

<https://doi.org/10.1038/s41538-025-00690-7>

Phenotypic feature-based identification of tea geographical origin using lightweight deep learning

Check for updates

Guoquan Pei¹, Bing Zhou², Xueying Qian¹, Baijuan Wang³, Wei Chen¹ & Wendou Wu¹ ✉

Accurate identification of the geographical origin of tea leaves is crucial for ensuring quality assurance and traceability within the tea industry. This study introduces Origin-Tea, a novel lightweight convolutional neural network that innovatively combines depthwise separable convolutions with squeeze-and-excitation (SE) attention mechanisms to effectively capture subtle phenotypic variations while minimizing computational costs. Unlike prior approaches that depend on heavy architectures or handcrafted features, Origin-Tea is explicitly designed for efficiency and interpretability in agricultural applications. Comprehensive ablation studies confirm the significant contribution of each architectural component to the model's robust performance. The dataset comprises 900 high-resolution RGB images of Yunkang 10 tea leaves, independently collected from seven distinct regions in Yunnan Province. A 10-fold stratified nested cross-validation (CV) was employed, with one-fold designated for testing, one for validation, and the remaining eight for training in each iteration. Data augmentation techniques, including flipping, rotation, and exposure adjustments, were applied solely to the training set to enhance model robustness without compromising the intrinsic phenotypic features. Origin-Tea achieved an average overall accuracy (OA) of 0.92 ± 0.03 and a Kappa coefficient of 0.90 ± 0.03 , outperforming the best-performing baseline, CoAtNet (OA = 0.89 ± 0.03), by 3.37% accuracy while reducing parameters by over 90% (1.7 M versus 17 M). Furthermore, in an independent test on 1788 scanner-captured images from four villages, Origin-Tea demonstrated excellent generalization with an OA of 0.97. These results highlight the model's potential as a scalable, field-deployable solution for intelligent tea provenance verification and precision phenotyping.

Tea, one of the most widely consumed beverages worldwide, is enjoyed by over two-thirds of the global population¹. Although the tea plant originated in China, its cultivation has since expanded to numerous countries, including India, Indonesia, Sri Lanka, Kenya, Turkey, Vietnam, Japan, Iran, and Argentina. Tea is particularly rich in polyphenolic compounds, many of which possess strong antioxidant properties.

Currently, tea origin authentication primarily relies on multispectral technologies, blockchain integration, and physicochemical testing. Zhang et al.² developed a tea origin classification model for white tea using near-infrared spectroscopy (NIR)³. They applied Principal Component Analysis (PCA)⁴, Linear Discriminant Analysis (LDA)⁵, and the Successive Projections Algorithm (SPA) for dimensionality reduction, followed by training k-Nearest Neighbor (KNN)⁶, Random Forest (RF), and Support Vector Machine (SVM)⁷ models on spectral data. Their results demonstrated the

feasibility of NIR spectroscopy for rapid and accurate origin verification. Wu et al.⁸ integrated machine learning with blockchain technology to enhance the accuracy of source data in the tea supply chain, ensuring secure traceability and improving data reliability. Jin et al.⁹ applied Fourier-transform NIR (FT-NIR) spectroscopy and an Extreme Learning Machine (ELM)¹⁰ model to classify 114 Taiping Houkui (TPHK) green tea samples from four production areas, achieving a classification accuracy of 95.35%. Liu et al.¹¹ analyzed stable isotopes and 31 mineral elements in green tea samples from different regions, using PCA and LDA to identify key classification features, effectively distinguishing green teas from various origins. Yun et al.¹² employed headspace gas chromatography-mass spectrometry (HS-GC/MS) combined with chemometric analysis on 306 black tea samples, identifying 48 volatile compounds. Using k-NN and RF models, they achieved high-precision origin classification, with k-NN reaching 100%.

¹College of Big Data, Yunnan Agricultural University, Kunming, China. ²College of Science, Yunnan Agricultural University, Kunming, China. ³College of Tea, Yunnan Agricultural University, Kunming, China. ✉e-mail: wuwd2004@126.com

Their study highlighted the geraniol-to-linalool ratio is a key indicator for distinguishing black tea origins.

Despite their high classification accuracy, these methods have several limitations. Physicochemical testing is costly and time-consuming, hindering large-scale implementation. Spectral data acquisition is complex and influenced by instrument parameters, environmental conditions, and sample preparation techniques, limiting practical applicability in real-world scenarios. Additionally, while blockchain technology enhances data security, its widespread adoption in agricultural supply chains faces challenges related to data standardization, computational costs, and technical barriers. Future research should focus on developing more efficient, cost-effective, and scalable tea traceability technologies to support sustainable industry growth.

Phenotypic data are essential for modern crop production and intelligent agriculture, as they encompass morphological and physiological traits influenced by both genetics and the growing environment^{13,14}. Environmental factors such as altitude, temperature, sunlight intensity, soil composition, and humidity induce observable variations in plant morphology. For instance, in tea, higher altitudes often result in smaller, thicker leaves with darker green coloration, while leaves grown at lower altitudes tend to be larger and lighter in color^{15–18}. Variations in sunlight exposure can influence leaf pigmentation and vein prominence, and differences in soil nutrient content influence leaf size, shape, and texture¹⁹. These traits provide critical information for applications such as variety identification, growth monitoring, disease detection, and geographic origin determination.

Recent advances in computer vision and deep learning have significantly improved the automation and accuracy of phenotypic analysis. RGB imaging, a widely adopted, non-destructive technique in plant phenotyping, captures digital information across three primary spectral channels: red (R), green (G), and blue (B). Each channel encodes information about leaf pigmentation and surface reflectance, preserving essential morphological and color features such as leaf area, aspect ratio, color histograms, venation patterns, and texture^{20,21}. High-resolution imaging under controlled illumination, employing diffuse lighting, neutral backgrounds, and fixed camera positions, ensures geometric consistency while minimizing shadows, glare, and color distortion. By combining RGB phenotypic imaging with computer vision and deep learning, subtle environmental effects on leaf morphology can be quantified and encoded into discriminative features, enabling accurate and scalable tea origin identification²².

For example, CucumberAI²³ integrated U-Net, MobileNetV2, ResNet50, and YOLOv5 to extract phenotypic features of cucumber fruits. Ai et al.²⁴ applied Inception-ResNet-v2 to RGB images for accurate crop disease classification. Zhu et al.²⁵ employed EfficientNet-B4-CBAM to identify oil tea varieties with high accuracy, showcasing the power of deep models in plant trait recognition.

However, deep learning has been rarely explored for geographic origin identification of tea leaves, a task highly relevant to traceability and quality control. Traditional methods rely on spectroscopy or chromatography (e.g., NIR, Fourier transform infrared (FTIR), MS), which, despite their accuracy, are costly, time-consuming, and unsuitable for field deployment.

To address these limitations, this study proposes Origin-Tea, a lightweight convolutional neural network specifically designed for tea origin recognition based on RGB phenotypic images. The model tackles the fine-grained classification challenge of distinguishing subtle morphological differences arising from regional environmental factors. It employs depthwise separable convolutions to reduce computational complexity and incorporates squeeze-and-excitation (SE) attention mechanisms²⁶ to enhance channel-wise feature recalibration. The hybrid architecture combines MobileNetV3-style bottlenecks and CoAtNet-inspired attention blocks, enabling the network to capture both local and global features while maintaining compactness and efficiency.

A regionally diverse dataset of Yunkang 10 tea leaf images was compiled, comprising samples from seven major tea-producing regions in Yunnan Province. This dataset captures variations in environmental

conditions and provides a robust foundation for deep learning-based tea traceability research.

The main contributions of this study are summarized as follows: First, we introduce Origin-Tea, a lightweight deep learning framework for non-invasive tea origin identification that combines depthwise separable convolutions, SE modules, and attention-enhanced blocks to achieve efficient and discriminative feature extraction. Second, a comprehensive ablation study quantifies the contribution of each architectural component, offering valuable insights for designing compact yet effective networks tailored to agricultural classification tasks. Furthermore, the model's generalization capability is evaluated under domain shift using a scanner-acquired dataset, demonstrating its adaptability to diverse real-world imaging conditions. Finally, a high-quality tea origin dataset has been developed and made publicly available to support future research in image-based crop traceability and smart agriculture.

Results

Parameter sensitivity analysis

To investigate the impact of key training hyperparameters on model performance, a parameter sensitivity analysis was conducted using the baseline architecture, MobileNetV3-Small. The experiments focused on two critical parameters: the number of training epochs and the learning rate. By systematically varying these factors, we aimed to evaluate their effects on convergence behavior, training stability, and overall classification accuracy.

To ensure experimental consistency and computational efficiency, the sensitivity analysis was conducted using the specific data partition from the first iteration of the 10-fold CV, where one fold was designated for testing, one for validation, and the remaining eight for training. This approach provides a controlled evaluation of hyperparameter effects without variability introduced by inter-fold differences, while maintaining the same data distribution ratios as the full CV scheme.

Table 1 summarizes the detailed results of these experiments. Evaluation metrics include overall accuracy (OA), average accuracy (AA), Kappa coefficient (κ), Matthews correlation coefficient (MCC), precision (P), recall (R), and F1-score, offering a comprehensive assessment of model performance under each hyperparameter configuration.

As shown in Table 1, the model achieved its best performance when trained for 100 epochs with a learning rate of 0.001, attaining an OA of 0.82, AA of 0.85, and a Kappa coefficient of 0.79. Increasing the learning rate to 0.01 at the same number of epochs resulted in a substantial drop in accuracy (OA = 0.63), indicating training instability and over-adjustment during optimization.

When the number of epochs was increased to 200 or 300, model performance exhibited noticeable fluctuations. Specifically, extended training at the lower learning rate (0.001) led to performance degradation, likely caused by overfitting or insufficient gradient updates. Conversely, a higher learning rate (0.01) produced relatively more stable results over longer training (OA ranging from 0.78 to 0.76), but did not surpass the optimal performance observed at 100 epochs with a 0.001 learning rate.

Overall, this sensitivity analysis indicates that MobileNetV3-Small achieves optimal convergence and generalization with a moderate training duration of 100 epochs and a low learning rate of 0.001. Based on these

Table 1 | Results of parameter sensitivity experiments using MobileNetV3-Small on the first fold of the dataset

Epochs	Learning rate	OA	AA	κ	MCC	Precision	F1
100	0.001	0.82	0.85	0.79	0.79	0.85	0.82
100	0.01	0.63	0.66	0.56	0.57	0.65	0.63
200	0.001	0.56	0.54	0.47	0.48	0.54	0.53
200	0.01	0.78	0.79	0.74	0.74	0.79	0.77
300	0.001	0.69	0.73	0.63	0.63	0.70	0.68
300	0.01	0.76	0.81	0.71	0.72	0.78	0.76

Table 2 | Performance of Origin-Tea with different block depths (S012)

Model	OA	AA	κ	Precision	MCC	Recall	F1	Params (M)
Origin-Tea-S012-B0	0.90 ± 0.03	0.90 ± 0.02	0.88 ± 0.03	0.91 ± 0.03	0.79 ± 0.03	0.90 ± 0.03	0.90 ± 0.03	2.0
Origin-Tea-S012-B1	0.88 ± 0.03	0.88 ± 0.03	0.85 ± 0.04	0.89 ± 0.03	0.76 ± 0.03	0.88 ± 0.03	0.88 ± 0.03	2.0
Origin-Tea-S012-B2	0.89 ± 0.02	0.88 ± 0.03	0.86 ± 0.03	0.89 ± 0.02	0.77 ± 0.02	0.89 ± 0.02	0.89 ± 0.02	2.0
Origin-Tea-S012-B3	0.89 ± 0.03	0.89 ± 0.03	0.86 ± 0.04	0.89 ± 0.03	0.77 ± 0.03	0.89 ± 0.03	0.89 ± 0.03	2.1
Origin-Tea-S012-B4	0.88 ± 0.03	0.88 ± 0.04	0.86 ± 0.04	0.89 ± 0.03	0.76 ± 0.03	0.88 ± 0.03	0.88 ± 0.03	2.1
Origin-Tea-S012-B5	0.89 ± 0.02	0.89 ± 0.02	0.86 ± 0.02	0.89 ± 0.02	0.77 ± 0.02	0.89 ± 0.02	0.89 ± 0.02	2.1
Origin-Tea-S012-B6	0.88 ± 0.03	0.87 ± 0.03	0.85 ± 0.04	0.88 ± 0.03	0.76 ± 0.03	0.88 ± 0.03	0.87 ± 0.03	2.2
Origin-Tea-S012-B7	0.88 ± 0.03	0.88 ± 0.03	0.85 ± 0.04	0.89 ± 0.03	0.76 ± 0.03	0.88 ± 0.03	0.88 ± 0.03	2.5
Origin-Tea-S012-B8	0.88 ± 0.03	0.88 ± 0.03	0.85 ± 0.04	0.88 ± 0.03	0.76 ± 0.03	0.88 ± 0.03	0.88 ± 0.03	2.8
Origin-Tea-S012-B9	0.88 ± 0.03	0.88 ± 0.03	0.85 ± 0.04	0.88 ± 0.03	0.76 ± 0.03	0.88 ± 0.03	0.87 ± 0.03	3.1

findings from the first-fold evaluation, this parameter configuration was adopted in the subsequent ablation experiments to ensure consistent and reliable model performance across all comparative analyses.

Experimental design of ablation study

To identify the optimal architecture for the Origin-Tea model, we conducted an ablation study focusing on two key components: the shallow feature extraction modules (S1 and S2) and the number of Block layers in the backbone network (ranging from 0 to 9).

In the first part of the study, we systematically removed the S1 and S2 modules, which are responsible for extracting low-level features such as texture and edges. This allowed us to evaluate the impact of shallow feature extraction on the model's overall classification performance.

In the second part, we varied the depth of the model by adjusting the number of Block layers in the deep feature extraction stage. These layers capture higher-level semantic features crucial for distinguishing tea origins. Networks with 0 to 9 Block layers were tested to analyze how increasing depth affects accuracy. To ensure robust and reliable results, a 10-fold CV strategy was employed. Final performance metrics were averaged across all folds, with standard deviations reported to assess stability and generalizability.

This iterative process of module removal and layer variation, combined with CV, provided insights into the contribution of each component and informed the selection of the most effective Origin-Tea configuration.

Effect of block depth on Origin-Tea performance with s012 modules

To evaluate the impact of backbone depth on classification performance and model complexity, we fixed the shallow feature extraction modules and varied the number of Block layers from 0 to 9 (denoted B0–B9). Table 2 summarizes the results, including classification metrics, MCC, and parameter counts (in millions).

The results show that the configuration without any Block layers (Origin-Tea-S012-B0) achieves the best overall performance, with an OA of 0.90 ± 0.03, AA of 0.90 ± 0.02, Kappa coefficient of 0.88 ± 0.03, F1-score of 0.90 ± 0.03, and MCC of 0.79 ± 0.03, while maintaining a compact model size of 2.0 M parameters. MCC, which accounts for true and false positives and negatives, provides a balanced measure, especially under class imbalance, corroborating the trends seen in OA and F1-score.

As Block layers are added, a consistent but modest decline in performance is observed across all metrics, including MCC. For example, OA decreases to 0.88 ± 0.03 with one Block (B1), and MCC drops to 0.76 ± 0.03, indicating slightly reduced correlation between predictions and ground truth. Meanwhile, the parameter count grows moderately from 2.0 M (B0) to 3.1 M (B9).

More detailed analysis shows that OA, F1, and MCC peak at B0 and gradually decline with increasing depth. P follows a similar trend, starting at 0.91 ± 0.03 and decreasing, reflecting reduced classification confidence. The

F1-score decreases from 0.90 ± 0.03 to 0.87 ± 0.03 across B0 to B9, indicating a less balanced trade-off between P and R. AA slightly drops from 0.90 ± 0.02 and stabilizes after B3, suggesting performance saturation. R and Kappa coefficient exhibit similar downward trends, reflecting decreased generalizability and agreement with true labels as depth increases. MCC trends mirror these observations, underscoring diminished predictive reliability at greater depths.

Supplementary Figure 1 illustrates these degradation patterns for P, R, F1-score, and MCC during training under 10-fold CV, reinforcing the consistency of these findings.

In conclusion, for tea origin recognition based on phenotypic features, a shallow configuration without additional Block layers (Origin-Tea-S012-B0) provides the best balance between classification performance and model efficiency. Increasing network depth does not improve, and may even impair, generalization, highlighting the importance of careful architectural design in lightweight models tailored for agricultural applications.

Effect of block depth on Origin-Tea performance with S01 module

This section examines how varying the Block depth affects the classification performance of the Origin-Tea model when only the S01 module is retained. As shown in Table 3, the model achieves its best performance without any Block layers. Specifically, the B0 configuration attains an OA of 0.92 ± 0.03, AA of 0.91 ± 0.03, Kappa coefficient of 0.90 ± 0.03, F1-score of 0.92 ± 0.03, and MCC of 0.80 ± 0.03, with a compact parameter count of only 1.7 M. The MCC metric, accounting for true and false positives and negatives, confirms the model's robust predictive ability even under potential class imbalance.

Across all metrics, including P, R, AA, Kappa, and MCC, the B0 configuration consistently outperforms deeper variants. For instance, P decreases from 0.92 ± 0.03 (B0) to 0.90 ± 0.03 (B1), and MCC similarly drops from 0.80 ± 0.03 to 0.78 ± 0.03, indicating a slight decline in prediction-ground truth correlation with increased depth.

These findings suggest that under the S01-only configuration, deeper architectures offer no performance benefits and may introduce overfitting or redundancy. The shallow structure sufficiently captures relevant phenotypic features, supporting the principle of model parsimony. Supplementary Figure 2 further illustrates slower convergence and widened training-validation gaps in deeper models, consistent with overfitting symptoms.

In summary, the Origin-Tea-S01 model achieves optimal accuracy and efficiency with no Block layers. The MCC trend reinforces that minimalistic architectures are well-suited for agricultural classification tasks, particularly in real-time or resource-constrained scenarios.

Effect of block depth on Origin-Tea performance with S0 module

This section evaluates the performance of the Origin-Tea model when both S1 and S2 modules are removed, while varying the Block depth from B0 to B9. As summarized in Table 4, the S0B0 configuration (no Blocks) achieves a

Table 3 | Performance of Origin-Tea with different block depths (S01)

Model	OA	AA	κ	Precision	MCC	Recall	F1	Params (M)
Origin-Tea-S01-B0	0.92 ± 0.03	0.91 ± 0.03	0.90 ± 0.03	0.92 ± 0.03	0.80 ± 0.03	0.92 ± 0.03	0.92 ± 0.03	1.7
Origin-Tea-S01-B1	0.89 ± 0.03	0.89 ± 0.03	0.87 ± 0.04	0.90 ± 0.03	0.78 ± 0.03	0.89 ± 0.03	0.89 ± 0.03	1.7
Origin-Tea-S01-B2	0.89 ± 0.04	0.89 ± 0.04	0.87 ± 0.05	0.90 ± 0.04	0.78 ± 0.04	0.89 ± 0.04	0.89 ± 0.04	1.8
Origin-Tea-S01-B3	0.89 ± 0.03	0.89 ± 0.03	0.87 ± 0.03	0.90 ± 0.02	0.77 ± 0.03	0.89 ± 0.03	0.89 ± 0.03	1.9
Origin-Tea-S01-B4	0.89 ± 0.02	0.89 ± 0.02	0.87 ± 0.03	0.90 ± 0.02	0.77 ± 0.02	0.89 ± 0.02	0.89 ± 0.03	1.9
Origin-Tea-S01-B5	0.89 ± 0.03	0.89 ± 0.03	0.87 ± 0.03	0.90 ± 0.02	0.77 ± 0.03	0.89 ± 0.03	0.89 ± 0.03	1.9
Origin-Tea-S01-B6	0.88 ± 0.03	0.88 ± 0.03	0.86 ± 0.04	0.89 ± 0.03	0.75 ± 0.03	0.88 ± 0.03	0.88 ± 0.03	2.0
Origin-Tea-S01-B7	0.87 ± 0.04	0.87 ± 0.04	0.85 ± 0.04	0.88 ± 0.03	0.76 ± 0.04	0.87 ± 0.04	0.87 ± 0.04	2.3
Origin-Tea-S01-B8	0.88 ± 0.04	0.88 ± 0.04	0.85 ± 0.04	0.88 ± 0.03	0.75 ± 0.04	0.88 ± 0.04	0.88 ± 0.04	2.6
Origin-Tea-S01-B9	0.89 ± 0.03	0.89 ± 0.03	0.87 ± 0.04	0.90 ± 0.03	0.77 ± 0.03	0.89 ± 0.03	0.89 ± 0.03	2.9

Table 4 | Performance of Origin-Tea under S0 configuration with different block depths (B0--B9)

Model	OA	AA	κ	Precision	MCC	Recall	F1	Params (M)
Origin-Tea-S0-B0	0.71 ± 0.13	0.79 ± 0.09	0.65 ± 0.16	0.80 ± 0.08	0.60 ± 0.11	0.71 ± 0.13	0.69 ± 0.16	0.8
Origin-Tea-S0-B1	0.91 ± 0.02	0.91 ± 0.02	0.89 ± 0.03	0.91 ± 0.02	0.79 ± 0.02	0.91 ± 0.02	0.90 ± 0.02	0.9
Origin-Tea-S0-B2	0.90 ± 0.03	0.90 ± 0.04	0.88 ± 0.03	0.91 ± 0.03	0.79 ± 0.03	0.90 ± 0.03	0.90 ± 0.03	0.9
Origin-Tea-S0-B3	0.91 ± 0.03	0.91 ± 0.03	0.90 ± 0.03	0.92 ± 0.03	0.79 ± 0.03	0.91 ± 0.03	0.91 ± 0.03	1.0
Origin-Tea-S0-B4	0.90 ± 0.02	0.90 ± 0.03	0.88 ± 0.03	0.91 ± 0.02	0.79 ± 0.02	0.90 ± 0.02	0.90 ± 0.02	1.0
Origin-Tea-S0-B5	0.91 ± 0.02	0.91 ± 0.02	0.89 ± 0.03	0.92 ± 0.02	0.80 ± 0.02	0.91 ± 0.02	0.91 ± 0.02	1.0
Origin-Tea-S0-B6	0.91 ± 0.04	0.90 ± 0.04	0.89 ± 0.05	0.91 ± 0.03	0.80 ± 0.04	0.91 ± 0.04	0.91 ± 0.04	1.1
Origin-Tea-S0-B7	0.89 ± 0.02	0.89 ± 0.02	0.87 ± 0.03	0.90 ± 0.02	0.78 ± 0.03	0.89 ± 0.02	0.89 ± 0.02	1.4
Origin-Tea-S0-B8	0.89 ± 0.04	0.89 ± 0.03	0.87 ± 0.05	0.90 ± 0.03	0.76 ± 0.04	0.89 ± 0.04	0.89 ± 0.04	1.7
Origin-Tea-S0-B9	0.90 ± 0.02	0.90 ± 0.02	0.88 ± 0.03	0.91 ± 0.02	0.78 ± 0.03	0.90 ± 0.02	0.90 ± 0.02	2.0

low OA of 0.71 ± 0.13 and an MCC of 0.60 ± 0.11, highlighting the critical role of the S1 and S2 modules in early feature extraction.

Introducing a single Block (S0B1) yields substantial improvements across all metrics: OA increases to 0.91 ± 0.02, F1-score to 0.90 ± 0.02, Kappa to 0.89 ± 0.03, and MCC to 0.79 ± 0.02. With moderate Block depths (S0B2–S0B4), these performance metrics remain relatively stable, indicating that a small number of Blocks can effectively compensate for the absence of shallow modules. However, beyond this range, further increases in Block depth lead to a gradual performance decline, with OA dropping to 0.89 ± 0.02 at S0B7 and MCC decreasing to 0.78 ± 0.03.

P and R exhibit similar trends: P improves from 0.80 ± 0.08 (S0B0) to 0.92 ± 0.03 (S0B5), while R rises from 0.71 ± 0.13 to 0.91 ± 0.04. The F1-score peaks between S0B2 and S0B4 before tapering off. The Kappa coefficient increases from 0.65 ± 0.16 (S0B0) to a maximum of 0.90 ± 0.03 (S0B3), with AA following a similar pattern. The MCC trend corroborates these observations, confirming that moderate Block depth restores prediction-ground truth correlation but that excessive depth offers no additional gains.

Supplementary Figure 3 monitors the evolution of Precision, Recall, and F1-score over the course of 100 training epochs. As observed, the shallowest configuration exhibits significant instability and slow convergence, with metrics fluctuating heavily throughout the training process. In contrast, introducing Block layers accelerates convergence, allowing the models to reach high performance metrics at earlier epochs. However, further increasing the depth does not yield statistically significant improvements in the final metrics, indicating that performance saturation is reached quickly. This confirms that a moderate network depth is sufficient to capture the necessary phenotypic features, while excessive depth leads to redundancy.

In summary, when shallow modules are excluded, a moderate Block depth (B2–B4) effectively recovers model performance. However, excessive

depth yields diminishing returns and may introduce redundancy or overfitting. The MCC analysis reinforces the essential role of shallow feature extractors and supports the design of lightweight architectures tailored for small-sample agricultural classification tasks.

Overall comparison and optimal configuration analysis

This section presents a comprehensive comparison and synthesis of results across three architectural settings: retaining both S1 and S2 modules (S012), retaining only the S1 module (S01), and removing both S1 and S2 modules (S0). The evaluation integrates quantitative metrics from Table 2, Table 3, and Table 4 alongside visual analyses in Fig. 1. The goal is to identify the optimal configuration that balances classification accuracy with computational efficiency for practical tea origin recognition applications.

When both S1 and S2 modules are retained, the shallowest model configuration (Origin-Tea-S1S2-B0) achieves the highest performance, with an OA of 0.90 ± 0.03, AA of 0.90 ± 0.02, and a Kappa coefficient of 0.88 ± 0.03. This model also boasts the smallest parameter count (2.0 M) among all S1S2 variants. Increasing the number of Block layers results in a gradual decline in performance, OA drops to 0.88 ± 0.03 for B1 and remains below the baseline at greater depths. These results indicate that the S1 and S2 modules provide sufficient feature extraction capacity, and additional Blocks neither enhance representation nor improve performance, likely due to redundancy or overfitting.

In the S01 setting, where only the S1 module is retained, the 0-Block variant again outperforms deeper counterparts, achieving an OA of 0.92 ± 0.03, AA of 0.91 ± 0.03, and Kappa of 0.90 ± 0.03 with just 1.7 M parameters. Adding even a single Block layer (S01-B1) causes a performance decline to 0.89 ± 0.03 OA, with similar or lower results observed for deeper models (B2–B9). These findings confirm that S1 alone sufficiently captures spatial and channel features for this task, while deeper Block structures offer

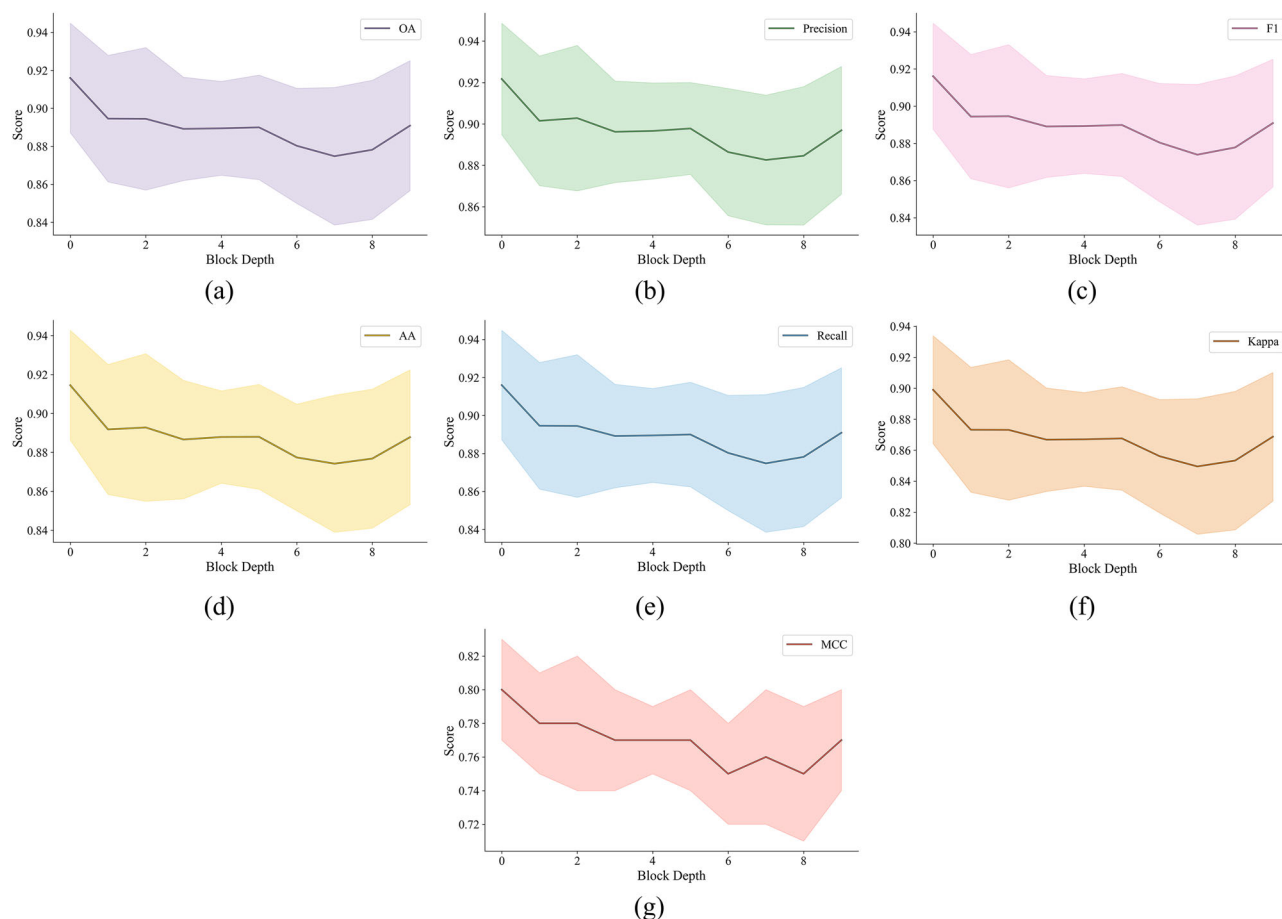


Fig. 1 | Impact of backbone depth on the classification performance of the Origin-Tea model. The panels illustrate how varying the number of MBCConv Blocks (from 0 to 9) within the optimal S01 configuration affects key evaluation metrics. The sub-figures correspond to: (a) Overall Accuracy (OA), (b) Precision, (c) F1-score, (d) Average Accuracy (AA), (e) Recall, (f) Kappa Coefficient, and (g) Matthews Correlation Coefficient (MCC). The solid lines represent the mean performance scores

averaged across the 10-fold cross-validation, while the shaded regions indicate the standard deviation, reflecting the stability of the model. The trends demonstrate that a shallow architecture (Block depth = 0) achieves the highest stability and accuracy, whereas increasing depth leads to performance fluctuations and diminishing returns.

diminishing returns. The minimalistic S01-B0 model thus offers an excellent trade-off between accuracy and efficiency, making it well-suited for light-weight deployment.

Conversely, removing both S1 and S2 modules (S0 configuration) severely compromises performance in the shallow setting (S0-B0: OA = 0.71 ± 0.13), underscoring the importance of early feature extractors for discriminative pattern recognition in tea leaf images. However, performance improves progressively as more Block layers are added, peaking at S0-B3 with an OA of 0.91 ± 0.03 . This suggests that deeper Blocks partially compensate for the absence of S1 and S2. Nonetheless, this gain accompanies increased model complexity and less stable convergence; parameter counts rise from 0.8 M (B0) to 1.0 M (B3), and standard deviations remain relatively large, indicating inconsistent training outcomes.

Synthesizing these observations, the Origin-Tea-S01-B0 model emerges as the optimal architecture. It achieves the highest OA (0.92 ± 0.03), surpassing the best S1S2 model by 2.22% and the best S0 model by 1.10%, while maintaining the smallest parameter size among top-performing variants (1.7 M). Moreover, it consistently outperforms alternatives across additional metrics including P, R, F1-score, AA, MCC, and Kappa, as illustrated in Fig. 1. These findings collectively demonstrate that the S1-only configuration without additional Block layers provides the most effective and efficient architecture for tea origin recognition. Accordingly, this configuration is adopted as the final Origin-Tea model throughout the remainder of this study.

Model comparison experiment

To comprehensively evaluate the effectiveness of the proposed Origin-Tea model for tea geographical origin recognition, we conducted comparative experiments involving a range of mainstream deep learning architectures. These included lightweight models, ResMLP, StartNet, MobileNetV3, MobileNetV4, GhostNetV3, and EfficientNet, as well as larger networks such as InceptionNeXt, ResNet50, and CoAtNet. This diverse selection covers various architectural paradigms and computational complexities, providing a robust benchmark for fair and representative assessment.

Each model was trained using optimization strategies tailored to its architecture, ensuring fair convergence despite differences in depth and capacity. A 10-fold CV scheme was employed to enhance result reliability and generalization across data splits. The accompanying loss curves represent the average convergence behavior of each model under these customized training conditions.

Table 5 presents a quantitative comparison across seven key evaluation metrics, including the MCC. Origin-Tea achieves the highest OA of 0.92 ± 0.03 , outperforming ResNet50 (0.88 ± 0.03) and CoAtNet (0.89 ± 0.03) by 3.54% and 2.44%, respectively. The Kappa coefficient of Origin-Tea stands at 0.90 ± 0.03 , indicating strong agreement with ground truth labels. P (0.92 ± 0.03), R (0.92 ± 0.03), MCC (0.80 ± 0.03), and F1-score (0.92 ± 0.03) are consistently high, reflecting balanced and stable classification performance.

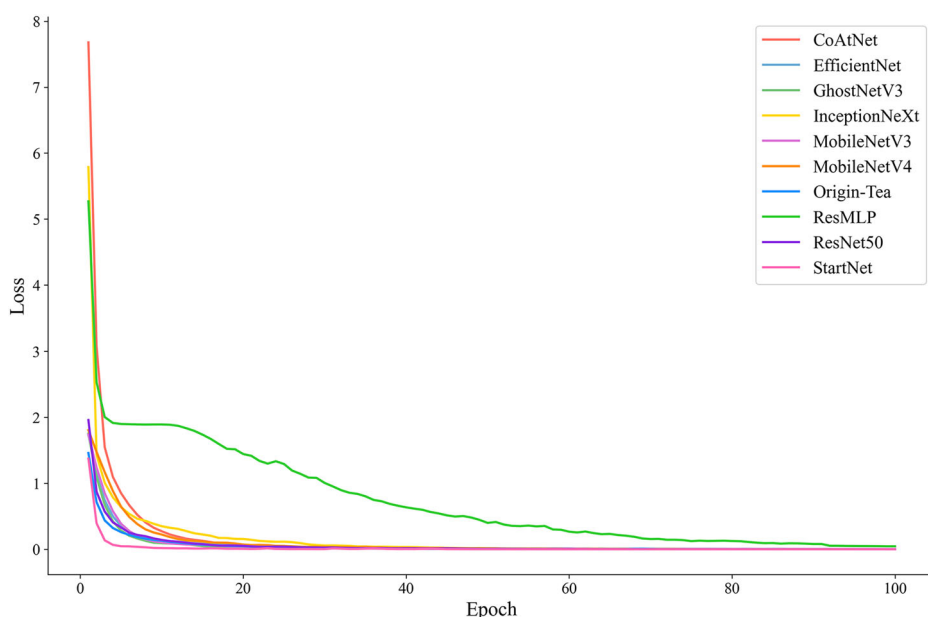
Among the lightweight models, MobileNetV3, which has the same parameter count (1.7 M) as Origin-Tea, achieves a notably lower OA of

Table 5 | Performance comparison of different models

Model	OA	AA	κ	Precision	Recall	MCC	F1	Parameters (M)
ResMLP	0.74 ± 0.03	0.75 ± 0.04	0.69 ± 0.04	0.76 ± 0.03	0.76 ± 0.03	0.70 ± 0.03	0.74 ± 0.03	15.0
StartNet	0.83 ± 0.03	0.82 ± 0.03	0.80 ± 0.04	0.84 ± 0.03	0.83 ± 0.03	0.80 ± 0.04	0.83 ± 0.03	2.7
MobileNetV3	0.71 ± 0.04	0.71 ± 0.05	0.66 ± 0.05	0.72 ± 0.04	0.71 ± 0.04	0.58 ± 0.04	0.71 ± 0.05	1.7
GhostNetV3	0.77 ± 0.04	0.77 ± 0.04	0.73 ± 0.05	0.78 ± 0.04	0.77 ± 0.04	0.64 ± 0.04	0.77 ± 0.04	6.9
MobileNetV4	0.78 ± 0.05	0.78 ± 0.06	0.74 ± 0.06	0.79 ± 0.05	0.78 ± 0.05	0.65 ± 0.06	0.78 ± 0.06	2.5
EfficientNet	0.85 ± 0.03	0.85 ± 0.03	0.82 ± 0.04	0.86 ± 0.03	0.85 ± 0.03	0.73 ± 0.03	0.85 ± 0.03	4.0
InceptionNeXt	0.78 ± 0.12	0.82 ± 0.07	0.74 ± 0.15	0.82 ± 0.07	0.78 ± 0.12	0.67 ± 0.11	0.77 ± 0.14	49.4
ResNet50	0.88 ± 0.03	0.88 ± 0.04	0.86 ± 0.04	0.89 ± 0.03	0.88 ± 0.03	0.76 ± 0.03	0.88 ± 0.03	23.5
CoAtNet	0.89 ± 0.03	0.89 ± 0.03	0.87 ± 0.04	0.90 ± 0.03	0.89 ± 0.03	0.76 ± 0.03	0.89 ± 0.03	17.0
Origin-Tea	0.92 ± 0.03	0.91 ± 0.03	0.90 ± 0.03	0.92 ± 0.03	0.92 ± 0.03	0.80 ± 0.03	0.92 ± 0.03	1.7

Fig. 2 | Comparative training loss convergence analysis across different deep learning models.

The plot displays the training loss curves averaged over the 10-fold cross-validation for the proposed Origin-Tea model and nine comparative baselines (CoAtNet, EfficientNet, GhostNetV3, InceptionNeXt, MobileNetV3, MobileNetV4, ResMLP, ResNet50, and StartNet). The X-axis represents the training epochs, and the Y-axis represents the loss value. The smooth and rapid descent of the curves indicates that all models successfully converged and were effectively optimized under the specified hyperparameters, with Origin-Tea demonstrating competitive convergence speed and stability comparable to larger architectures.



0.71 ± 0.04. GhostNetV3 and MobileNetV4 improve on this baseline with OA scores of 0.77 ± 0.04 and 0.78 ± 0.05, respectively, but still do not match Origin-Tea’s performance. EfficientNet attains higher accuracy (0.85 ± 0.03) but requires 4.0 M parameters, more than twice the size of Origin-Tea, highlighting a trade-off between accuracy and efficiency. ResMLP and StartNet deliver moderate OA scores of 0.74 ± 0.03 and 0.83 ± 0.03, respectively, underscoring the importance of network design tailored to dataset characteristics.

Among larger models, ResNet50 (23.5 M parameters) and CoAtNet (17.0 M parameters) demonstrate strong performance with OA values of 0.88 ± 0.03 and 0.89 ± 0.03, respectively; however, their substantial computational demands limit practical deployment in resource-constrained settings. InceptionNeXt, despite its large size (49.4 M parameters), achieves a relatively modest OA of 0.78 ± 0.12, further emphasizing the efficiency-performance trade-off in deep learning model design.

As shown in Fig. 2, each curve represents the mean training loss averaged over the ten folds. The losses decrease rapidly during the initial epochs and gradually stabilize as training progresses, indicating consistent and stable convergence across all folds and architectures.

Generalization performance under alternative acquisition conditions

To further assess the robustness and generalization ability of the proposed Origin-Tea model, experiments were conducted on the scanner-acquired

dataset. This dataset represents a domain-shift scenario, featuring differences in acquisition modality and imaging conditions such as lighting uniformity, color fidelity, spatial resolution, and background consistency. These domain discrepancies pose significant challenges and offer a rigorous benchmark for evaluating the adaptability of models to heterogeneous data sources.

All models were trained from scratch on the scanner dataset without using pre-trained weights or domain adaptation techniques. This setup ensures that observed performance differences stem solely from the architectural capacity and inductive biases of the models, eliminating confounding effects from transfer learning or prior knowledge.

As shown in Table 6, Origin-Tea achieved the highest performance across all evaluation metrics, attaining an OA of 0.97. This slightly surpasses CoAtNet (OA = 0.96) and InceptionNeXt (OA = 0.93). Notably, Origin-Tea maintained consistent P, R, F1-score, and AA values of 0.97, reflecting balanced classification performance and strong class-wise discrimination.

Despite having only 1.7 million parameters, Origin-Tea outperformed substantially larger models such as CoAtNet (17.0 M) and InceptionNeXt (49.4 M), highlighting the efficiency and effectiveness of its architectural design. Larger networks trained on relatively small datasets are prone to overfitting, which partially explains the lower performance of these high-capacity models. Among lightweight architectures, MobileNetV3 (1.7 M) achieved a considerably lower OA of 0.69, while EfficientNet (4.0 M) reached only 0.60 OA, emphasizing that Origin-Tea’s superior results derive from architectural suitability rather than mere compactness.

Table 6 | Performance of different models on the scanner-based additional validation dataset

Model	OA	AA	κ	Precision	MCC	Recall	F1	Parameters (M)
ResMLP	0.67	0.66	0.56	0.70	0.56	0.67	0.68	15.0
StartNet	0.87	0.87	0.82	0.88	0.83	0.87	0.87	2.7
MobileNetV3	0.69	0.68	0.58	0.71	0.59	0.69	0.69	1.7
GhostNetV3	0.74	0.70	0.65	0.73	0.64	0.74	0.73	6.9
MobileNetV4	0.76	0.75	0.67	0.76	0.67	0.76	0.76	2.5
EfficientNet	0.60	0.58	0.46	0.62	0.46	0.60	0.60	4.0
InceptionNeXt	0.93	0.91	0.90	0.93	0.90	0.93	0.93	49.4
ResNet50	0.88	0.88	0.84	0.89	0.84	0.88	0.88	23.5
CoAtNet	0.96	0.95	0.94	0.96	0.95	0.96	0.96	17.0
Origin-Tea	0.97	0.97	0.96	0.97	0.96	0.97	0.97	1.7

Models such as StartNet and ResMLP further illustrate how both architecture design and parameter scale influence performance under domain-shifted conditions. StartNet achieved solid performance (OA = 0.87, MCC = 0.83) with a relatively low parameter count (2.7 M), whereas ResMLP, despite having a larger parameter count (15.0 M), delivered moderate performance (OA = 0.67, MCC = 0.56). This demonstrates that larger capacity alone does not guarantee better generalization.

These findings validate the strong generalization capacity of Origin-Tea in practical agricultural settings, particularly where image data originate from diverse and uncontrolled sources. Additionally, Supplementary Fig. 4 illustrates the changes in model performance during training, further demonstrating the stability and convergence of the training process.

Discussion

The experimental results presented in this study comprehensively validate the effectiveness and practical utility of the proposed Origin-Tea model for tea origin identification based on leaf phenotype images. This section synthesizes the key findings from multiple perspectives, including model performance, architectural contributions, generalization ability, and real-world applicability.

Origin-Tea consistently outperforms a broad spectrum of baseline models across benchmark evaluations. These include conventional convolutional neural networks (CNNs) such as StartNet and ResNet50, lightweight architectures like MobileNetV3/V4 and GhostNetV3, as well as hybrid models such as CoAtNet and InceptionNeXt. As shown in Table 5, Origin-Tea achieves the highest OA, AA, and F1-score, alongside strong Kappa coefficient and class-wise consistency. Remarkably, these advantages come with only 1.7 million parameters, minimizing overfitting risk on the relatively small Origin-Tea dataset. This demonstrates that Origin-Tea's superior performance stems from its efficient and well-suited architectural design, rather than from mere model compactness.

Ablation studies reveal that the hybrid design, integrating MobileNetV3-style bottleneck layers with attention-augmented S-Blocks inspired by CoAtNet, drives this performance. MobileNetV3 components efficiently capture local features with minimal computational overhead, while the S-Blocks, which incorporate depthwise separable convolutions, SE modules, and residual connections, enhance global context modeling and channel recalibration. This combination enables Origin-Tea to effectively extract both fine-grained and large-scale phenotypic features crucial for discriminating tea leaf origin. Removing any key architectural element consistently reduces accuracy and generalization, confirming their necessity.

To assess robustness under domain shift, Origin-Tea was tested on a scanner-acquired dataset differing markedly in acquisition device, lighting, and background. Without domain adaptation or pretraining, Origin-Tea achieved an OA of 0.97 and Kappa of 0.96, outperforming larger models such as CoAtNet (OA = 0.96) and InceptionNeXt (OA = 0.93). This strong generalization highlights the model's ability to adapt to heterogeneous

imaging conditions, underscoring its reliability for real-world deployment in uncontrolled or non-curated environments.

Origin-Tea's combination of high accuracy, low parameter count, and strong generalization positions it as a practical solution for agricultural applications, including in-field tea origin verification, mobile quality assessment, and digital traceability. Unlike spectral or biochemical fingerprinting techniques that require specialized and costly instruments, Origin-Tea operates solely on RGB images, acquirable via common devices such as smartphones or flatbed scanners, greatly simplifying deployment and reducing costs.

Compared with existing approaches, the advantages of Origin-Tea become even more pronounced. For instance, Zhang et al.² utilized NIR spectroscopy combined with PCA/LDA and traditional classifiers such as KNN and SVM to classify tea origins, achieving competitive accuracy. However, their approach depends on complex instrumentation and extensive preprocessing, which limits accessibility and practical deployment. In contrast, Origin-Tea leverages inexpensive RGB imaging alongside a lightweight deep learning model, enabling rapid, cost-effective deployment with minimal technical expertise required.

Lin et al.²⁷ employed laser-induced breakdown spectroscopy (LIBS) combined with transfer learning to achieve high accuracy (93.81%) under small-sample conditions. However, their approach relies on specialized spectroscopic equipment and domain adaptation techniques, which can complicate deployment. In contrast, Origin-Tea attains higher accuracy (96.91%) using only raw RGB images and a single-stage classification model without the need for domain adaptation, highlighting its greater practicality and ease of use.

Li et al.²⁸ proposed a FTIR-based fingerprinting workflow combined with SVM and KNN classifiers, achieving perfect accuracy (100%) on a large dataset of black tea samples. Despite this impressive performance, their approach depends on specialized FTIR instrumentation and carefully controlled sampling protocols, which limit its practicality for real-time or field-level applications. In contrast, Origin-Tea, while slightly lower in accuracy, offers a more flexible and cost-effective solution suitable for on-site deployment.

Peng et al.²⁹ utilized metabolomics analysis via UHPLC-Q/TOF-MS to discriminate fine-grained tea origins, achieving 100% classification accuracy with models such as feedforward neural networks and RFs. Although these methods provide valuable biochemical insights, they are not easily scalable for widespread use due to the need for sophisticated laboratory equipment, extensive sample preparation, and lengthy processing times. By comparison, Origin-Tea presents an efficient, rapid, and non-destructive alternative tailored for real-time agricultural applications.

While the current model demonstrates strong performance and generalization within the study's scope, several limitations should be acknowledged. First, the model was trained and validated on tea samples from seven specific geographical regions in China, collected during a single week in 2025. Although scanner-based testing indicates some domain

Table 7 | Geographical Information of Sampling Locations

Region	Latitude	Longitude	Altitude (m)
Nanjian	24°50'37.8908"N	100°15'55.6685"E	2167.971
Hejiashai	22°44'53.2949"N	100°52'22.6209"E	1397.708
Reservoir No. 8	21°59'53.2863"N	100°24'29.3776"E	1261.477
Daxinzai	22°2'39.0325"N	100°25'1.2329"E	1211.367
Aluo Village	23°44'17.8999"N	101°32'13.6789"E	1863.967
Mangda	22°0'37.2774"N	100°21'59.5661"E	1276.149
Guangjingmeng	22°1'23.2681"N	100°24'37.8038"E	1281.397

robustness, its effectiveness in distinguishing teas from other regions within China or internationally remains to be validated. Expanding the dataset to include additional geographic regions, diverse cultivation practices, and international samples would help evaluate the model's ability to capture subtle phenotypic variations and address challenges such as overlapping morphological features among different regions or cultivars.

Second, the dataset covers only one harvest season and fixed tea cultivars. Future research should incorporate multiple seasons, production years, and a wider variety of cultivars to ensure the model's robustness against seasonal, environmental, and genetic variability. Such expansion is crucial for developing a truly reliable tea provenance verification system.

Third, the current workflow assumes clean, pre-segmented leaf images, which may not always be feasible in field conditions. Integrating automated segmentation, noise reduction, and occlusion handling strategies could further improve model robustness in real-world scenarios.

Finally, while this study focuses on static classification, future work could explore dynamic or incremental learning techniques to adapt the model to new tea varieties or evolving data distributions without requiring full retraining. Additionally, approaches like unsupervised domain adaptation and few-shot learning could enhance practical applicability in rapidly changing or data-scarce environments, facilitating broader deployment across diverse agricultural contexts.

In summary, this study proposes Origin-Tea, a lightweight and efficient deep learning model tailored for tea leaf origin identification using phenotypic images. By integrating large-kernel convolutional blocks within an optimized architecture, the model strikes an effective balance between high accuracy and low computational cost. Extensive experiments on benchmark datasets and a scanner-based validation set demonstrate that Origin-Tea not only achieves state-of-the-art classification performance but also exhibits strong generalization across heterogeneous acquisition conditions. Compared to conventional CNNs and contemporary hybrid models, Origin-Tea delivers superior accuracy with substantially fewer parameters, making it highly suitable for practical deployment in real-world agricultural applications.

The key contributions and significance of this work are summarized as follows. First, regarding task-specific, lightweight model design, Origin-Tea is specifically tailored for tea origin recognition, featuring only 1.7 million parameters to balance high classification accuracy with computational efficiency. Second, in terms of multi-metric performance, Origin-Tea consistently surpasses both traditional and state-of-the-art models across key evaluation metrics, including OA, AA, P, R, F1-score, and Kappa, demonstrated on both the primary dataset and scanner-acquired test sets. Third, the model shows robust generalization to diverse conditions, demonstrating strong adaptability and consistent performance on images captured using different acquisition modalities. Finally, regarding applicability to intelligent agriculture, this work advances smart agricultural systems by delivering a lightweight, easily deployable solution for traceability, quality control, and origin verification in tea production.

In future work, we plan to extend the model's applicability to a broader range of crops and diverse environmental conditions. We will also explore model compression techniques, real-time deployment pipelines, and multi-modal fusion frameworks to enhance inference efficiency and robustness in

complex agricultural settings. These advancements will lay the groundwork for wider adoption of AI-powered traceability technologies, supporting sustainable production practices and brand protection across global tea markets.

Methods

Research location

The geographic coordinates of the research sites were obtained using the onboard GPS system of the DJI Mavic 3 M drone. Sampling was conducted across seven representative tea-producing regions in Yunnan Province, including Nanjian (Dali), Hejiashai (Simao), Aluo Village (Mojiang), and four locations in Menghai County, Reservoir No. 8, Daxinzai, Guangjingmeng, and Mangda. These regions were selected to capture the ecological and environmental diversity of Yunnan's tea-growing areas, which vary significantly in climate, soil types, and altitude. Detailed geographical information for each sampling site is provided in Table 7.

Nanjian, located in Dali Prefecture, sits at the highest elevation among all sampling sites (2167.971 m). Its highland terrain creates a cool climate with significant diurnal temperature variations, which promote the synthesis of secondary metabolites in tea leaves, compounds often linked to enhanced flavor and aroma quality³⁰. In contrast, Hejiashai in Pu'er, situated at 1397.708 m, represents a mid-altitude site characterized by a warm, humid subtropical climate. This moderate elevation provides optimal sunlight and rainfall, fostering favorable conditions for producing high-yield and high-quality tea³¹.

Menghai County, a major tea production area in Yunnan, encompasses several lower-elevation sampling sites. Reservoir No. 8 (1261.477 m), Daxinzai (1211.367 m), Guangjingmeng (1281.397 m), and Mangda (1276.149 m) share similar ecological traits, warm temperatures, abundant rainfall, and humid conditions that support rapid vegetative growth and rich biochemical development in tea leaves.

Aluo Village in Mojiang, centrally located in Yunnan at an elevation of 1863.967 m, serves as a transitional zone both geographically and climatically. It bridges the high-altitude environment of Nanjian and the lowland climate of Menghai, offering moderate conditions that combine the benefits of both extremes, stability in growth and richness in chemical constituents.

Together, these seven sites span a broad altitudinal and climatic gradient, reflecting the geographic and ecological complexity of Yunnan's tea-growing regions. This diversity provides a robust foundation for investigating how environmental factors influence tea leaf morphology and facilitates a comprehensive assessment of tea origin classification through image-based phenotypic analysis^{32,33}. The geographical distribution of the sampling sites, along with representative views of the tea gardens, are illustrated in Fig. 3.

Materials

Yunkang No. 10 tea³⁴, identified by the reference number GS13020-1987, was selected as the experimental subject in this study. As a variety propagated asexually, it ensures genetic uniformity across individuals, thereby minimizing variability in phenotypic traits. This genetic stability is crucial for controlled scientific research, allowing for more precise analysis of environmental influences on tea leaf characteristics³⁵. Yunkang No. 10 is widely recognized for its superior quality, characterized by a robust flavor, refined aroma, and a rich profile of bioactive compounds, making it particularly well-suited for phenotypic analysis and classification tasks.

Tea samples were manually harvested following the "one bud and two leaves" standard³⁶, a widely accepted criterion representing the optimal developmental stage for evaluating tea quality. This standard ensures consistency in leaf maturity and biochemical composition across samples. Harvesting was conducted over a one-week period between March and April 2025, coinciding with the spring tea season³⁷. Spring tea is generally considered the most representative and highest-quality harvest of the year due to favorable climatic conditions during early growth³⁸. By standardizing both the picking criteria and the harvesting window, this approach

Fig. 3 | Geographical distribution of the sampling locations. The central map illustrates the research sites in Yunnan Province, China, where Yunkang No. 10 tea samples were collected. The colored regions indicate the four main administrative areas: Nanjian (blue), Simao (yellow), Mojiang (pink), and Menghai (purple). Surrounding images display representative aerial views of the seven specific tea plantations (Hejiazhai, Nanjian, Aluo Village, Reservoir No. 8, Daxinzhai, Mangda, and Guangjingmeng), reflecting the diverse ecological environments and terrain variations of the data acquisition sites.

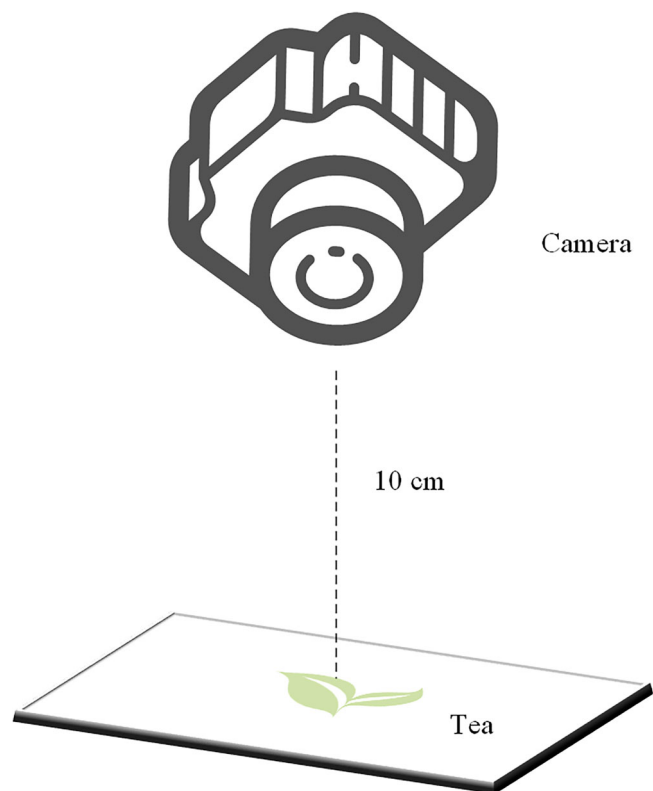
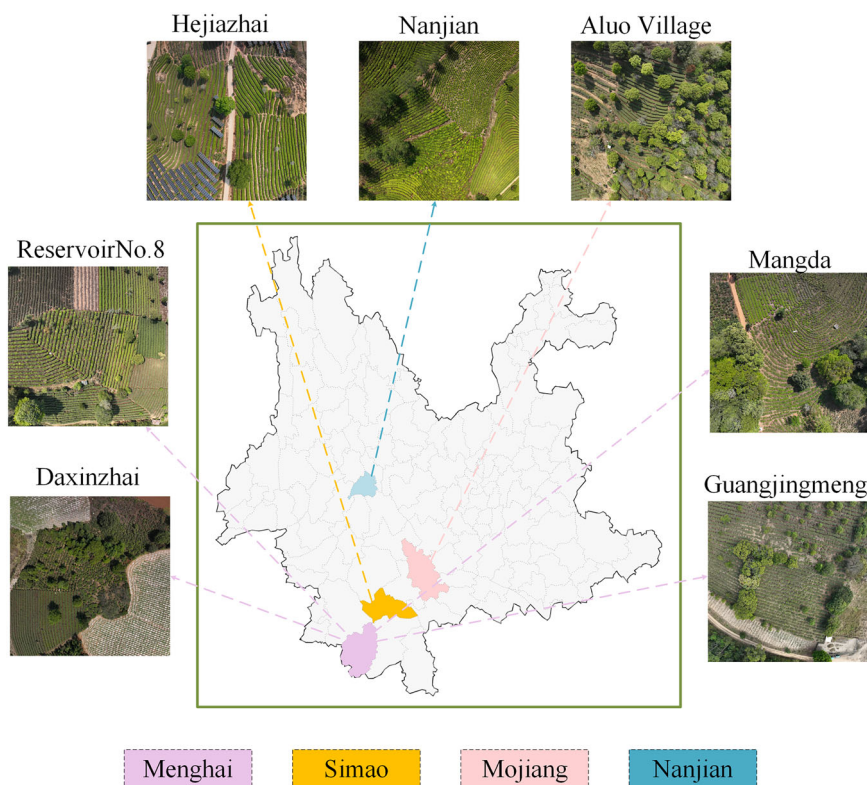


Fig. 4 | Schematic diagram of the RGB image acquisition setup. The diagram illustrates the standardized photography environment used for collecting the primary dataset. A digital camera (Canon EOS RP) was positioned vertically above the tea leaf samples at a fixed distance of 10 cm to ensure consistent spatial resolution and geometric fidelity. Individual tea shoots, selected according to the “one bud and two leaves” standard, were placed on a flat surface for high-resolution RGB imaging under controlled settings.

effectively reduces variability caused by developmental stage or seasonal factors, thereby enhancing the reliability and comparability of subsequent analyses.

A depiction of the data collection scenario in the tea plantation is shown in Supplementary Fig. 5.

Data acquisition

In this study, tea leaf samples were collected using a five-point sampling method, adapted to accommodate the irregular shapes of the tea gardens. Prior to sampling, a DJI drone was employed to observe and assess the spatial layout of each plantation. Based on the aerial imagery, five representative sampling points were selected to capture the heterogeneity within each tea garden. At each point, fresh tea leaves were hand-picked from the canopy layer of the tea trees to ensure consistency in leaf maturity and environmental exposure³⁹.

Immediately after harvesting at each site, the collected tea leaves were individually photographed to preserve their original morphological and color features. Each tea shoot, consisting of one bud and two leaves, was placed separately on a flat tabletop. Images were captured using a Canon EOS RP camera equipped with an RF 50 mm f/1.8 lens. The camera was positioned vertically above the leaf at a fixed distance of 10 cm. Camera settings were configured as follows: aperture *f*/4, ISO 300, and shutter speed 1/100 s. These settings ensured an adequate depth of field to keep all leaf features in sharp focus, while minimizing motion blur and controlling exposure under consistent lighting conditions. The fixed distance and lens choice provided high-resolution images with minimal distortion, accurately preserving the leaves’ morphological and color characteristics. The complete imaging setup is illustrated in Fig. 4.

A total of 900 high-resolution RGB images of tea leaves were collected from seven distinct tea-growing regions to form the primary dataset. The number of images acquired from each region is summarized in Table 8. Among these, Hejiazhai and Mangda contributed the largest shares, with 200 images each, reflecting their larger plantation areas and higher yield capacities. The remaining regions, Nanjian, Reservoir No. 8, Daxinzai, Aluo Village, and Guangjingmeng, each contributed 100 images, ensuring a

balanced and diverse representation of tea phenotypes across various geographical locations. The collected images capture the natural heterogeneity in leaf shape, texture, and coloration caused by environmental and agronomic factors, providing a rich foundation for training and evaluating the proposed model.

Data processing

To effectively extract tea leaf regions from RGB images, a semi-automated, color-based segmentation approach was employed⁴⁰. For each geographic region, 10 representative images were manually selected and imported into Adobe photoshop (PS). Using the eyedropper tool, dominant color values of the tea leaf surfaces were sampled to estimate the characteristic color distribution for each target region.

Based on these sampled values and through a series of empirical evaluations, a unified threshold range in the RGB color space was defined to isolate tea leaf pixels from the background. After multiple trials under varying lighting conditions and environmental backgrounds, the optimal threshold range was determined as follows:

$$T_{\text{lower}} = \begin{bmatrix} 20 \\ 30 \\ 0 \end{bmatrix}, T_{\text{upper}} = \begin{bmatrix} 170 \\ 180 \\ 120 \end{bmatrix} \quad (1)$$

This relatively broad threshold range was carefully chosen to accommodate intra-class variability in leaf pigmentation and diverse image acquisition conditions, while effectively excluding non-leaf elements such as varying lighting effects, color fluctuations, and background structures. The thresholds were refined through iterative adjustments and visual validation

Table 8 | Number of RGB images acquired by region

Region	Number of Images
Nanjian	100
Hejiazhai	200
Reservoir No.8	100
Daxinzai	100
Aluo Village	100
Mangda	200
Guangjingmeng	100
Total	900

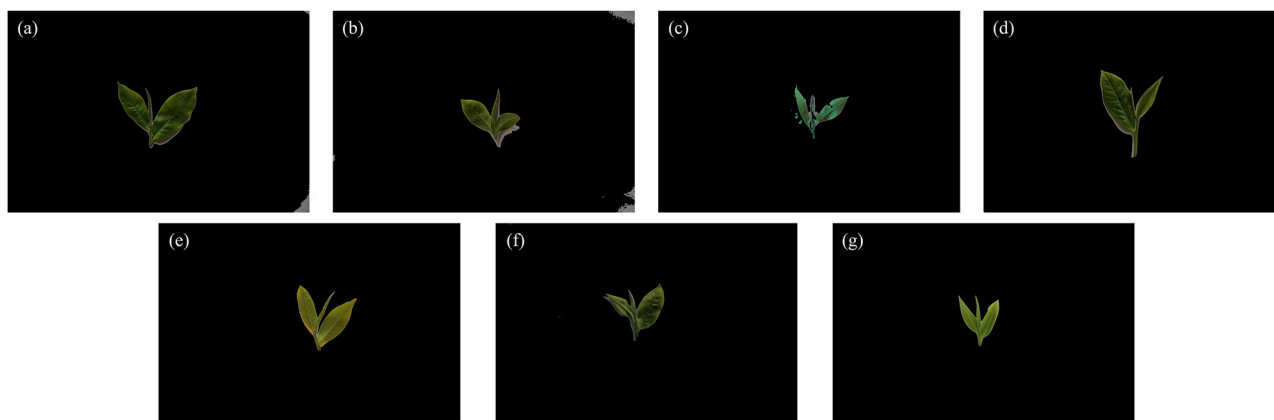


Fig. 5 | Representative examples of segmented tea leaf images from the seven study regions. The panel displays the results of the semi-automated color-based segmentation process used to isolate tea leaf regions from the original field-captured images. By applying specific RGB threshold ranges, background interference was

effectively removed while preserving the morphological integrity and color fidelity of the “one bud and two leaves” samples. The sub-figures correspond to samples from

to maximize the inclusion of complete leaf regions and preserve key morphological features essential for downstream phenotypic analysis. By applying this RGB-based thresholding technique, tea leaf regions were consistently segmented from all input images through color range selection⁴¹. This process effectively eliminated background interference while maintaining the structural integrity of the leaf contours. Representative segmentation results from the seven geographic regions are shown in Fig. 5, demonstrating the robustness and practicality of the proposed method for subsequent feature extraction and model development.

To further improve model robustness and generalization, a series of data augmentation techniques were applied to the segmented tea leaf images⁴². The original dataset comprised 900 manually segmented images collected from seven geographic regions.

To ensure proportional representation of each region across all subsets, a stratified split was performed. Data were shuffled within each region prior to splitting to randomize sample selection.

For model evaluation, a 10-fold cross-validation (CV) scheme was employed. The dataset was first partitioned into 10 equally sized stratified folds, guaranteeing that each fold contained a representative proportion of images from each region. In each CV iteration, one fold was designated as the test set, while the remaining nine folds were further split into training (eight folds) and validation (one fold) sets. This design ensures each fold serves exactly once as the test set, while validation folds rotate in a nested manner for hyperparameter tuning and early stopping. Within each training fold, five augmentation operations, horizontal flip, vertical flip, random rotation, exposure enhancement, and exposure reduction, were applied exclusively to training images, generating five additional variants per original image. This expanded the training set size and enhanced model robustness.

These augmentations were implemented using Python-based image processing libraries to ensure consistency and efficiency. Representative examples of the applied augmentations are shown in Fig. 6, illustrating how each transformation preserves essential tea leaf features while introducing realistic variability. The augmented dataset served as input for model training and evaluation, effectively enhancing the model's ability to recognize phenotypic features⁴³.

Additional validation dataset

To further evaluate the effectiveness and generalizability of the proposed method, an additional batch of Yunkang No. 10 tea leaf samples was collected from four distinct tea plantations: Dele Village, Mangdan Village, Santa Village, and Xiaomenglong Village. Unlike the original image acquisition process, which utilized handheld camera photography, this supplementary dataset was acquired using a flatbed scanner (Canon

effectively removed while preserving the morphological integrity and color fidelity of the “one bud and two leaves” samples. The sub-figures correspond to samples from (a) Reservoir No. 8, (b) Daxinzai, (c) Guangjingmeng, (d) Hejiazhai, (e) Aluo Village, (f) Mangda, and (g) Nanjian.

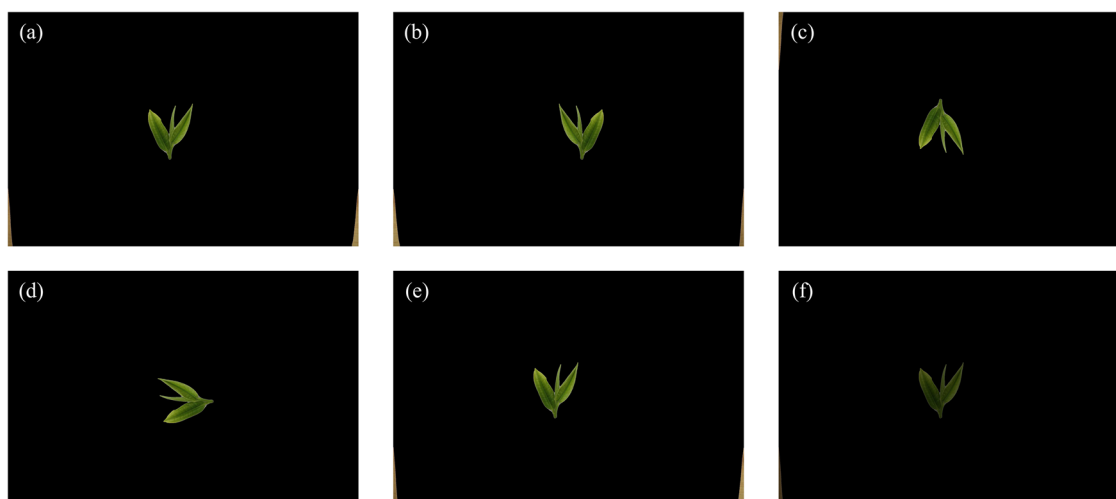
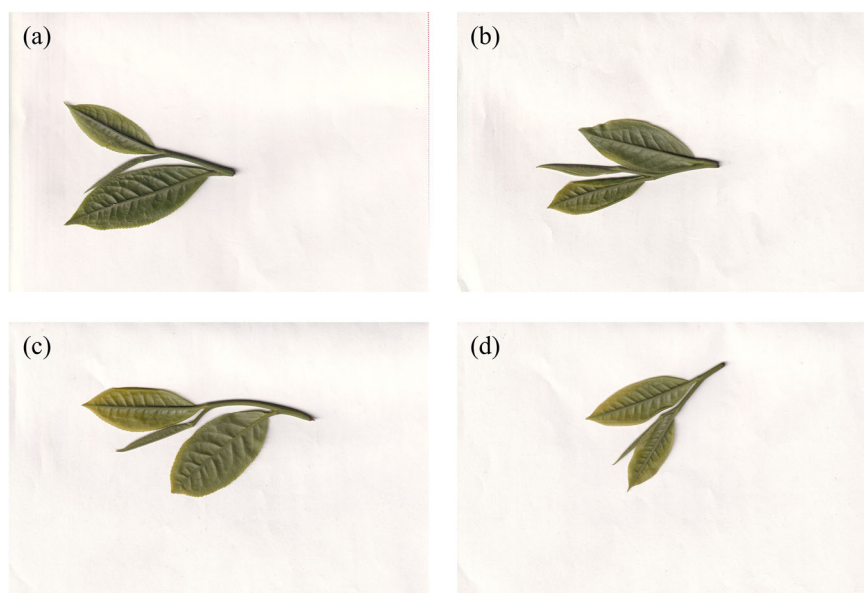


Fig. 6 | Illustration of data augmentation techniques applied to the training dataset. To enhance model robustness and prevent overfitting, five distinct augmentation strategies were applied to the original segmented images (a). The

generated variations include: (b) Horizontal flip, (c) Vertical flip, (d) Random rotation, (e) Exposure enhancement, and (f) Exposure reduction.

Fig. 7 | Representative scanner-acquired tea leaf samples from the additional validation dataset. To evaluate the model's generalization capability under domain-shift conditions, an independent dataset was collected from four villages using a flatbed scanner (Canon CanoScan LiDE 400). Unlike the camera-captured images, these samples feature uniform illumination and a consistent white background, eliminating the need for segmentation. The sub-figures display representative samples from: (a) Dele Village, (b) Mangdan Village, (c) Santa Village, and (d) Xiaomenglong Village.



CanoScan LiDE 400)⁴⁴. The scanner operated in its default high-resolution RGB scanning mode at 300 dpi. The use of a scanner was motivated by its ability to produce uniformly illuminated, shadow-free, and geometrically consistent images, thereby eliminating common issues associated with camera-based imaging such as lighting variability, white balance inconsistencies, and lens distortion. This controlled imaging environment provides a more reproducible basis for model validation⁴⁵.

Unlike the camera-based acquisition, which required color-based background removal, the scanner-acquired images were captured under uniform illumination and on a consistent white background. Therefore, no RGB threshold segmentation was necessary, as the scanning process inherently eliminated background interference, resulting in clean, shadow-free, and geometrically consistent images.

To ensure color fidelity and consistency, a standardized color calibration card was included in each scan, allowing for accurate post-processing color correction and normalization. Following image acquisition, the same data augmentation strategies were applied to the scanned samples, including horizontal and vertical flips, random

rotations, and exposure adjustments. These augmentations help simulate diverse visual conditions, further enhancing the model's generalization capabilities.

Representative scanned samples from the four geographic regions are shown in Fig. 7. These examples highlight the visual differences in leaf morphology and color characteristics among growing regions, providing important cues for model-based origin discrimination.

As summarized in Table 9, the RGB dataset was partitioned into training, validation, and testing subsets. Data augmentation techniques were applied exclusively to the training set to improve model robustness. The final validation dataset contains 1788 RGB images, with Mangdan Village contributing 588 images, Dele Village 480 images, and Santa Village and Xiaomenglong Village each providing 360 images, ensuring balanced representation across locations. By introducing a different imaging modality from that used during training, this scanner-based dataset supports comprehensive performance evaluation and lays the groundwork for future studies on model generalization across heterogeneous imaging conditions.

Model architecture: Origin-Tea

The proposed Origin-Tea model is designed for efficient and accurate tea origin identification, emphasizing lightweight deployment without compromising classification performance. The architecture integrates the strengths of MobileNetV3⁴⁶ and CoAtNet⁴⁷, forming a novel hybrid network that combines convolutional locality with attention mechanisms.

The motivation behind this hybrid design is twofold: first, MobileNetV3 provides an efficient convolutional architecture with low computational cost, making it well-suited for deployment on resource-constrained devices; second, CoAtNet incorporates attention mechanisms that enable the model to capture rich global contextual information, thereby enhancing overall classification accuracy. By combining these approaches, Origin-Tea achieves a balanced trade-off between computational efficiency and feature expressiveness, critical for robust tea image recognition.

As illustrated in Fig. 8a, which depicts the main structure of the Origin-Tea model, the overall pipeline consists of five primary components: data

preprocessing, an initial convolutional stem (S0), spatial attention-enhanced feature extraction blocks (S-Blocks), a sequence of lightweight MobileNetV3 bottlenecks (MBConv), and a fully connected classifier head.

In the preprocessing stage, input images are resized to 224 × 224 pixels and normalized using ImageNet statistics:

$$x' = \frac{x - \mu}{\sigma} \tag{2}$$

where x denotes the raw image pixel value, and μ and σ represent the channel-wise mean and standard deviation derived from the ImageNet dataset, respectively.

The normalized images are then fed into the convolutional stem (S0), as detailed in Fig. 8c. This stem consists of two consecutive 3 × 3 convolutional layers. The first layer applies a convolution operation with a stride of 2 to downsample the spatial resolution:

$$Y_{i,j,k} = \sum_{m,n,c} W_{m,n,c,k} \cdot X_{s-i+m,s-j+n,c} + b_k, s = 2 \tag{3}$$

where X and Y represent the input and output feature maps, respectively; W denotes the convolution kernel weights; b is the bias term; and s is the stride. This convolutional layer expands the input channels from 3 to 64. It is followed by batch normalization (BN) and GELU activation⁴⁸:

$$\text{GELU}(x) = x \cdot \Phi(x), \Phi(x) = \frac{1}{2} \left[1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right) \right] \tag{4}$$

The second convolutional layer uses a stride of 1, preserving both spatial resolution and the number of channels, and is again followed by BN and GELU activation.

Table 9 | Number of RGB Images in the Additional Validation Dataset

Region	Number of Images
Dele Village	480
Mangdan Village	588
Santa Village	360
Xiaomenglong Village	360
Total	1788

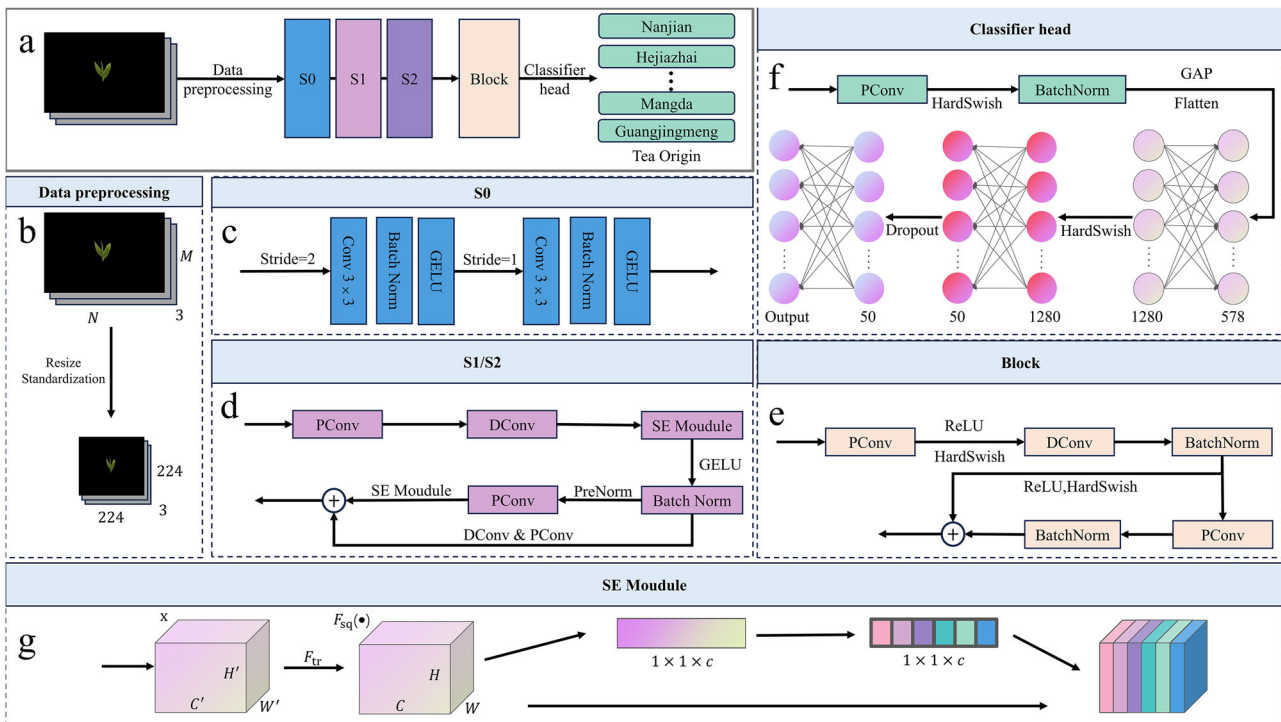


Fig. 8 | Schematic architecture of the proposed Origin-Tea model. The figure details the complete pipeline and internal components of the lightweight network designed for tea origin identification. **a** The overall network architecture, showing the sequential flow from data preprocessing through the stem (S0), spatial attention stages (S1, S2), backbone blocks, to the final classifier head. **b** Data preprocessing module for image resizing and standardization. **c** The convolutional stem (S0) responsible for initial feature extraction and downsampling. **d** The spatial attention-

enhanced feature extraction block (S1/S2), incorporating depthwise separable convolutions and residual connections. **e** The lightweight backbone Block (MBConv) designed to balance efficiency and representation. **f** The classifier head, comprising pointwise convolution, global average pooling (GAP), and fully connected layers. **g** The Squeeze-and-Excitation (SE) module employed for channel-wise feature recalibration.

The spatial attention-enhanced feature extractor (S-Blocks), illustrated in Fig. 8d, comprises two stages: S1 and S2. Each stage begins with a pointwise convolution to expand the channel dimensions:

$$Y_{i,j,k} = \sum_c W_{c,k} \cdot X_{i,j,c} + b_k \tag{5}$$

Next, a depthwise convolution with stride 2 (in S1) performs spatial filtering and downsampling:

$$Y_{i,j,c} = \sum_{m,n} W_{m,n,c} \cdot X_{s+i+m,s+j+n,c} + b_c, s = 2 \tag{6}$$

Following this, the SE module recalibrates channel-wise feature responses:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_{i,j,c} \tag{7}$$

$$s = \sigma(W_2 \cdot \delta(W_1 \cdot z)) \tag{8}$$

where z is the channel-wise global average pooled feature, W_1, W_2 are the learned weights of two fully connected layers, δ denotes the ReLU activation function, and σ represents the sigmoid activation⁴⁹. The recalibrated features are then obtained by channel-wise multiplication:

$$\hat{X}_{i,j,c} = s_c \cdot X_{i,j,c} \tag{9}$$

The block follows the SE module and adopts PreNorm, applying normalization before the convolutional layers. It includes residual skip connections⁵⁰ to facilitate gradient flow and improve training stability.

Pointwise and depthwise convolutions within the block ensure proper dimension alignment. BN and GELU activation are applied after the SE module and convolutional operations:

$$Y = X + \text{Conv}_{\text{skip}}(X) \tag{10}$$

Following the S-Blocks, a sequence of MBCConv blocks constitutes the backbone of the model. Each MBCConv block comprises an expansion layer, a depthwise convolution, an SE module, and a projection layer, balancing computational efficiency with rich feature representation.

The classifier head consists of a pointwise convolution, followed by BN and HardSwish activation:

$$\text{HardSwish}(x) = x \times \frac{\text{ReLU6}(x + 3)}{6} \tag{11}$$

Global average pooling aggregates spatial features, which are then passed through two fully connected layers with dropout regularization.

The first fully connected layer reduces the feature dimension to 50, serving as an intermediate embedding representation.

The second fully connected layer maps these 50-dimensional features to the seven tea origin classes, Nanjian, Hejiazhai, Aluo Village, Reservoir No. 8, Daxinzai, Mangda, and Guangjingmeng, using a Softmax activation for final classification.

The overall classifier head architecture is illustrated in Fig. 8f.

In summary, Origin-Tea effectively integrates convolutional and attention mechanisms with SE modules and spatial attention blocks, resulting in a lightweight yet powerful model well-suited for resource-constrained tea origin classification tasks.

Experimental Conditions

All experiments in this study were conducted within a consistent hardware and software environment to ensure reproducibility and minimize variability due to computational differences. The hardware setup includes an NVIDIA GeForce RTX 4090 GPU and an NVIDIA RTX 4090D GPU, each

Table 10 | Training Hyperparameters Used in Sensitivity Experiments

Hyperparameter	Values
Batch size	30
Optimizer	AdamW
Weight decay	0.0001
Number of epochs	100, 200, 300
Learning rate	0.01, 0.001

equipped with 24 GB of memory, enabling multi-GPU parallel training for enhanced computational efficiency. The system is further supported by an Intel(R) Xeon(R) Silver 4214R CPU and 128 GB of RAM, providing ample resources for large-scale image processing and model training.

All deep learning experiments were implemented using PyTorch 2.2.1, with Python 3.9.7 as the programming environment. Key libraries utilized include NumPy 1.24.3 and Pandas 2.0.3 for data processing and model development. Model optimization employed the AdamW optimizer with a weight decay of 0.0001 to prevent overfitting and improve generalization. AdamW was chosen for its combination of adaptive learning rate benefits and decoupled weight decay, which enhances convergence stability and generalization, especially for small to medium-sized datasets typical in agricultural image analysis.

Training was conducted with a batch size of 30, balancing computational efficiency with gradient stability. This moderate batch size effectively utilizes GPU memory while maintaining sufficient stochasticity to improve generalization and reduce overfitting.

To assess the model's sensitivity to key hyperparameters, experiments varied the number of training epochs (100, 200, and 300) and learning rates (0.01 and 0.001). These values were selected based on preliminary trials to evaluate their effects on convergence, stability, and overall performance. The chosen range reflects common practice in deep learning and provides insight into model robustness under different training conditions.

The detailed hyperparameter settings used in this sensitivity analysis are summarized in Table 10, ensuring consistent and controlled comparisons across configurations.

Assessment indicators

To comprehensively evaluate the performance of the Origin-Tea model in the multi-class tea origin classification task, several widely used metrics were adopted: overall accuracy (OA), average accuracy (AA), Kappa coefficient (κ), P, recall (R), F1-score (F_1), Matthews correlation coefficient (MCC), and the number of model parameters. These metrics collectively capture classification accuracy, class-wise balance, and computational efficiency:

$$OA = \frac{\sum_{i=1}^C TP_i}{\sum_{i=1}^C (TP_i + FP_i + FN_i + TN_i)} \tag{12}$$

$$AA = \frac{1}{C} \sum_{i=1}^C \frac{TP_i + TN_i}{TP_i + FP_i + FN_i + TN_i} \tag{13}$$

$$\kappa = \frac{OA - Pe}{1 - Pe} \tag{14}$$

$$P = \frac{1}{C} \sum_{i=1}^C \frac{TP_i}{TP_i + FP_i} \tag{15}$$

$$R = \frac{1}{C} \sum_{i=1}^C \frac{TP_i}{TP_i + FN_i} \tag{16}$$

$$F_1 = \frac{1}{C} \sum_{i=1}^C \frac{2 \cdot TP_i}{2 \cdot TP_i + FP_i + FN_i} \tag{17}$$

$$MCC = \frac{\sum_{i,k} TP_i \cdot TN_i - FP_i \cdot FN_i}{\sqrt{(TP_i + FP_i)(TP_i + FN_i)(TN_i + FP_i)(TN_i + FN_i)}} \quad (18)$$

where C represents the total number of classes. For the i -th class, TP_i , FP_i , FN_i , and TN_i denote the counts of true positives, false positives, false negatives, and true negatives, respectively. Specifically, these terms correspond to the number of samples correctly predicted as class i (TP_i), incorrectly assigned to class i (FP_i), belonging to class i but predicted as others (FN_i), and correctly identified as not belonging to class i (TN_i). Finally, Pe indicates the expected accuracy by chance, calculated based on the confusion matrix.

Data availability

The dataset used in this study, including data and relevant experimental materials, has been publicly released and is openly accessible via a dedicated GitHub repository: <https://github.com/GuoquanPei/Origin-Tea>.

Received: 28 July 2025; Accepted: 24 December 2025;

Published online: 09 January 2026

References

- Shang, A., Li, J., Zhou, D. D., Gan, R. Y. & Li, H. B. Molecular mechanisms underlying health benefits of tea compounds. *Free Radic. Biol. Med.* **172**, 181–200 (2021).
- Zhang, L. et al. A study on origin traceability of white tea (White Peony) based on near-infrared spectroscopy and machine learning algorithms. *Foods* **12**, 499 (2023).
- Zhang, W., Kasun, L. C., Wang, Q. J., Zheng, Y. & Lin, Z. A review of machine learning for near-infrared spectroscopy. *Sensors* **22**, 9764 (2022).
- Greenacre, M. et al. Principal component analysis. *Nat. Rev. Methods Prim.* **2**, 100 (2022).
- Zhao, S., Zhang, B., Yang, J., Zhou, J. & Xu, Y. Linear discriminant analysis. *Nat. Rev. Methods Prim.* **4**, 70 (2024).
- Bansal, M., Goyal, A. & Choudhary, A. A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning. *Decis. Analytics J.* **3**, 100071 (2022).
- Kok, Z. H., Shariff, A. R. M., Alfatni, M. S. M. & Khairunniza-Bejo, S. Support vector machine in precision agriculture: a review. *Comput. Electron. Agric.* **191**, 106546 (2021).
- Wu, Y. et al. Blockchain-based internet of things: machine learning tea sensing trusted traceability system. *J. Sens.* **2022**, 8618230 (2022).
- Jin, G. et al. Rapid identification of the geographic origin of Taiping Houkui green tea using near-infrared spectroscopy combined with a variable selection method. *J. Sci. Food Agric.* **102**, 6123–6130 (2022).
- Wang, J., Lu, S., Wang, S. H. & Zhang, Y. D. A review on extreme learning machine. *Multimed. Tools Appl.* **81**, 41611–41660 (2022).
- Liu, W. et al. Authentication of the geographical origin of Guizhou green tea using stable isotope and mineral element signatures combined with chemometric analysis. *Food Control* **125**, 107954 (2021).
- Yun, J. et al. Use of headspace GC/MS combined with chemometric analysis to identify the geographic origins of black tea. *Food Chem.* **360**, 130033 (2021).
- Tardieu, F., Cabrera-Bosquet, L., Pridmore, T. & Bennett, M. Plant phenomics, from sensors to knowledge. *Curr. Biol.* **27**, R770–R783.736 (2017).
- McCouch, S. et al. Feeding the future. *Nature* **499**, 23–24 (2013).
- Wang, T. S., Liang, A. R. D., Ko, C. C. & Lin, J. H. The importance of region of origin and geographical labeling for tea consumers: The moderating effect of traditional tea processing method and tea prices. *Asia Pac. J. Mark. Logist.* **34**, 1158–1177 (2022).
- Seow, W. J. et al. Coffee, black tea, and green tea consumption in relation to plasma metabolites in an Asian population. *Mol. Nutr. Food Res.* **64**, 2000527 (2022).
- ZHANG, L. et al. Changes in physiological characteristics and cellular structures of tea plant leaves during overwintering at different altitudes in Xinyang. *J. Henan Agric. Sci.* **54**, 61 (2025).
- Bassiony, A. et al. Differential accumulation patterns of flavor compounds in Longjing 43 and Qunti fresh leaves and during processing responding to altitude changes. *Food Res. Int.* **187**, 114392 (2024).
- Yue, C., Wang, Z. & Yang, P. The effect of light on the key pigment compounds of photosensitive etiolated tea plant. *Botanical Stud.* **62**, 21 (2021).
- Mochida, K. et al. Computer vision-based phenotyping for improvement of plant productivity: a machine learning perspective. *GigaScience* **8**, giy153 (2019).
- Li, M., Frank, M. H. & Migicovsky, Z. Colourquant: A high-throughput technique to extract and quantify color phenotypes from plant images. In *High-Throughput Plant Phenotyping: Methods and Protocols* (eds Lorence, A. & Medina Jimenez, K.) 77–85 (Springer, 2022).
- Hodač, L. et al. Deep learning to capture leaf shape in plant images: Validation by geometric morphometrics. *Plant J.* **120**, 1343–1357 (2024).
- Xue, W. et al. CucumberAI: Cucumber fruit morphology identification system based on artificial intelligence. *Plant Phenomics* **6**, 0193 (2015).
- Ai, Y., Sun, C., Tie, J. & Cai, X. Research on recognition model of crop diseases and insect pests based on deep learning in harsh environments. *IEEE Access* **8**, 171686–171693 (2020).
- Zhu, X. et al. Identification of oil tea (*Camellia oleifera* C. Abel) cultivars using efficientNet-B4 CNN model with attention mechanism. *Forests* **13**, 1 (2021).
- Hu, J., Shen, L. & Sun, G. Squeeze-and-excitation networks. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* 7132–7141 (IEEE, 2018).
- Lin, P. et al. Rapid identification of the geographical origins of crops using laser-induced breakdown spectroscopy combined with transfer learning. *Spectrochim. Acta B: At. Spectrosc.* **206**, 106729 (2023).
- Li, Y. et al. Fingerprinting black tea: When spectroscopy meets machine learning a novel workflow for geographical origin identification. *Food Chem.* **438**, 138029 (2024).
- Peng, C. Y. et al. A comparative UHPLC-Q/TOF-MS-based metabolomics approach coupled with machine learning algorithms to differentiate Keemun black teas from narrow-geographic origins. *Food Res. Int.* **158**, 111512 (2022).
- Wu, Y. et al. Comparative analysis of volatiles difference of Yunnan sun-dried Pu-erh green tea from different tea mountains: Jingmai and Wuliang mountain by chemical fingerprint similarity combined with principal component analysis and cluster analysis. *Chem. Cent. J.* **10**, 1–11.770 (2016).
- Ran, W. et al. Comprehensive analysis of environmental factors on the quality of tea (*Camellia sinensis* var. *sinensis*) fresh leaves. *Sci. Horticulturae* **319**, 112177 (2023).
- Wei, G. & Zhou, R. Comparison of machine learning and deep learning models for evaluating suitable areas for premium teas in Yunnan, China. *PLoS ONE* **18**, e0282105 (2023).
- Widyaningrum, N. et al. The effect of altitude against total phenolic and Epigallocatechin Gallate (EGCG) content in green tea leaves. In *Proc. IConSSE FSM SWCU 2015* 34–40 (Satya Wacana Christian University, 2015).
- Wei, W. et al. Isolation, diversity, and antimicrobial and immunomodulatory activities of endophytic actinobacteria from tea cultivars Zijuan and Yunkang-10 (*Camellia sinensis* var. *assamica*). *Front. Microbiol.* **9**, 1304 (2018).

35. Li, N. et al. The influence of processing methods on polyphenol profiling of tea leaves from the same large-leaf cultivar (*Camellia sinensis* var. *assamica* cv. Yunkang-10): nontargeted/targeted polyphenomics and electronic sensory analysis. *Food Chem.* **460**, 140515 (2024).
36. Meng, J. et al. Tea bud and picking point detection based on deep learning. *Forests* **14**, 1188 (2023).
37. Liu, Y. et al. Comprehensive analysis and expression profiles of the AP2/ERF gene family during spring bud break in tea plant (*Camellia sinensis*). *BMC Plant Biol.* **23**, 206 (2023).
38. Zhang, C. et al. A comprehensive investigation of macro-composition and volatile compounds in spring-picked and autumn-picked white tea. *Foods* **11**, 3628 (2022).
39. Zhang, Z. et al. Identification of tea plant cultivars based on canopy images using deep learning methods. *Sci. Horticulturae* **339**, 113908 (2025).
40. Waldamichael, F. G., Debelee, T. G. & Ayano, Y. M. Coffee disease detection using a robust HSV color-based segmentation and transfer learning for use on smartphones. *Int. J. Intell. Syst.* **37**, 4967–4993 (2022).
41. Lingeswari, S. & Gomathi, P. and Kailasam SP. A Study of tomato fruit disease detection using RGB color thresholding and Kmeans clustering. *Int. J. Comput. Sci. Mob. Comput.* **10**, 51–9.798 (2021).
42. Maharana, K., Mondal, S. & Nemade, B. A review: Data pre-processing and data augmentation techniques. *Glob. Transit. Proc.* **3**, 91–99 (2022).
43. Su, D., Kong, H., Qiao, Y. & Sukkarieh, S. Data augmentation for deep learning based semantic segmentation and crop-weed classification in agricultural robotics. *Comput. Electron. Agric.* **190**, 106418 (2021).
44. Ichsan, C. & Rodiah, S. The best performing color space and machine learning regression algorithm for the accurate estimation of chromium (VI) and iron (III) in aqueous samples using low-cost and portable flatbed scanner colorimetry. *J. Iran. Chem. Soc.* **21**, 2335–2349 (2024).
45. Ropelewska, E., Rady, A. M. & Watson, N. J. Apricot stone classification using image analysis and machine learning. *Sustainability* **15**, 9259 (2023).
46. Howard, A. et al. Searching for mobilenetv3. In *Proc. IEEE/CVF Int. Conf. on Computer Vision. 2019* 1314–1324 (IEEE, 2019).
47. Dai, Z., Liu, H., Le, Q. V. & Tan, M. Coatnet: marrying convolution and attention for all data sizes. *Adv. neural Inf. Process. Syst.* **34**, 3965–3977 (2021).
48. Lee, M. Mathematical analysis and performance evaluation of the gelu activation function in deep learning. *J. Math.* **2023**, 4229924 (2023).
49. Pratiwi, H. et al. Sigmoid activation function in selecting the best model of artificial neural networks. *J. Phys.: Conf. Ser.* **1471**, 012010 (2020).
50. Liu, F., Ren, X., Zhang, Z., Sun, X. & Zou, Y. Rethinking skip connection with layer normalization. In *Proc. 28th Int. Conf. on Computational Linguistics (COLING)* 3587–3598 (2020).

Acknowledgements

The present study was funded by the Major Science and Technology Special Programs in Yunnan Province (202302AE09002001, 202402AE090014), the Basic Research Special Program in Yunnan Province (202401AS070006), and the Yunnan Tea Industry Technology Innovation Center (202505AK340010).

Author contributions

G. Pei and W. Wu conceived the study and led the writing of the manuscript. B. Zhou, B. Wang and W. Wu provided overall guidance and critical revisions. X. Qian, and W. Chen were responsible for data processing, conducting experiments, and preparing figures. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41538-025-00690-7>.

Correspondence and requests for materials should be addressed to Wendou Wu.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

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

© The Author(s) 2026