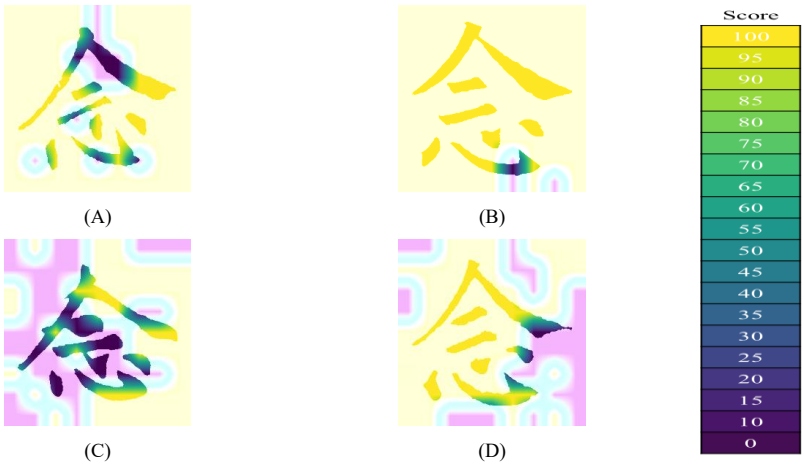


Fig. 6. EfficientNetB0: Key regions of calligrapher’s 念: (a) Liu, (b) Ouyang, (c) Zhao, and (d) Yan.

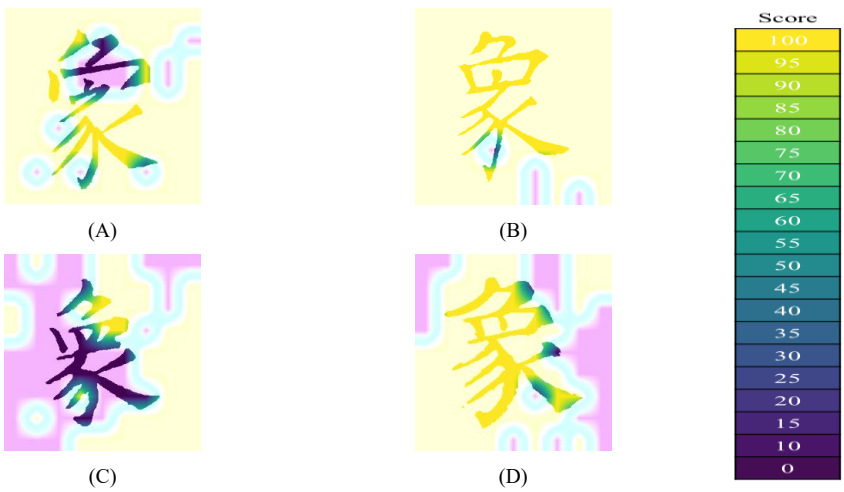


Fig. 7. EfficientNetB0: Key regions of calligrapher’s 象: (a) Liu, (b) Ouyang, (c) Zhao, and (d) Yan.

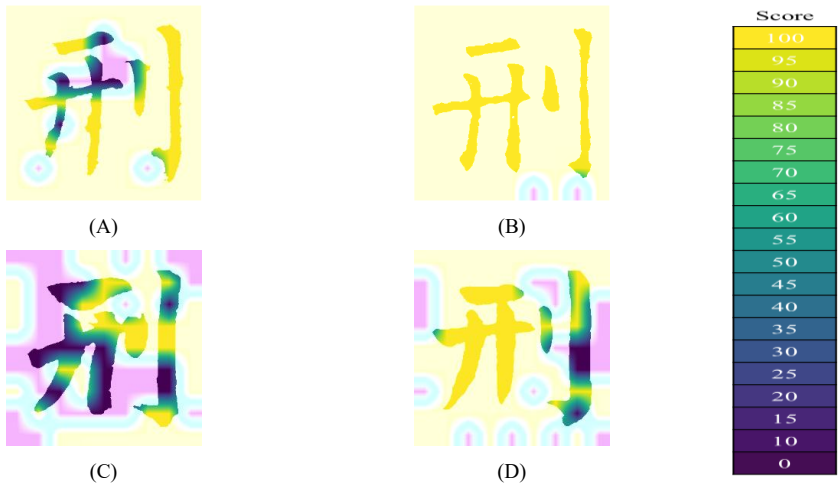


Fig. 8. EfficientNetB0: Key regions of calligrapher's 刑: (a) Liu, (b) Ouyang, (c) Zhao, and (d) Yan.

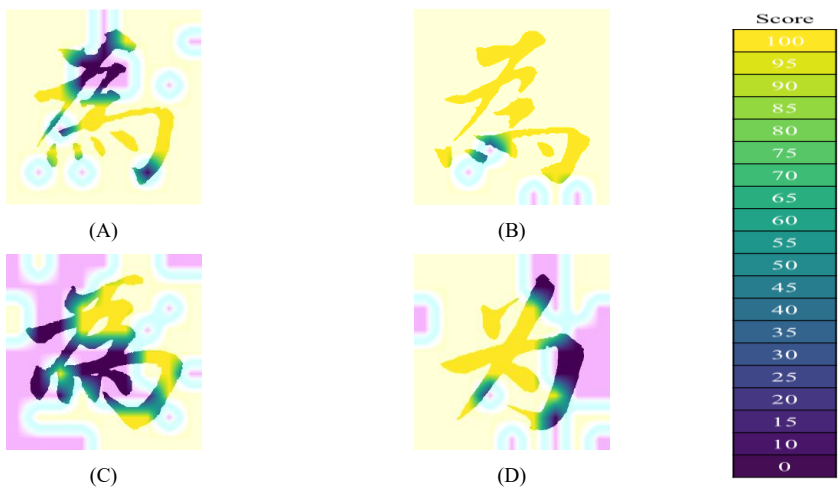


Fig. 9. EfficientNetB0: Key regions of calligrapher's 為: (a) Liu, (b) Ouyang, (c) Zhao, and (d) Yan.

Figure 10 presents the attribution results for characters written by Zhenqing Yan, demonstrating a consistent interpretability pattern. Across multiple characters, the heatmaps produced using EfficientNetB0 consistently highlight regions corresponding to brush strokes rather than background areas. High-importance regions cluster along stroke contours, turning points, and structural junctions, indicating that the model relies on meaningful calligraphic components for classification. This consistency suggests that the learned representations were stable across characters in Yan's stylistic system. The model repeatedly emphasizes variations in stroke thickness, angular transitions, and structural forceful elements, all of which are characteristic of Yan's bold and vigorous writing style.

The recurrence of these attention patterns across characters supports the reliability of the attribution mechanism and indicates that the heatmaps reflect structurally relevant features rather than random activations.



Fig. 10. EfficientNetB0: Key regions of calligraphic characters by Yan.

5.DISCUSSION

The model with six CNNs is ineffective in predicting book recognition using LeNet-5 and VGG-16. The two models ended early at the fifth epoch with a poor prediction accuracy of 46.75%. The MobileNetV1 model presented a good prediction effect compared with the LeNet-5 and VGG-16 models, but it only improved to 51.00%. The InceptionV3, ResNet-101, and EfficientNetB0 models exhibit better prediction in calligraphic font recognition, with 82.19%, 88.04%, and 98.34% prediction accuracy values. Among them, EfficientNetB0 achieved the best prediction results.

This study employs the EfficientNetB0 prediction results in the XAI module to explore the classification of the model in predicting where the critical areas of each calligrapher are correct. The model can determine the relative importance of each calligrapher's font, rather than classifying it based on the nontextual aspects.

In Chinese calligraphy, calligraphers can express their own unique style of writing of each character in the brush stroke. Calligraphers develop a style of writing different from others after learning from the spirit of the brushwork of their predecessors. Greater originality historically results in a higher reputation in calligraphy. The calligraphy of Xizhi Wang, the sage of calligraphy, was already famous in the Eastern Jin dynasty, and people of that era imitated his calligraphy to improve their own strokes. In the Tang dynasty and even later, almost every great calligrapher studied Wang's calligraphy. Famous calligraphers, such as Xun Ouyang (557–641) in the early Tang dynasty, Zhenqing Yan (709–785) and Gongquan Liu (778–865) in the late Tang dynasty, and Mengfu Zhao (1254–1322) in the Yuan dynasty, all studied Wang's calligraphy and imitated Wang's handwriting as if it were printed by a photocopier, but they still developed their own unique style of calligraphy, which is the reason for their high status in calligraphy history.

Next, this study employs interpretable AI technology to determine the calligraphy of

the calligraphers Liu Gongquan, Ouyang Qu, Zhao Mengfu, and Yan Zhenqing. The same Chinese characters, for example, “念,” “象,” “刑,” and “為,” are written in different calligraphers’ styles. The emphasis of Liu’s style is on the importance of the beginning and end of each stroke, which can also be observed in Liu’s calligraphy with the naked eye. Liu’s writing style focuses on the beginning of the strokes, where an obvious fold exists, making an extra twist in the straight lines, and heavy pressure occurs before closing.

The emphasis of the Ouyang style of writing is on learning the structure of the entire character. Each stroke appears thin and bony, yet contains spirit, and the layout of the main strokes should be extended outward, while the structure of the character should be tightened toward the center.

Zhao’s style of writing emphasizes the connection between strokes and the turning method. Zhao was a native of the Yuan dynasty after the Tang and Song dynasties. The Tang dynasty was dominated by the development of the regular script, and the Song dynasty by the development of the running script. The greatest difference between the running script and the regular script is that the speed of writing in the running script is faster. In terms of brushwork, Zhao was of a later generation that absorbed the brushwork of his predecessors to an extent; hence, fewer special features are exhibited in his brushwork.

The Yan style of writing emphasizes the left side of the character, requiring attention to the writing style. Yan’s calligraphy style exhibits good left-light and right-heavy layouts, and each character opens and closes in a flat and even manner.

The core of STEAM education covers cross-domain, hands-on, life application, problem solving, and five-sense learning. This study posits combining calligraphy and AI to realize cross-domain applications to apply in the field of education. This approach offers a new method of research primarily because every teacher and learner may learn and write in a value-based way when learning calligraphy and cannot learn writing fast enough. If this method can be applied to learn the calligraphy writing style, the critical points of the text produced by each calligrapher at the time of writing could enhance the student’s problem-solving and produce more meaningful value. In education, learning can be accelerated. Furthermore, outside the teaching field, when people have problems understanding calligraphy work, AI image recognition could provide viewers or appreci-ators with hints or analysis for appreciation.

Compared with previous studies focusing on calligraphy style classification and model efficiency (e.g., a lightweight network design and transfer learning optimization [40]), this study adopts a different research orientation. Earlier research has emphasized improving recognition accuracy and lowering computational cost, treating calligraphy analysis as a pattern recognition problem. In contrast, this study extends beyond performance metrics by integrating XAI to explore how CNNs capture the stroke structure and compositional features at the character level. Rather than focusing solely on style categorization, the proposed approach evaluates whether the model attention aligns with pedagogically meaningful brushwork principles. This shift from efficient classification to interpretable structural analysis underscores the educational value of AI-assisted calligraphy learning and offers a deeper understanding of how deep learning models represent artistic features.

The study data comprise publicly accessible calligraphic images available to users via online searches. Although this ensures accessibility and reproducibility, it may limit

the diversity and consistency of the image quality and historical coverage. In addition, this study focuses on regular script and includes four representative calligraphers in a controlled multiclass setting. Although this configuration allows a systematic evaluation of classification performance and interpretability, the scope is limited in stylistic breadth and calligrapher diversity. With respect to scalability, the EfficientNetB0 model performs stably under the current four-class framework. However, expanding the dataset to include additional calligraphers or writing styles could increase the complexity of the classification task and may introduce greater interclass similarity. As the number of categories increases, additional training data per class may be necessary to maintain discriminative performance. Exploring higher-capacity network variants, hierarchical classification strategies, or metric learning approaches could support robustness under expanded multiclass conditions.

Future research should extend this framework beyond the current four calligraphers to include additional historical and contemporary calligraphers from diverse dynasties and stylistic traditions. Expanding the dataset to incorporate a broader range of representative figures fosters a more comprehensive comparative analysis of stylistic variations and structural characteristics. In addition, this study focuses only on the regular script. Subsequent work should explore other Chinese script styles, including oracle, bronze, seal, clerical, cursive, and running scripts. By integrating a broader spectrum of calligraphers and script forms, the proposed approach can facilitate large-scale stylistic modeling and provide richer interpretability for academic research and educational applications, helping learners better understand stroke techniques, compositional structures, and stylistic evolution across Chinese calligraphy periods.

6. CONCLUSIONS

This study compares well-established image recognition algorithms for the classification of calligraphic characters and demonstrates that the EfficientNetB0 model performs best overall, reaching an accuracy of 98.34% in identifying calligraphers. The comparative evaluation highlights the effectiveness of compound scaling in EfficientNetB0, which balances the model depth, width, and resolution to capture fine-grained stylistic features (e.g., stroke thickness, curvature, structural balance, and brush dynamics). The high classification accuracy indicates that deep CNNs can learn discriminative stylistic representations from calligraphic images, even when subtle structural variations exist between calligraphers.

This achievement has practical implications. For educators and learners, the model can serve as an intelligent analytical tool to identify characteristic stroke structures and compositional patterns associated with various calligraphers. When combined with XAI visualization, the system enables users to observe which spatial regions most influence the prediction, supporting the more structured learning of stylistic techniques and artistic features. The proposed framework may also offer a foundation for automated calligrapher identification systems. Such systems could assist researchers in art history, digital humanities, and cultural preservation by supporting large-scale stylistic analysis, authenticity verification, and archival categorization. Furthermore, integrating high-performance classification and interpretable visualization creates opportunities for developing intelligent

tutoring systems and interactive platforms that enhance engagement in traditional calligraphy studies. The results suggest that deep learning models, particularly EfficientNetB0, offer a reliable and scalable approach to computational calligraphy analysis, bridging traditional artistic scholarship using modern AI techniques.

REFERENCES

1. Y. Xiao, W. Lei, L. Lu, X. Chang, X. Zheng, X. Chen, "CS-GAN: Cross-Structure Generative Adversarial Networks for Chinese calligraphy translation, " *Knowledge-Based Systems*, Vol. 229, pp. 107334.
2. G. Lin, Z. Guo, F. Chao, L. Yang, X. Chang, C. M. Lin, C. Shang, "Automatic stroke generation for style-oriented robotic Chinese calligraphy, " *Future Generation Computer Systems*, Vol. 119, pp. 20-30.
3. B. Zhao, J. Tao, M. Yang, Z. Tian, C. Fan, Y. Bai, "Deep imitator: Handwriting calligraphy imitation via deep attention networks, " *Pattern Recognition*, Vol. 104, pp. 107080.
4. R. Wu, C. Zhou, F. Chao, L. Yang, C. M. Lin, C. Shang, "Integration of an actor-critic model and generative adversarial networks for a Chinese calligraphy robot, " *Neurocomputing*, Vol. 388, pp. 12-23.
5. T. A. Zucker, J. Montroy, A. Master, M. Assel, C. McCallum, G. Yeomans-Maldonado, G., "Expectancy-value theory and preschool parental involvement in informal STEM learning, " *Journal of Applied Developmental Psychology*, Vol. 76, pp. 101320.
6. D. Mutambara, A. Bayaga, A., "Determinants of mobile learning acceptance for STEM education in rural areas, " *Computers & Education*, Vol. 160, pp. 104010.
7. J. Morris, E. Slater, M. T. Fitzgerald, G. W. Lummis, E. van Etten, "Using local rural knowledge to enhance STEM learning for gifted and talented students in Australia, " *Research in Science Education*, Vol. 51, No. 1, pp. 61-79.
8. L. Sun, L. Hu, W. Yang, D. Zhou, X. Wang, "STEM learning attitude predicts computational thinking skills among primary school students, " *Journal of Computer Assisted Learning*, Vol. 37, No. 2, pp. 346-358.
9. J. Zheng, W. Xing, G. Zhu, G. Chen, H. Zhao, C. Xie, C., "Profiling self-regulation behaviors in STEM learning of engineering design, " *Computers & Education*, Vol. 143, pp. 103669.
10. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, "Rethinking the inception architecture for computer vision, " in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818-2826.
11. R. Castro-Zunti, K. J. Chae, Y. Choi, G. Y. Jin, S. B. Ko, "Assessing the speed-accuracy trade-offs of popular convolutional neural networks for single-crop rib fracture classification, " *Computerized Medical Imaging and Graphics*, Vol. 91, pp. 101937.
12. F. S. Ishengoma, I. A. Rai, R. N. Said, "Identification of maize leaves infected by fall armyworms using UAV-based imagery and convolutional neural networks, " *Computers and Electronics in Agriculture*, Vol. 184, pp. 106124.
13. Z. Liu, C. Yang, J. Huang, S. Liu, Y. Zhuo, X. Lu, "Deep learning framework based on integration of S-Mask R-CNN and Inception-v3 for ultrasound image-aided diagnosis of prostate cancer, " *Future Generation Computer Systems*, Vol. 114, pp. 358-367.
14. I. D. Apostolopoulos, D. I. Apostolopoulos, T. I. Spyridonidis, N. D. Papathanasiou, G. S. Panayiotakis, "Multi-input deep learning approach for Cardiovascular Disease diagnosis using Myocardial Perfusion Imaging and clinical data, " *Physica Medica*, Vol. 84, pp. 168-177.
15. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications, " *arXiv preprint*, arXiv:1704.04861,
16. Y. W. Hong, J. S. Leu, M. Faisal, S. W. Prakosa, "Analysis of Model Compression Using Knowledge Distillation, " *IEEE Access*, Vol. 10, pp. 85095-85105.

17. J. Chen, D. Zhang, M. Suzauddola, "A. Zeb, Identifying crop diseases using attention embedded MobileNet-V2 model, " *Applied Soft Computing*, Vol. 113, pp. 107901.
18. G. N. Elwirehardja, J. S. Prayoga, "Oil palm fresh fruit bunch ripeness classification on mobile devices using deep learning approaches, " *Computers and Electronics in Agriculture*, Vol. 188, pp. 106359.
19. E. Goceri, "Diagnosis of skin diseases in the era of deep learning and mobile technology, " *Computers in Biology and Medicine*, Vol. 134, pp. 104458.
20. Y. Nan, J. Ju, Q. Hua, H. Zhang, B. Wang, "A-MobileNet: An approach of facial expression recognition, " *Alexandria Engineering Journal*, Vol. 61, No. 6, pp. 4435-4444, June, 2022.
21. H. Wang, F. Lu, X. Tong, X. Gao, L. Wang, Z. Liao, "A model for detecting safety hazards in key electrical sites based on hybrid attention mechanisms and lightweight Mobilenet, " *Energy Reports*, Vol. 7, pp. 716-724, November, 2021.
22. K. Cai, X. Miao, W. Wang, H. Pang, Y. Liu, J. Song, "A modified YOLOv3 model for fish detection based on MobileNetv1 as backbone, " *Aquacultural Engineering*, Vol. 91, pp. 102117, November, 2020.
23. M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks, " in *International Conference on Machine Learning*, 2019, pp. 6105-6114.
24. Ü. Atila, M. Uçar, K. Akyol, E. Uçar, "Plant leaf disease classification using EfficientNet deep learning model, " *Ecological Informatics*, Vol. 61, pp. 101182.
25. K. Y. Ho, Y. J. Tsai, Y. F. Kuo, "Automatic monitoring of lactation frequency of sows and movement quantification of newborn piglets in farrowing houses using convolutional neural networks, " *Computers and Electronics in Agriculture*, Vol. 189, pp. 106376.
26. S. Poudel, S. W. Lee, "Deep multi-scale attentional features for medical image segmentation, " *Applied Soft Computing*, Vol. 109, pp. 107445.
27. L. Yang, H. Yu, Y. Cheng, S. Mei, Y. Duan, D. Li, Y. Chen, "A dual attention network based on EfficientNet-B2 for short-term fish school feeding behavior analysis in aquaculture, " *Computers and Electronics in Agriculture*, Vol. 187, pp. 106316.
28. G. Marques, D. Agarwal, I. de la Torre Díez, "Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network, " *Applied Soft Computing*, Vol. 96, pp. 106691.
29. W. Samek, T. Wiegand, K. R. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models , " *arXiv preprint*, arXiv:1708.08296.
30. N. Bussmann, P. Giudici, D. Marinelli, J. Papenbrock, J., "Explainable machine learning in credit risk management, " *Computational Economics*, Vol. 57, No. 1, pp. 203-216, January, 2021.
31. A. Dikshit, B. Pradhan, "Interpretable and explainable AI (XAI) model for spatial drought prediction, " *Science of the Total Environment*, Vol. 801, pp. 149797.
32. Y. W. Ma, J. L. Chen, Y. J. Chen, Y. H. Lai, "Explainable deep learning architecture for early diagnosis of Parkinson's disease, " *Soft Computing*, Vol. 27, No. 5, pp. 2729-2738.
33. A. Raza, K. P. Tran, L. Koehl, S. Li, "Designing ECG monitoring healthcare system with federated transfer learning and explainable AI, " *Knowledge-Based Systems*, Vol. 236, pp. 107763.
34. L. Prechelt, "Early stopping - but when?, " in *Neural Networks: Tricks of the trade*, Springer, 1998.
35. M. D. Zeiler, R. Fergus, "Visualizing and understanding convolutional networks, " *European Conference on Computer Vision*, Zurich, Switzerland, 2014, pp. 818-833.
36. Jiang, Z., Yu, J., Jin, Y., Ginn, A., Chen, J., & Sun, G, "When artificial intelligence comes to the Chinese calligraphic landscape: The coming transformation, " *Geography Compass*, 17(1), e12670.
37. Du, Y. (2024), "The innovative path of calligraphy education in universities in the era of artificial intelligence, " *Revista Ibérica de Sistemas e Tecnologias de Informação*, (E67), 560-569.
38. Lin, T. T., She, J., Wang, Y. A., & Zhang, K, "Future ink: The collision of AI and Chinese calligraphy, " *ACM Journal on Computing and Cultural Heritage*, 18(1), 1-17.

39. Valeri, F., Nilsson, P., & Cederqvist, A. M, "Exploring students' experience of ChatGPT in STEM education, " *Computers and Education: Artificial Intelligence*, 8, 100360.
40. Zhu, Y., & Zhu, Y, "PagodaNet: Light Weight AI Model for Chinese Calligraphy Styles Recognition, " *In 2024 IEEE 4th International Conference on Software Engineering and Artificial Intelligence (SEAI)*, 2024, pp. 22-27.



Wen-Ian Chen (陳文妍) graduated with a Master's and Doctoral degree in Chinese Literature from National Chung Hsing University in Taiwan. Her research specialties include classical Chinese literature, Chinese Etymology, and Chinese calligraphy.



Ying-Lin Hsu (許英麟) received his Ph.D. degree in Statistics from the National Chiao Tung University, Taiwan in 2002. He is currently a professor with the Department of Applied Mathematics and the Institute of Statistics at National Chung Hsing University, Taichung, Taiwan. His major research interests include statistical applications in industrial engineering, finance and bioinformatics. In recent years, the main research has been on using statistics combined with AI to apply to smart medical care.



Jia-Lang Xu (許家朗) received the Ph.D. degree in Doctoral Program in Data Science and Industrial Analytics from the National Chung Hsing University, Taiwan in 2022. He is currently an assistant professor with the Department of Applied Statistics, National Taichung University of Science and Technology, Taichung, Taiwan. His current research interests include smart healthcare, deep learning, data mining, machine learning, and multivariate analysis. He has served as Editorial Board on Scientific Reports and PLOS Digital Health.Associate Editor on International Journal of Big Data and Analytics in Healthcare and International Journal of Biomedical Engineering and Clinical

Science, and Editor Review Board on Health Informatics Journal and International Journal of Public and Private Perspectives on Healthcare, Culture, and the Environment.