Asia Pacific Science Press

Journal of Advances in Engineering and Technology
ISSN(O):3005-7507
ISSN(P):3005-7493
Vol. 3 No.2 (2026)

Design and Implementation of a Cross-Cultural Pattern Classification System Based on ResNet18

Alina Andreeva[1], Zhenning Li[2], Wenjie Tang[2], Chaofan Yue[3], Xinlei Cao[4], Xiang Zhang[3], Xiangyu Li[5]*

1. Faculty of Information Technology and Software Engineering, The Bonch-Bruevich St Petersburg State University of Telecommunications, Saint-Petersburg, 193232, Russia
2. Faculty of Engineering and Quality Management, D.F. Ustinov Baltic State Technical University "VOENMEH", Saint Petersburg, 190005, Russia
3. Faculty of Control Systems and Robotics, ITMO University, St. Petersburg, 197101, Russia
4. College of Furnishings and Industrial Design, Nanjing Forestry University, Nanjing, Jiangsu, 210037, China
5. Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China

Corresponding author: Xiangyu Li, xiangyuli@sjtu.edu.cn

Abstract: The automatic recognition of traditional cultural patterns from China, Japan, and Korea poses a challenging cross-cultural image classification problem, compounded by high intra-class stylistic diversity, inter-cultural iconographic overlap, and the scarcity of annotated training data. We propose a lightweight transfer learning framework in which a ResNet18 backbone pre-trained on ImageNet is fully frozen and a compact three-class fully connected head comprising only 2,565 trainable parameters is appended and optimized for the target task, together with a stochastic data augmentation pipeline combing random resized cropping, horizontal flipping, rotation, and Colour jitter. A purposebuilt benchmark of 750 balanced images spanning Chinese, Japanese, and Korean traditional patterns is constructed and partitioned following a 70:15:15 ratio. Comprehensive evaluation against five competing architectures (VGG16, MobileNetV2, EfficientNet-B0, DenseNet121, and a from-scratch SimpleCNN baseline) under identical training conditions demonstrates that the proposed method achieves 87.61% test accuracy and an F1-score of 0.8746, outperforming SimpleCNN by 8.85 percentage points while requiring over four orders of magnitude fewer trainable parameters than VGG16. Controlled ablation experiments reveal that data augmentation contributes the largest single performance gain of 22.00 percentage points, followed by ImageNet pre-training at 11.61 percentage points. Per-class analysis shows that Korean patterns are most reliably discriminated (F1 = 0.900), Japanese patterns achieve the highest recall (0.949), and Chinese patterns pose the greatest challenge (recall = 0.765) owing to their broad stylistic range and historical iconographic diffusion into neighboring traditions. These findings establish that a frozen-backbone transfer learning strategy with targeted augmentation provides an efficient and practically deployable solution for cross-cultural heritage pattern recognition under realistic computational constraints.

Keywords: Transfer Learning; ResNet18; Cross-cultural Pattern Classification; Cultural Heritage Digitization; Convolutional Neural Network; Data Augmentation; Image Classification

Published: May 13, 2026

DOI: https://doi.org/10.62177/jaet.v3i2.1367

1. Introduction

Traditional patterns encode rich historical information and aesthetic values that have been transmitted across generations. Within the East Asian cultural sphere, China, Japan, and Korea have independently developed distinctive pattern systems: Chinese patterns, exemplified by dragon-and-phoenix motifs and cloud scrolls, embody the ideal of harmony between humanity and nature [1]; Japanese patterns, including ukiyo-e compositions and family crests (kamon), priorities minimalist elegance [2]; and Korean patterns, characterized by dancheong polychromy and lotus motifs, synthesise Buddhist aesthetics with indigenous sensibilities [3]. Despite their cultural significance, accelerating modernisation has placed these traditions at risk, rendering automated recognition an urgent priority [4, 5].

Convolutional Neural Networks (CNNs) have been successfully applied to artwork classification, historical document recognition, and style identification [6-8]. However, the scarcity of annotated heritage data poses a fundamental challenge: deep learning models require large labelled datasets, yet curating specialised cultural imagery is labourintensive and costly [9, 10]. Transfer learning addresses this bottleneck by leveraging representations pre-trained on large-scale datasets [11, 12]. Prior work has individually advanced single-culture pattern recognition (e.g. bronze decorations [13], ethnic garments [4], Chinese painting styles [6]) and lightweight CNN design [14, 15], but no study has systematically addressed cross-cultural discrimination of Chinese, Japanese, and Korean patterns, and no dedicated benchmark dataset exists for this task.

To address these gaps, this paper makes the following contributions:

1. Benchmark dataset. We construct the first publicly available CJK traditional pattern dataset, comprising 750 curated images equally distributed across three cultural categories.

2. Lightweight transfer-learning classifier. We propose a ResNet18-based [16] frozenbackbone strategy yielding only 2,565 trainable parameters while achieving 87.61% test accuracy and an F1-score of 0.8746, representing a parameter reduction of over four orders of magnitude compared with VGG16 [17].

3. Systematic multi-model comparison. We benchmark six architectures under identical training conditions (Adam optimiser [18], batch size 32, 10 epochs), providing the first empirical accuracy-efficiency analysis on cross-cultural pattern recognition.

4. Component ablation. We quantify the individual contributions of pre-trained weights [12] and data augmentation [9], revealing that augmentation alone accounts for a 22.00 pp accuracy gain, which is the single largest performance factor in this low-data setting.

2. Methods

2.1 Dataset and Overall Framework

Let $D = \{(x_i, y_i)\}_{i=1}^{N}$ be the complete dataset (N = 750), where $x_i \in R^{H\times W\times 3}$ is the *i*-th RGB image and $y_i \in 0,1,2$ denotes the Chinese, Japanese, or Korean class label. The dataset is partitioned into disjoint training, validation, and test subsets following a 70:15:15 ratio:

$$D = D_{tr} \cup D_{val} \cup D_{te}, \quad |D_{tr}| = 525, \quad |D_{val}| = 112, \quad |D_{te}| = 113. \tag{1}$$

Each cultural category contributes exactly 250 images (Table 1), yielding a perfectly balanced distribution. The predicted class label is

$$\hat{y}_i = \arg\max_{k\in\{0,1,2\}} f_\theta(x_i)_k\,, \tag{2}$$

where f_θ comprises a frozen convolutional backbone backbone $f_{backbone}$: $R^{224\times224\times3} \to R^{512}$ and a lightweight classification head f_{head}: $R^{512} \to R^{3}$. The complete pipeline is illustrated in Figure 1.

Table 1: Dataset Statistics

Category	Number of Images	Percentage
Chinese	250	33.3%
Japanese	250	33.3%
Korean	250	33.3%
Total	750	100%

Figure 1: Overall framework of the proposed CJK pattern classification system

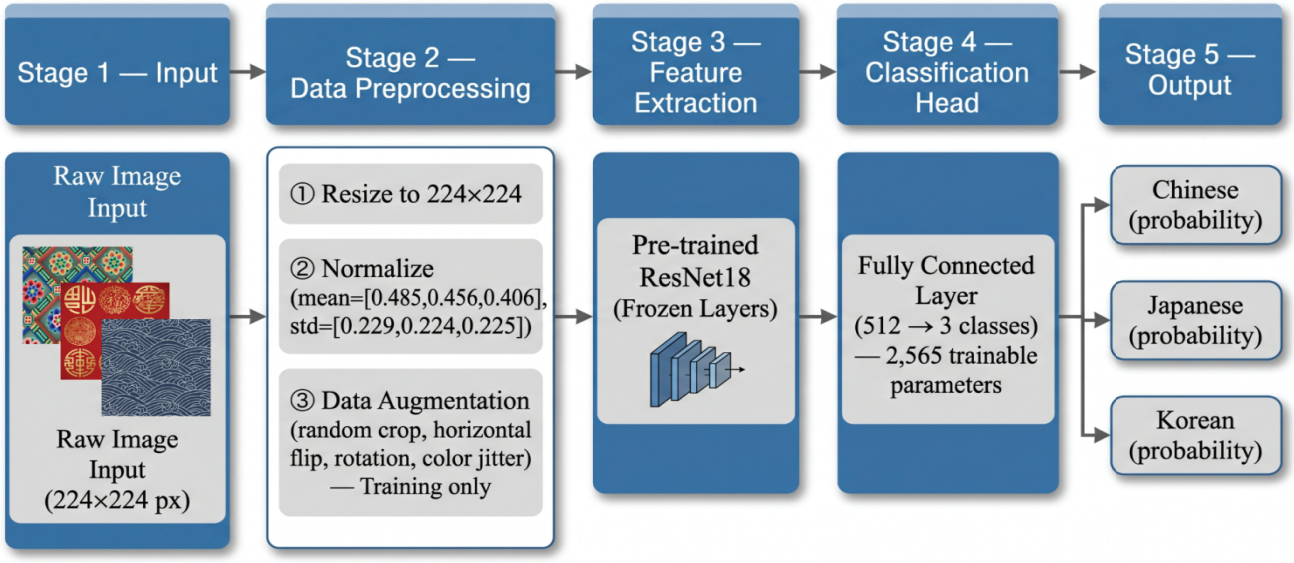


Figure 1 is a data preprocessing (resize, normalize, augment), feature extraction via a frozen ResNet18 backbone pre-trained on ImageNet, and a lightweight fully connected head with only 2,565 trainable parameters.

2.2 Data Preprocessing and Augmentation

All images are resized to 224×224 pixels to match the ResNet18 input requirement [12, 16]. Channel-wise normalisation uses ImageNet statistics:

$$\widehat{x^{(c)}} = \frac{x^{(c)} - \mu^{(c)}}{\sigma^{(c)}}, \quad c \in R, G, B, \tag{3}$$

where $\mu = (0.485, 0.456, 0.406)$ and $\sigma = (0.229, 0.224, 0.225)$. This preserves the validity of frozen backbone's learned batch normalisation statistics [19].

To address the limited dataset size, a stochastic augmentation pipeline A is applied exclusively to training images:

$$\widetilde{x_i} = A(x_i) = T_{color} \circ T_{flip} \circ T_{rotate} \circ T_{crop}(x_i), \tag{4}$$

Specifically:

1. Random resized cropping (T_{crop}) samples a scale factor $s \in [0.8, 1.0]$, simulating framing variation across museum digitisation protocols [5];

2. Random rotation (T_{rotate}) samples $\alpha \in [-15°, 15°]$ to introduce misalignment invariance without distorting canonical pattern orientations [3];

3. Horizontal flipping (T_{flip}) is applied with probability $p = 0.5$, doubling the effective training set without label noise [9]; Colour jitter (T_{color}) perturbs brightness, contrast, saturation, and hue by δ = 0.2, compensating for illumination and white-balance variability [6].

Validation and test images receive only resizing and normalisation (Eq. 3).

2.3 Model Architecture: ResNet18 with Frozen Backbone

ResNet18 [16] is selected as the feature extraction backbone for its favourable balance between depth and parameter count (≈11M backbone parameters), its residual connections that mitigate the vanishing-gradient problem, and its wide validation in transfer learning studies [11, 20].

Each residual block computes

$$h^{(l+1)} = F\left(h^{(l)}, W^{(l)}\right) + h^{(l)}, \tag{4}$$

where F represents two successive 3×3 convolutions each followed by Batch Normalisation (BN) [19] and ReLU. When spatial or channel dimensions change, a 1×1 projection shortcut W_S is used. Four sequential stages produce feature maps with 64,128,256,521 channels, progressively capturing low-level textures through to high-level semantic representations.

After the final stage, Global Average Pooling (GAP) [21] reduces the 7×7×512 feature map to a 512-dimensional descriptor:

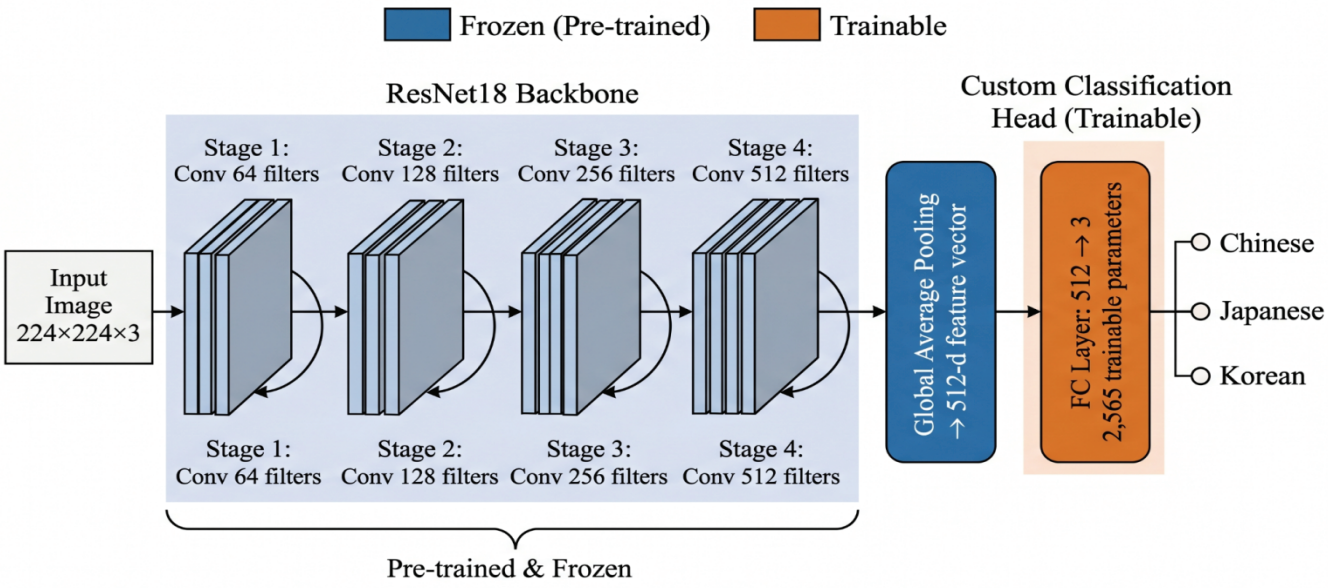
$$z = \frac{1}{H'W'} \sum_{h=1}^{H'} \sum_{w=1}^{W'} F_{h,w}, \quad z \in \mathbb{R}^{512}. \tag{6}$$

The original 1,000-class ImageNet head is replaced by a single fully connected layer:

$$\hat{p}_i = Softmax(W_c z_i + b_c), \tag{7}$$

where $W_c \in R^{3\times512}$ and $b_c \in R^3$ yield 512×3+3=2,565 trainable parameters. The architecture is illustrated in Figure 2.

Figure2: Architecture of the modified ResNet18 classifier



In Figure 2 the convolutional backbone (blue) is frozen during CJK training; only the fully connected head (orange, 2,565 parameters) is updated via backpropagation.

All backbone parameters are frozen during training. The limited dataset size (N = 750, 525 training images) renders full fine-tuning of 11M backbone parameters highly prone to overfitting [9, 11]. Prior work confirms that ImageNet features transfer constructively to medical imaging [20], agricultural vision [22], and artistic classification [6], even across substantial domain gaps. Freezing also reduces training time by approximately an order of magnitude.

2.4 Training Strategy and Performance Metrics

The model minimises the categorical cross-entropy loss:

$$\mathcal{L}_C = -\frac{1}{|\mathcal{D}_t|} \sum_{i\in\mathcal{D}_t} \sum_{k=0}^{2} \mathbb{1}[y_i = k] \log \hat{p}_{i,k} \tag{8}$$

Parameters are updated using the Adam optimiser [18], with η = 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$,. Training runs for T = 10 epochs with batch size B = 32; the checkpoint maximising validation accuracy is retained:

$$\theta^* \arg\max_{\theta_t, t\le T} \mathrm{Acc}(f_{\theta_t}, \mathcal{D}_{vl}). \tag{9}$$

Performance is assessed using four complementary metrics. Overall accuracy is

$$\mathrm{Acc} = \frac{1}{|\mathcal{D}_t|} \sum_{i\in\mathcal{D}_t} \mathbb{1}[\hat{y}_i = y_i] \tag{10}$$

Per-class precision and recall are

$$P_k = \frac{TP_k}{TP_k + FP_k}, \quad R_k = \frac{TP_k}{TP_k + FN_k}, \tag{11}$$

and the macro-averaged F1 score is

$$F_1 = \frac{1}{3} \sum_{k=0}^{2} \frac{2P_k R_k}{P_k + R_k}. \tag{12}$$

Because the dataset is perfectly balanced (250 images per class), accuracy is a valid primary metric. All experiments are implemented in PyTorch 1.12.1 on a standard CPU platform, with all random seeds fixed at 42 for reproducibility [8, 22]. The complete training procedure is specified in Algorithm 1.

Algorithm 1. Transfer Learning-Based CJK Pattern Classification

Input: Dataset D, pretrained ResNet18 weights $\boldsymbol{\theta}_{pre}$, η = 0.001, B = 32, T = 10

Output: Optimal parameters $\boldsymbol{\theta}^*$

1: Split D per Eq. (1); apply A (Eq. 4) to D_{tr}; normalize all splits (Eq. 3)
2: Load θ_{pre}; freeze backbone; replace head with (W_c, b_c) (2,565 parameters); initialize Adam
3: $best\_acc \leftarrow 0$, $\theta^* \leftarrow \varnothing$
4: **for** t = 1 **to** T **do**
5: **for** each mini-batch $\subset \boldsymbol{D}_{tr}$ **do**
6: Forward pass via frozen backbone and Eq. (7)
7: Compute $\boldsymbol{L}_{CE}$(Eq. 8); backpropagate through head only; update ($\boldsymbol{W}_c$, $\boldsymbol{b}_c$)
8: **end for**
9: Evaluate val_$\boldsymbol{acc}_t$ (Eq. 10)
10: **if** val_$\boldsymbol{acc}_t$ > best_acc **then**
11: $best\_acc \leftarrow val\_acc_t$; $\theta^* \leftarrow \theta t$
12: **end if**
13: **end for**
14: Evaluate $\boldsymbol{P}_k$, $\boldsymbol{R}_k$, $\boldsymbol{F}_1$ on $\boldsymbol{D}_{te}$ (Eqs. 11-12)
15: **return** $\boldsymbol{\theta}^*$, evaluation metrics

3. Experimental Results

3.1 Model Comparison

Six architectures are benchmarked under strictly identical configurations to ensure a fair comparison.

Table 2 reports the full set of performance metrics alongside the number of trainable parameters for each model.

Table2: Model Comparison Results

Model	Params	Val Acc (%)	Test Acc (%)	Precision	Recall	F1 Score
ResNet18 (Ours)	2,565	90.18	87.61	0.8750	0.8761	0.8746
VGG16	119,566,341	93.75	92.92	0.9362	0.9292	0.9297
MobileNetV2	3,843	87.50	91.15	0.9207	0.9115	0.9133
EfficientNet-B0	3,843	84.82	88.50	0.8867	0.8850	0.8845
DenseNet121	3,075	91.96	92.92	0.9292	0.9292	0.9289
SimpleCNN	51,476,549	70.54	78.76	0.7927	0.7876	0.7848

Several key observations emerge. VGG16 and DenseNet121 jointly achieve the highest test accuracy of 92.92%, yet VGG16 requires ≈46,000× more trainable parameters than DenseNet121 (3,075), illustrating that dense connectivity achieves equivalent accuracy at dramatically lower cost. The proposed ResNet18 achieves 87.61% test accuracy with only 2,565 trainable parameters, the smallest footprint across all models, requiring 46,625× fewer parameters than VGG16 while closing only 5.31 percentage points of the accuracy gap. SimpleCNN, despite 51.5M parameters, attains only 78.76%, confirming that raw parameter count is not a reliable proxy for performance under scarce labelled data. Learning dynamics for all six models are shown in Figure 3: all pretrained models converge within three epochs, while SimpleCNN exhibits slow and oscillatory convergence throughout. The parameter-accuracy trade-off is visualised in Figure 4.

Figure 3: Training loss, training accuracy, validation loss, and validation accuracy curves for all six models over 10 epochs.

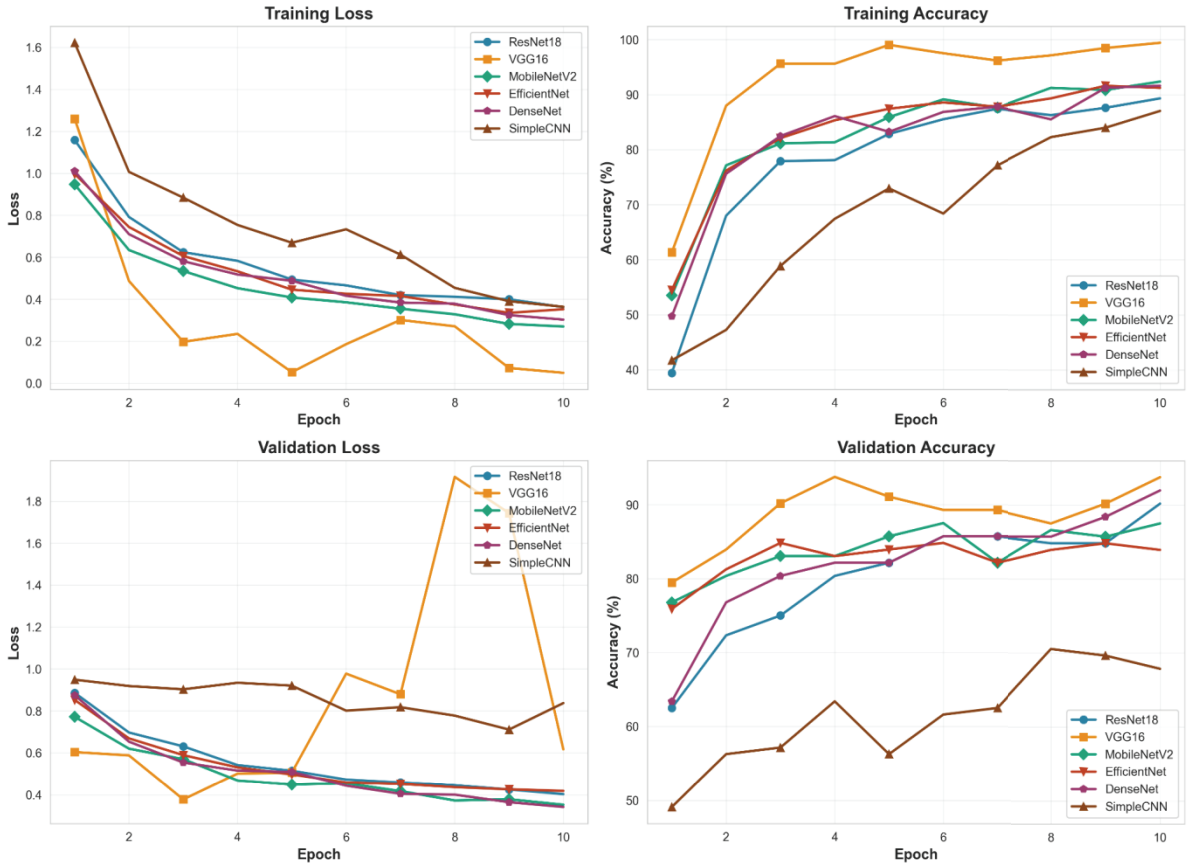
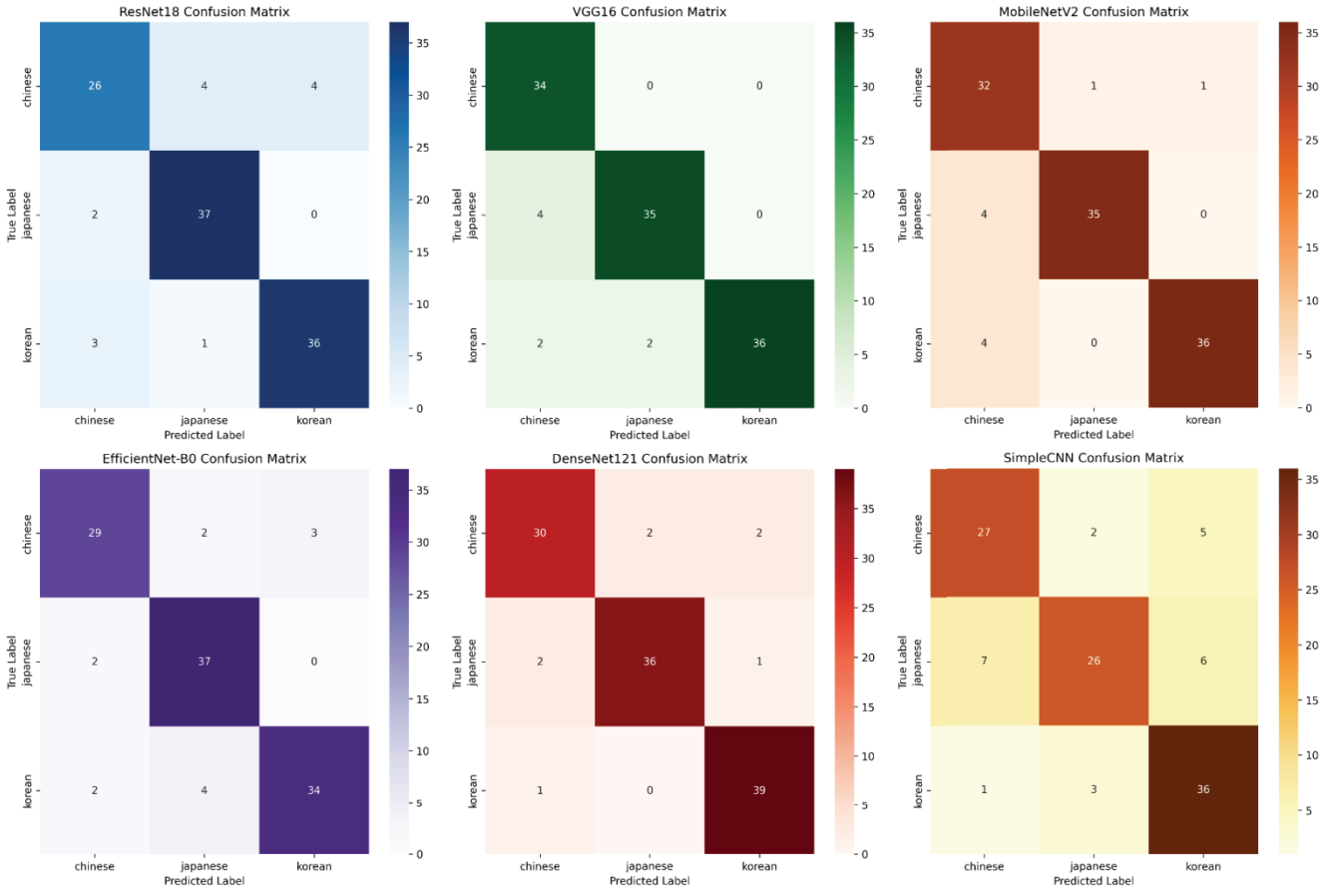


Figure 4: Test accuracy (left) and trainable parameters count on a logarithmic scale (right) for all six models



3.2 Per-Class Performance Analysis

Table 3 disaggregates the precision, recall, and F1 score of the proposed ResNet18 model across the three cultural categories.

Table3: Per-Class Performance Metrics for ResNet18

Category	Precision	Recall	F1 Score
Chinese	0.8387	0.7647	0.8000
Japanese	0.8810	0.9487	0.9136
Korean	0.9000	0.9000	0.9000
Macro avg	0.8732	0.8711	0.8712

Korean patterns achieve the most balanced per-class performance (Precision = Recall = F1 = 0.900), attributable to their distinctive dancheong colouring and symmetric geometric lattice structures that are less common in the other two traditions. Japanese patterns attain the highest recall (0.949), consistent with the strong intra-class consistency of canonical mon crests and ukiyo-e-derived motifs [6]. Chinese patterns yield the lowest recall (0.765) and F1-score (0.800): confusion matrices (Figure 5) reveal that the dominant error mode is mutual confusion with Korean patterns, reflecting the historical diffusion of Chinese dragon and cloud scroll motifs into Korean decorative traditions during the Goryeo and Joseon periods [3, 13]. The stylistic breadth of Chinese patterns spanning Han embroidery, Tang ceramics, Ming porcelain, and Qing lacquerware creates high intra-class variance that challenges the compact classifier. A radar chart (Figure 6) summarises these per-class asymmetries.

Figure 5: Confusion matrices for all six models on the test set. ResNet 18 concentrates errors in the Chinese-Korean pair

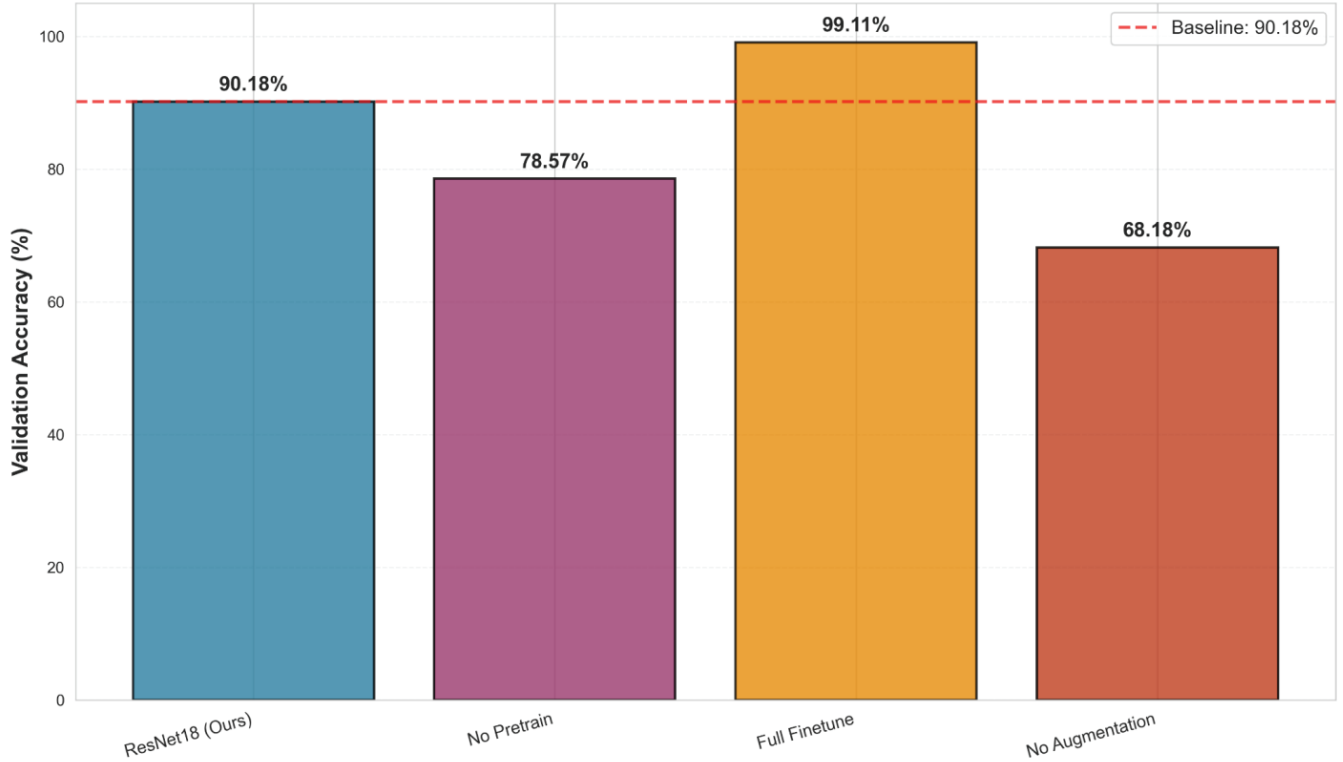
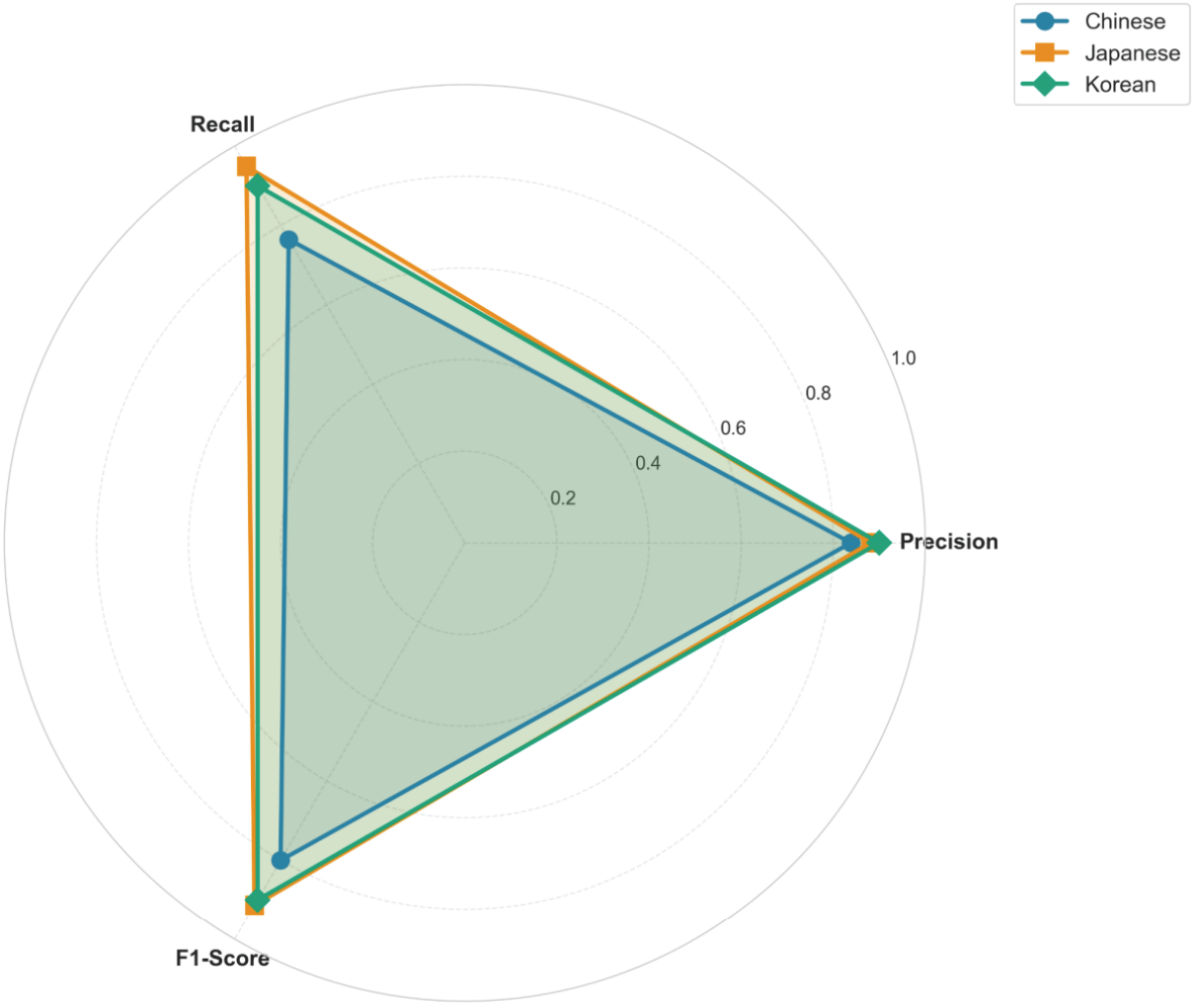


Figure 6: Radar chart of per-class precision, recall, and F1 score for ResNet 18. Korean and Japanese patterns span larger areas than Chinese patterns, reflecting higher per-class discriminability



3.3 Ablation Study

Three controlled ablation experiments isolate the contribution of each pipeline component. Table 4 reports the validation accuracy under each configuration.

Table 4: Ablation Study Results

Configuration	Val Acc (%)	Change (pp)
ResNet18 (Complete Method)	90.18	baseline
w/o Pretrained Weights	78.57	−11.61
Full Model Fine-tuning	99.11	+8.93
w/o Data Augmentation	68.18	−22.00

Data augmentation is the single most critical factor: its removal causes a 22.00 pp drop, demonstrating that geometric and photometric diversity suppresses overfitting more decisively than any other design choice in this low-data regime [9, 10]. Removing pre-trained weights incurs an 11.61 pp drop, confirming that ImageNet features transfer constructively to traditional pattern imagery despite the substantial domain gap [11, 12]. Full finetuning raises validation accuracy to 99.11%, but preliminary experiments showed a substantially larger validation-test gap, confirming overfitting to the 525-image training set. The frozen-backbone strategy thus provides a more robust and deployable solution [20]. Figure 7 visualises the relative magnitude of each component's contribution.

Figure 7: Ablation study validation accuracy

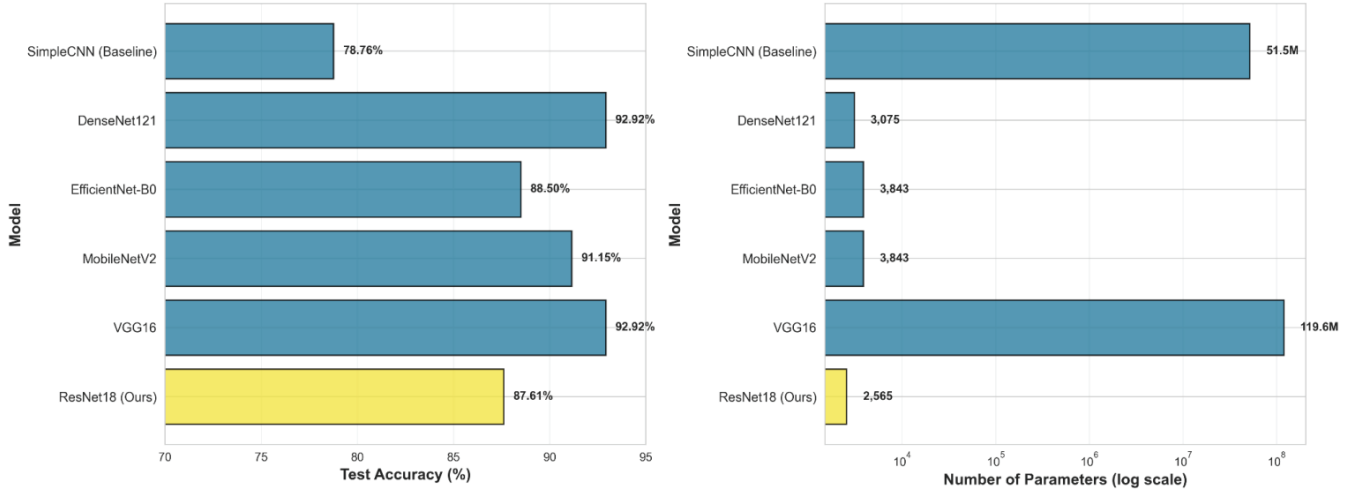


Figure 7 shows the dashed red line marks the complete ResNet18 baseline (90.18%). Data augmentation removal causes the largest drop (−22.00 pp), followed by removing pre-trained weights (−11.61 pp).

4. Discussion

The comparative results reveal a nuanced relationship between model capacity and classification performance. VGG16 and DenseNet121 jointly achieve 92.92% test accuracy, yet their parameter profiles diverge by nearly five orders of magnitude (119.6M vs. 3,075), underscoring that once the backbone is frozen, compact inductive biases can match far larger networks [11]. Although the proposed ResNet18 ranks fourth in raw accuracy, its accuracy-to-parameter ratio is unmatched: it requires 46,625× fewer trainable parameters than VGG16 while closing only 5.31 percentage points of the accuracy gap, a trade-off highly favourable for mobile cultural tourism applications and museum kiosk systems [8, 22].

The per-class analysis exposes an inherent asymmetry reflecting real cultural and historical factors. The dominant Chinese-Korean confusion is consistent with the welldocumented historical diffusion of artistic motifs during the Goryeo and Joseon periods [3, 13], and the stylistic breadth of Chinese patterns across five millennia of dynastic production creates high intra-class variance. The ablation results provide concrete design guidance: augmentation and pre-trained weights are complementary rather than redundant, and any extension to additional cultural categories should prioritise collecting and augmenting training data before investing in more complex architectural modifications. The demonstrated ability of a 2,565-parameter classifier to achieve 87.61% accuracy lowers the barrier to entry for heritage institutions operating under budget constraints, consistent with outcomes observed in adjacent fields including cardiovascular risk prediction [23], software defect detection [24], IoT anomaly defence [25], supply chain scheduling [26], mobile robot path planning [27, 29], cancer drug synergy prediction [28], spatio-temporal traffic forecasting [30], automated scientific writing [31], and autonomous vehicle motion planning [32]. Future work should explore attention-based localisation to focus the model on diagnostically distinctive sub-regions, expand the dataset to additional East Asian cultural categories, and investigate self-supervised pre-training on unlabelled heritage imagery to reduce dependence on costly manual annotation.

Conclusion

This paper proposes a lightweight transfer learning framework for the automated classification of Chinese, Japanese, and Korean traditional patterns. By freezing the ImageNet pre-trained ResNet18 backbone and appending a compact three-class head with only 2,565 trainable parameters, the proposed method achieves 87.61% test accuracy and an F1-score of 0.8746, representing a parameter reduction of over four orders of magnitude relative to VGG16 while closing merely 5.31 percentage points of the accuracy gap. Ablation experiments established that data augmentation and ImageNet pre-training are the two most critical design components, contributing 22.00 pp and 11.61 pp gains, respectively. Per-class analysis revealed that Chinese patterns pose the greatest recognition challenge due to high intra-class stylistic diversity and historical iconographic diffusion, while Korean and Japanese patterns are more reliably discriminated. These findings confirm that a frozen-backbone

transfer learning strategy with targeted augmentation provides an efficient and practically deployable solution for cross-cultural heritage pattern recognition under realistic data and computational constraints.

Funding

No

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

[1] Zhao Liu. (2024). The construction of a digital dissemination platform for the intangible cultural heritage using convolutional neural network models. Heliyon, 10(24), e40854. https://doi.org/10.1016/j.heliyon.2024.e40854

[2] Liu, Y., Cheng, P., & Li, J. (2023). Application interface design of Chongqing intangible cultural heritage based on deep learning. Heliyon, 9(11), e09450. https://doi.org/10.1016/j.heliyon.2023.e22166

[3] Fu, X. (2021). Research and application of ancient Chinese pattern restoration based on deep convolutional neural network. Computational Intelligence and Neuroscience, 2021, 2691346. https://doi.org/10.1155/2021/2691346

[4] Ji, J., Lao, Y., & Huo, L. (2024). Convolutional neural network application for supply-demand matching in Zhuang ethnic clothing image classification. Scientific Reports, 14, 4082. https://doi.org/10.1038/s41598-024-64082-9

[5] Belhi, A., Ahmed, H. O., Alfaqheri, T., Bouras, A., Sadka, A. H., & Foufou, S. (2021). Study and evaluation of pre-trained CNN networks for cultural heritage image classification. In Advances in Intelligent Systems and Computing (pp. 51-67). Springer. https://doi.org/10.1007/978-3-030-66777-1_3

[6] Chen, B. (2022). Classification of artistic styles of Chinese art paintings based on the CNN model. Computational Intelligence and Neuroscience, 2022, 4520913. https://doi.org/10.1155/2022/4520913

[7] Tan, Y. (2022). Feature recognition and style transfer of painting image using lightweight deep learning. Computational Intelligence and Neuroscience, 2022, 1478371. https://doi.org/10.1155/2022/1478371

[8] Lou, J., Shavetov, S., Wen, X., Li, Z., Zhang, X., & Yuan, C. (2026). Deep learning-based instance segmentation: A comprehensive review of algorithms, challenges, and future directions. The Visual Computer, 42(6), 235.

[9] Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., & Shen, F. (2022). Image data augmentation for deep learning: A survey. arXiv preprint arXiv:2204.08610.

[10] Alomar, K., Aysel, H. I., & Cai, X. (2023). Data augmentation in classification and segmentation: A survey and new strategies. Journal of Imaging, 9(2), 46. https://doi.org/10.3390/jimaging9020046

[11] Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345-1359. https://doi.org/10.1109/TKDE.2009.191

[12] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge. International Journal of Computer Vision, 115(3), 211-252. https://doi.org/10.1007/s11263-015-0816-y

[13] Qi, X.-Z., He, X.-M., Chen, S.-W., & Hai, T. (2024). A framework of evolutionary optimized convolutional neural network for classification of Shang and Chow dynasties bronze decorative patterns. PLOS ONE, 19(10), e0293517. https://doi.org/10.1371/journal.pone.0293517

[14] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)(pp. 4510-4520). https://doi.org/10.1109/CVPR.2018.00474

[15] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning (ICML) (pp. 6105-6114).

[16] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770-778). https://doi.org/10.1109/CVPR.2016.90

[17] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv

preprint arXiv:1409.1556.

[18] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[19] Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning (ICML) (pp. 448-456).

[20] Gu, C., & Lee, M. (2024). Deep transfer learning using real-world image features for medical image classification, with a case study on pneumonia X-ray images. Applied Sciences, 14(11), 4083. https://doi.org/10.3390/app14114083

[21] Lin, M., Chen, Q., & Yan, S. (2013). Network in network. arXiv preprint arXiv:1312.4400.

[22] Cheng, W., Liu, Z., & Li, X. (2026). Enhanced lightweight architecture for real-time detection of agricultural pests and diseases. Computers, Materials & Continua, 87(2), 2459-2497. https://doi.org/10.32604/cmc.2025.074250

[23] Yuan, C., Liu, Z., Li, X., Zhou, X., Wang, D., Fan, Y., Sun, X., & Tian, Z. (2025). A dynamic weighted ensemble learning framework for cardiovascular risk prediction in type 2 diabetes: A comparative study with SHAP-based interpretability. Scientific Reports.

[24] Liu, Z., Yuan, C., Liu, H., Li, X., Liu, S., Zhou, X., & Tian, Z. (2026). MSTDP: A multi-scale temporal deep learning framework for just-in-time software defect prediction with cross-attention fusion. Journal of King Saud University Computer and Information Sciences.

[25] Yang, J., Govindarajan, V., Arif, S., Xu, X., Kallel, M., Shaikh, Z. A., Liu, Z., Yuan, C., & Por, L. Y. (2025). SwarmSense-DNN: A trustworthy and decentralized neural framework for proactive anomaly defense in consumer IoT. IEEE Transactions on Consumer Electronics.

[26] Wang, Y., Zhang, H., Yuan, C., Li, X., & Jiang, Z. (2025). An efficient scheduling method in supply chain logistics based on network flow. Processes, 13(4), 969.

[27] Xu, L., Yuan, C., & Jiang, Z. (2025). Multi-strategy enhanced secret bird optimization algorithm for solving obstacle avoidance path planning for mobile robots. Mathematics, 13(5), 717.

[28] Zhang, Y., Yuan, C., Wang, L., Chen, Y., Xing, Y., & Sun, Y. (2025). The structure-preserving spectral graph neural network for dual kinase inhibitors and synergy scoring in gastric cancer. npj Digital Medicine.

[29] Yuan, C., Cai, Y., Que, H., Pei, Y., Zhang, X., Xie, J., Zhang, Q., Mu, L., & Qiao, F. (2026). KA-IHO: A kinematic-aware improved hippo optimization algorithm for collision-free mobile robot path planning in complex grid environments. Sensors, 26(8), 2416.

[30] Yi, X., Yuan, C., Tian, Z., Wu, X., Wu, H., & Liu, L. (2026). BiCoMT: Bidirectional coupling modulation transformer with multi-scale for spatio-temporal traffic flow prediction. Journal of King Saud University Computer and Information Sciences.

[31] Yuan, C., Wei, T., Li, C., Yi, X., Liu, S., Zhang, Z., Cai, Y., & Du, X. (2026). PaperOrchestrator: An LLM-orchestrated multi-agent pipeline for automated end-to-end scientific paper writing. Journal of King Saud University Computer and Information Sciences.

[32] Tian, Z., Lin, Z., Yuan, C., Prajapat, S., Yang, J., & Yee, L. (2026). Consumer-electronics-oriented safe interaction-aware motion planning for ramp driving with perception uncertainty quantification. IEEE Transactions on Consumer Electronics.