agronomy

MDPI

Article

Lightweight Multi-View Fusion Network for Non-Destructive Chlorophyll and Nitrogen Content Estimation in Tea Leaves Using Front and Back RGB Images

Wendou Wu [1], Guoquan Pei [1], Ziqiang Lu [2], Bing Zhou [2], Xueying Qian [1], Baijuan Wang [3] and Linnan Yang [1,*]

[1] College of Big Data, Yunnan Agricultural University, Kunming 650201, China; wuwd2004@126.com (W.W.); peiguoquan@outlook.com (G.P.); 15887903354@163.com (X.Q.)
[2] College of Science, Yunnan Agricultural University, Kunming 650201, China; myynau@163.com (Z.L.); bingzhoukm@126.com (B.Z.)
[3] College of Tea, Yunnan Agricultural University, Kunming 650201, China; wangbaijuan123@126.com
* Correspondence: 1998006@ynau.edu.cn

Abstract

Accurate estimation of chlorophyll and nitrogen content in tea leaves is essential for effective nutrient management. This study introduces a proof-of-concept dual-view RGB regression framework developed under controlled scanner conditions. Paired adaxial and abaxial images of Yun Kang 10 tea leaves were collected from four villages in Lincang, Yunnan, alongside corresponding soil and plant analyzer development (SPAD) and nitrogen measurements. A lightweight dual-input CoAtNet backbone with streamlined Bneck modules was designed, and three fusion strategies, Pre-fusion, Mid-fusion, and Late-fusion, were systematically compared. Ten-fold cross-validation revealed that Mid-fusion delivered the best performance (R^2 = 94.19% ± 1.75%, root mean square error ($RMSE$) = 3.84 ± 0.65, MAE = 3.00 ± 0.45) with only 1.92 M parameters, outperforming both the single-view baseline and other compact models. Transferability was further validated on a combined smartphone–scanner dataset, where the framework maintained robust accuracy. Overall, these findings demonstrate a compact and effective system for non-destructive biochemical trait estimation, providing a strong foundation for future adaptation to field conditions and broader crop applications.

Keywords: tea; chlorophyll; nitrogen; non-destructive measurement; deep learning

check for updates

Academic Editor: Jayme Garcia Arnal Barbedo

Received: 5 September 2025
Revised: 1 October 2025
Accepted: 4 October 2025
Published: 8 October 2025

Citation: Wu, W.; Pei, G.; Lu, Z.; Zhou, B.; Qian, X.; Wang, B.; Yang, L. Lightweight Multi-View Fusion Network for Non-Destructive Chlorophyll and Nitrogen Content Estimation in Tea Leaves Using Front and Back RGB Images. *Agronomy* **2025**, *15*, 2355. https://doi.org/10.3390/agronomy15102355

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Tea is one of the most widely consumed and economically valuable crops worldwide. Its quality is shaped by a range of physiological and biochemical factors, among which chlorophyll and nitrogen are key indicators for assessing both the nutritional status of tea plants and the quality of fresh tea leaves [1,2]. As the primary pigment driving photosynthesis, chlorophyll not only determines the green coloration and visual appeal of tea leaves but also influences their tenderness and freshness, traits that are highly valued in premium teas [3]. Nitrogen, an essential constituent of chlorophyll, proteins, and enzymes [4], plays a pivotal role in the biosynthesis of amino acids, alkaloids, and aromatic compounds, thereby directly shaping the taste, aroma, and nutritional value of tea [5–7].

In the tea industry, chlorophyll and nitrogen are widely recognized as critical biochemical markers of leaf quality [8]. Their concentrations not only reflect the physiological condition and metabolic activity of tea plants but also provide valuable insights for quality

grading, fertilization strategies, and harvest scheduling. Consequently, accurate, efficient, and non-destructive estimation of these indicators is essential for advancing modern tea production and enabling intelligent plantation management.

Given their importance, considerable research has been devoted to developing methods for monitoring and predicting chlorophyll and nitrogen contents in crops. Sonobe et al. [9] systematically evaluated different preprocessing techniques and 14 variable selection methods using hyperspectral reflectance data from 38 tea cultivars. Their results showed that a detrending preprocessing method combined with regression coefficients for variable selection yielded the best model performance, achieving a ratio of performance to deviation (RPD) of 2.60 and a root mean square error ($RMSE$) [10] of only 3.21 μg/cm^2. Barman et al. [11] proposed a method for predicting tea leaf chlorophyll content using images captured under natural light, applying multiple machine learning (ML) [12] algorithms: multiple linear regression (MLR) [13], artificial neural networks (ANN) [14], support vector regression (SVR) [15], and K-nearest neighbors (KNN) [16]. Tea leaf images were collected with a DSLR camera, color features were extracted, and these were fitted against actual chlorophyll values measured by a soil and plant analyzer development (SPAD) meter. The ANN model delivered the highest accuracy, outperforming the other approaches and demonstrating its potential as a rapid, low-cost alternative to SPAD-based chlorophyll estimation. Zeng et al. [17] estimated relative chlorophyll content (SPAD values) in tea canopies using multispectral UAV imagery. Two field experiments were conducted at the tea garden of South China Agricultural University, with rectangular plots ensuring precise alignment between multispectral images and ground samples. A multivariate linear stepwise regression (MLSR) model was developed, achieving high accuracy and stability R^2 of 0.865, $RMSE$ = 3.5481. This approach effectively captured the spatial distribution of chlorophyll and provided technical support for dynamic growth monitoring. Tang et al. [18] investigated microalgae chlorophyll prediction using LR and ANN based on RGB image components. The ANN model again outperformed LR, with an R^2 of 0.66 and $RMSE$ of 2.36, offering a faster, more efficient, and less costly alternative to conventional spectrophotometry for non-destructive estimation. Sharabiani et al. [19] applied visible–near-infrared (VIS/NIR) [20] spectroscopy combined with ANN and partial least squares regression (PLSR) to predict chlorophyll content in winter wheat leaves. Preprocessing techniques such as standard normal variate (SNV) and Savitzky–Golay smoothing were employed. The ANN model outperformed PLSR, achieving correlation coefficients of 0.97 and 0.92, respectively, with a minimum $RMSE$ of 0.73, confirming the superiority of nonlinear approaches for chlorophyll estimation. Brewer et al. [21] estimated maize chlorophyll content at different growth stages using multispectral UAV imagery with a random forest algorithm. Near-infrared and red-edge spectral bands, along with derived vegetation indices (VIs), were identified as key predictors. The random forest model performed best during the early reproductive and late vegetative stages, achieving $RMSE$ values of 40.4 and 39 μmol/m^{-2}, respectively. These findings highlight the potential of UAV remote sensing combined with ML for crop health monitoring in smallholder farming systems. Wang et al. [22] predicted nitrogen content in tea plants using UAV imagery integrated with ML and deep learning methods. Pixels were classified in the HSV color space, and two experimental strategies were designed to extract weighted color features paired with nitrogen data. Models including ordinary least squares (OLS) [23], XGBoost, RNN-LSTM [24], CNN, and ResNet were tested. HSV-based features consistently outperformed RGB features, with ResNet achieving the best results. This approach demonstrated both accuracy and practicality, supporting intelligent tea field management. Finally, Barzin et al. [25] predicted nitrogen concentration in corn leaves and grain yield using a handheld Crop Circle multispectral sensor combined with ML techniques. Twenty-five VIs, including SCCCI and RER-VI, were

evaluated under different nitrogen fertilizer treatments, with both spectral bands and VIs tested as model inputs. Among the models, SVR delivered the most accurate predictions of leaf nitrogen concentration, underscoring the potential of multispectral sensing and ML for precision agriculture.

Despite significant progress, current methods for estimating chlorophyll and nitrogen still face notable challenges and limitations. Hyperspectral and VIS/NIR spectroscopic techniques, while highly accurate and capable of multi-index estimation, rely on expensive equipment and complex procedures, making them impractical for rapid, low-cost deployment in small- to medium-scale tea plantations. In addition, although some studies have explored the use of RGB imagery combined with ML or deep learning for chlorophyll or nitrogen estimation [11,18], most have been limited to single-index modeling, thereby restricting their generalization capacity and practical applicability.

In practice, plantation managers require simple, cost-effective, and efficient tools to simultaneously assess chlorophyll and nitrogen levels in tea leaves, enabling more precise fertilization, pest and disease control, and quality management. This creates an urgent need for an accessible RGB image–based approach for multi-index (chlorophyll and nitrogen) prediction, one that achieves a balance between accuracy and computational efficiency to support intelligent plantation management.

To address this gap, we propose a proof-of-concept dual-view RGB image regression framework, developed and tested under controlled scanner conditions with standardized illumination and contour-based background removal. A paired dataset of adaxial (front) and abaxial (back) tea leaf images was collected alongside corresponding chlorophyll and nitrogen measurements. Using this dataset, we designed a lightweight dual-input CoAtNet backbone with streamlined Bneck modules and systematically evaluated three multi-view fusion strategies: Pre-fusion, Mid-fusion, and Late-fusion. Comprehensive experiments, including ten-fold cross validation and comparisons with advanced deep learning and classical regression models, demonstrated that the Mid-fusion strategy consistently achieved the optimal balance between prediction accuracy and model compactness.

The main contributions of this study can be summarized as follows:

- Dual-view regression framework: We propose a novel proof-of-concept dual-view RGB image regression framework developed under controlled conditions. By leveraging paired adaxial and abaxial tea leaf images, the framework effectively captures complementary phenotypic and biochemical traits, an aspect rarely addressed in previous research.
- Lightweight dual-input backbone: A streamlined dual-input backbone is introduced by integrating MobileNetV3-style Bneck modules into the CoAtNet architecture. This design achieves an 89% reduction in parameters while enhancing prediction accuracy, thereby improving computational efficiency and providing a foundation for future edge deployment.
- Fusion strategy comparison: Three multi-view fusion strategies, pre-fusion, Mid-fusion, and Late-fusion, are systematically designed and evaluated. Experimental results show that Mid-fusion delivers superior feature disentanglement and integration, resulting in consistent performance improvements for both chlorophyll and nitrogen estimation.

2. Materials and Methods

2.1. RGB Data Collection

In this study, tea leaves of the cultivar Yunkang No. 10 were selected as the research material due to their widespread cultivation and commercial importance in Yunnan Province. To ensure consistency in physiological maturity and reduce variability associated with

growth stage differences, data collection was carried out over a short, controlled period (16–19 April 2025). Sampling was conducted in four representative tea-producing villages in Lincang, Yunnan Province: Dele, Mangdan, Santa, and Xiaomenglong.

To ensure representativeness, each tea garden in the four villages was subdivided into multiple rectangular plots, each approximately 5 m in length and 0.8–1.0 m in width. Each plot typically contained two to three tea plants. Within each plot, three categories of leaves were sampled: ten standard harvesting units (one bud with two leaves), five young leaves, and five mature leaves. This stratified sampling strategy across developmental stages ensured that both intra-plant and inter-plant variation in chlorophyll and nitrogen contents were adequately represented. By incorporating leaves of different physiological ages and multiple plants per plot, the dataset captured a broad biochemical spectrum, providing a solid foundation for the development of robust prediction models.

To obtain high-quality RGB image data, all sampled tea leaves were scanned using a flatbed scanner (CanoScan LiDE 400, Canon Inc., Tokyo, Japan), which supports high-resolution, distortion-free imaging up to 4800 dpi. Each leaf was scanned on both the adaxial (front) and abaxial (back) surfaces to generate paired dual-view RGB images. The images were systematically saved and organized for subsequent modeling. An example of the scanned adaxial and abaxial images is shown in Figure 1.

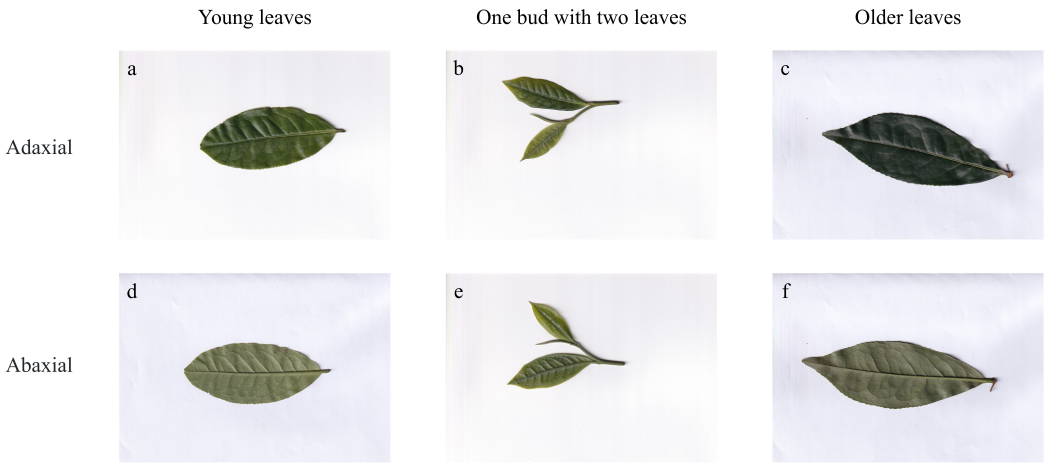


Figure 1. Scanned images of tea leaves showing different views and developmental stages: (**a**) Adaxial view of young leaves, (**b**) Adaxial view of one bud with two leaves, (**c**) Adaxial view of older leaves, (**d**) Abaxial view of young leaves, (**e**) Abaxial view of one bud with two leaves, (**f**) Abaxial view of older leaves. This figure highlights the variations in appearance across different physiological ages and leaf surfaces.

To minimize interference from background elements and ensure that the model focused exclusively on tea leaf phenotypic characteristics, a background removal procedure was applied to all RGB images. This preprocessing step isolated the leaf regions by eliminating non-leaf pixels, thereby reducing potential noise and bias in chlorophyll and nitrogen estimation.

Specifically, a contour-based image segmentation algorithm was applied. Each RGB image was first converted to grayscale and then binarized using an intensity threshold. The largest contour, corresponding to the tea leaf, was extracted using the cv2.findContours function, and a binary mask was generated. This mask was applied to the original image through a bitwise operation, preserving only the pixels within the leaf contour and effectively removing the background. The resulting leaf-only images were then saved in a separate directory for subsequent model training.

To validate the effectiveness of background removal and evaluate its impact on image statistics, intra-class variation was examined using color and illumination histograms before and after cropping. As shown in Figure 2, the original images with white backgrounds displayed a pronounced peak in the high-intensity region, which dominated the histogram distribution and introduced potential bias into model learning. Following contour-based segmentation, this peak disappeared, and the histograms reflected only leaf pixels, thereby providing a more accurate representation of color distributions. These results confirm that background removal effectively suppresses background-related artifacts while preserving intra-class variation relevant to chlorophyll and nitrogen estimation.

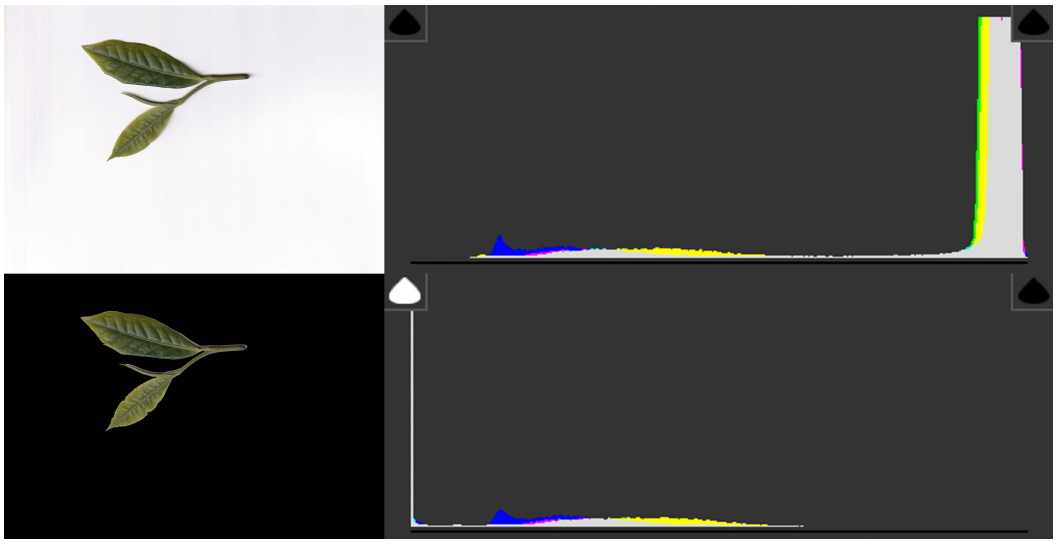


Figure 2. Comparison of color/illumination histograms before and after background removal. (**Top**) Original scanned image of tea leaves with a white background and its RGB histograms, showing a strong peak in the high-intensity region due to the white background. (**Bottom**) Background-removed image with corresponding histograms, where the background peak disappears and the distributions more faithfully reflect leaf pixel values. This confirms that contour-based segmentation effectively reduces background-induced bias and preserves intra-class variation in leaf color and illumination.

2.2. Acquisition of Chlorophyll and Nitrogen Data in Tea Leaves

In this study, chlorophyll and nitrogen contents of tea leaves were measured using a portable chlorophyll and nitrogen analyzer (Model: HM-YC, HENGMEI Intelligent Manufacturing). As shown in Figure 3, this device enables non-destructive and rapid estimation of chlorophyll levels through the SPAD index [26]. The SPAD value is derived from the transmittance ratio of light passing through the leaf at two wavelengths: 650 nm, which corresponds to chlorophyll absorption, and 940 nm, which serves as the reference. By comparing transmittance at these wavelengths, the analyzer estimates relative chlorophyll content. Because SPAD values are positively correlated with actual chlorophyll concentrations, they are widely used in agricultural research and practice as reliable indirect indicators of chlorophyll levels and overall plant health.

Given the well-established relationship between chlorophyll concentration and plant nitrogen availability, SPAD readings also provide a valuable proxy for estimating nitrogen content in tea leaves. In modern agricultural management, SPAD-based chlorophyll meters are widely employed to guide nitrogen fertilization, monitor plant nutritional status, and evaluate stress resistance across a variety of crops.



Figure 3. Portable chlorophyll and nitrogen analyzer (Model: HM-YC, Top Instrument Co., Ltd., Hangzhou, China) used for non-destructive measurement of leaf chlorophyll content. This instrument calculates the SPAD index based on light transmittance at 650 nm and 940 nm wavelengths, providing a rapid and reliable estimation of chlorophyll concentration and plant health.

To obtain accurate and representative measurements, a six-point sampling strategy was employed for each leaf. Specifically, the chlorophyll/nitrogen analyzer recorded values at six distinct positions: the leaf center, the apex, and four evenly distributed peripheral regions. As illustrated in Figure 4, the measurement points are indicated by red dots. The final chlorophyll and nitrogen contents for each leaf were calculated as the mean of these six readings, thereby minimizing local variability and enhancing data reliability.

For samples classified as "one bud with two leaves," measurements were taken only from the larger leaf. The smaller leaf was often too delicate or narrow to maintain stable contact with the measurement probe, making it unsuitable for reliable SPAD readings. This selective approach ensured data consistency while preserving the integrity of the sample set.

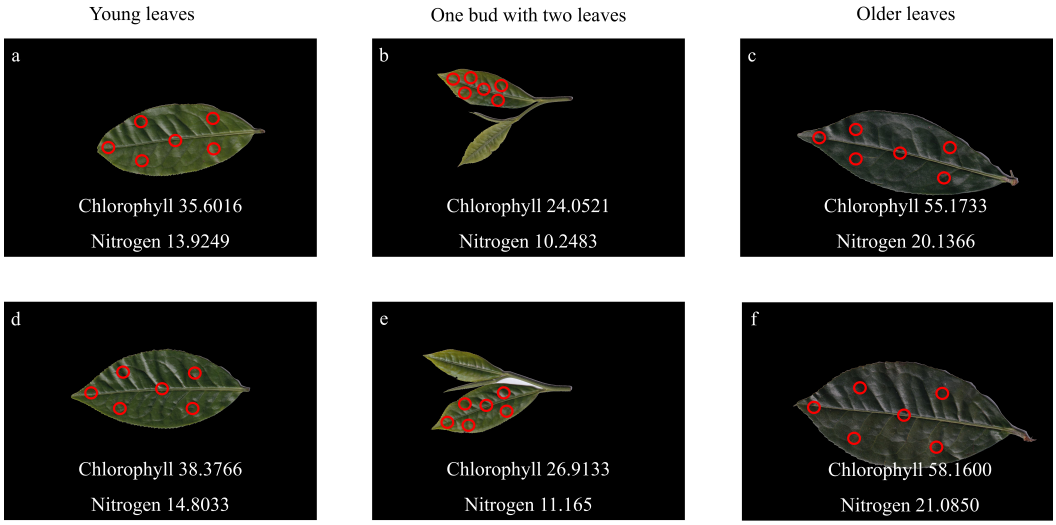


Figure 4. Six-point sampling method for chlorophyll and nitrogen content measurement in tea leaves at different developmental stages. The upper row (**a**–**c**) shows the adaxial leaf surfaces, and the lower row (**d**–**f**) shows the abaxial surfaces. Red circles indicate the sampling regions used for chlorophyll and nitrogen measurement. (**a**,**d**) Young leaves; (**b**,**e**) One bud with two leaves; (**c**,**f**) Older leaves.

To capture phenological variability across tea-growing regions, a stratified sampling strategy was employed. In total, 299 tea leaf samples were collected from four representative villages in Lincang, Yunnan Province: Dele, Mangdan, Santa, and Xiaomenglong. The samples were further categorized into three developmental stages: young leaves, one bud with two leaves, and older leaves. The detailed sample distribution is presented in Table 1. Notably, the "one bud with two leaves" category accounted for the largest proportion (149 samples), underscoring its economic significance in tea harvesting and processing. This stratified sampling approach ensured comprehensive data coverage, thereby enhancing the reliability and generalizability of the inversion modeling.

Table 1. Distribution of tea leaf samples collected from four regions in Lincang, Yunnan Province, categorized by developmental stages.

Region	Young Leaves	One Bud with Two Leaves	Older Leaves	Total
Dele	20	40	20	80
Mangdan	25	50	25	100
Santa	15	29	15	59
Xiaomenglong	15	30	15	60
Total	75	149	75	299

2.3. Model Architecture

2.3.1. CoAtNet Architecture

The CoAtNet architecture [27], shown in Figure 5, consists of five sequential stages (S0–S4), each progressively refining feature representations from low-level patterns to high-level concepts. The process begins with the stem stage (S0), where a 224 × 224 input image is passed through two consecutive 3 × 3 convolutional layers. The second convolution uses a stride of 2, reducing the spatial resolution to 112 × 112. This stage captures fundamental visual features and establishes the foundation for deeper feature extraction in subsequent stages.

Stage S1 consists of repeated L_1 blocks, each comprising a 1 × 1 convolution, a depthwise convolution with a 4× expansion ratio, and a final 1 × 1 convolution. This design strengthens local feature representation while maintaining computational efficiency. In Stage S2, the spatial resolution is further reduced to 28 × 28, and the network applies repeated L_2 blocks with increased channel capacity, enabling extraction of more complex and abstract features.

A major transition occurs at Stage S3, where the network shifts from local feature extraction to capturing global context. At this stage, the spatial resolution is reduced to 14 × 14, and the architecture incorporates relative attention mechanisms coupled with feed-forward networks (4× expansion ratio). This design allows the model to effectively capture long-range dependencies and complex relational interactions among features. Stage S4 extends this approach by further reducing the spatial dimensions to 7 × 7 and applying repeated L_4 blocks that combine relative attention with feed-forward networks, thereby enhancing high-level semantic representation and concept understanding.

Finally, the output from Stage S4 is passed through a global pooling layer, condensing the spatial feature maps into a compact vector representation. This vector is then fed into a fully connected layer to produce the final classification scores. Across all stages, CoAtNet seamlessly integrates convolutional and attention-based operations (Figure 5), enabling the model to handle both local detail and global context with high efficiency. By combining the strengths of convolution for spatial locality and attention for long-range dependency modeling, CoAtNet achieves superior accuracy and computational efficiency, making it a compelling architecture for advanced visual recognition tasks.

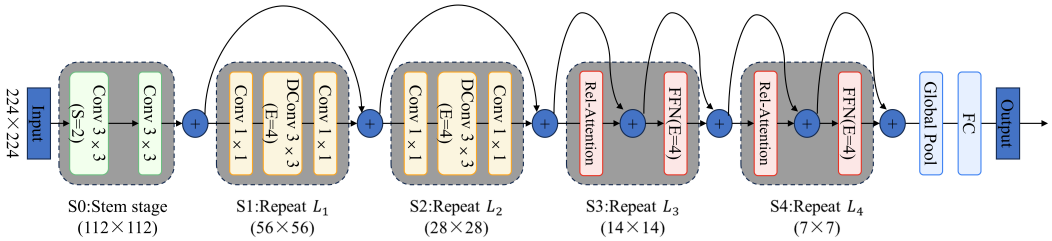


Figure 5. The diagram presents the complete network structure from the input layer to the output layer, encompassing convolutional stages (S0 to S1) and attention stages (S2 to S4). Notably, it demonstrates how the integration of depthwise convolution with relative attention mechanisms effectively captures local features and global contextual information, thereby achieving robust processing capabilities across various data scales.

2.3.2. Architecture of the Improved CoAtNet-Based Model

Although CoAtNet achieves strong performance across vision tasks through its hybrid integration of convolution and attention mechanisms, its high computational cost and large parameter count, particularly in the attention-heavy stages (S3 and S4), limit its suitability for real-time, resource-constrained applications such as mobile or edge-based tea leaf assessment. To overcome these challenges, we propose an improved CoAtNet-based architecture that significantly reduces model size while preserving predictive accuracy. As illustrated in Figure 6a, the redesigned framework retains the early convolutional stages (S0 and S1) for efficient low-level feature extraction, while replacing the later attention modules with lightweight Bneck blocks inspired by MobileNetV3 [28]. Each Bneck block incorporates an expansion layer, depthwise separable convolution, a Squeeze-and-Excitation (SE) module [29], and the h-swish activation function in place of GELU. This design, consistent with MobileNetV3 principles, achieves substantial computational efficiency and enables deployment on mobile and edge devices, as depicted in Figure 6b,c.

To enhance feature representation, the proposed model leverages adaxial and abaxial leaf surfaces as dual-view inputs [30]. Both views are first processed independently through shared S0 and S1 modules to extract modality-specific low-level features while preserving complementary phenotypic information. To integrate these features effectively, we design and evaluate three multi-view fusion strategies, each differing in the stage of integration. In the pre-fusion scheme, feature maps from the two views are concatenated immediately after S1 and unified via a 1×1 convolution before being passed into a shared sequence of Bneck blocks. In the mid-fusion scheme, each branch is further processed by its own Bneck module, and fusion occurs at the outputs of these modules, allowing deeper independent encoding before integration. In the late-fusion scheme, each branch is processed independently through the full pipeline, including its prediction head, and the outputs are subsequently combined using a learnable linear fusion layer. As illustrated in Figure 7, these strategies explicitly differ in their fusion points, after S1, after Bneck, or after the prediction heads.

In addition to the schematic comparison in Figure 7, the detailed feature map configurations at each fusion stage are summarized in Table 2. As shown, in the pre-fusion scheme, feature concatenation occurs immediately after S1, doubling the channel dimension to $192 \times 56 \times 56$. In the mid-fusion scheme, concatenation is performed at the Bneck output stage, producing a $192 \times 14 \times 14$ feature map. By contrast, the late-fusion strategy defers integration until the prediction head, where two high-level representations are merged to form an expanded feature space prior to regression.

Overall, the proposed improvements offer three key benefits. First, the integration of MobileNetV3-inspired lightweight modules substantially reduces computational cost and model complexity. Second, the use of dual-view leaf images captures complementary

phenotypic traits from adaxial and abaxial surfaces, thereby enriching feature diversity and enhancing the robustness of learned representations. Third, the introduction of flexible multi-stage fusion strategies enables systematic exploration of information integration at different levels of feature abstraction, providing insights into optimal fusion design.

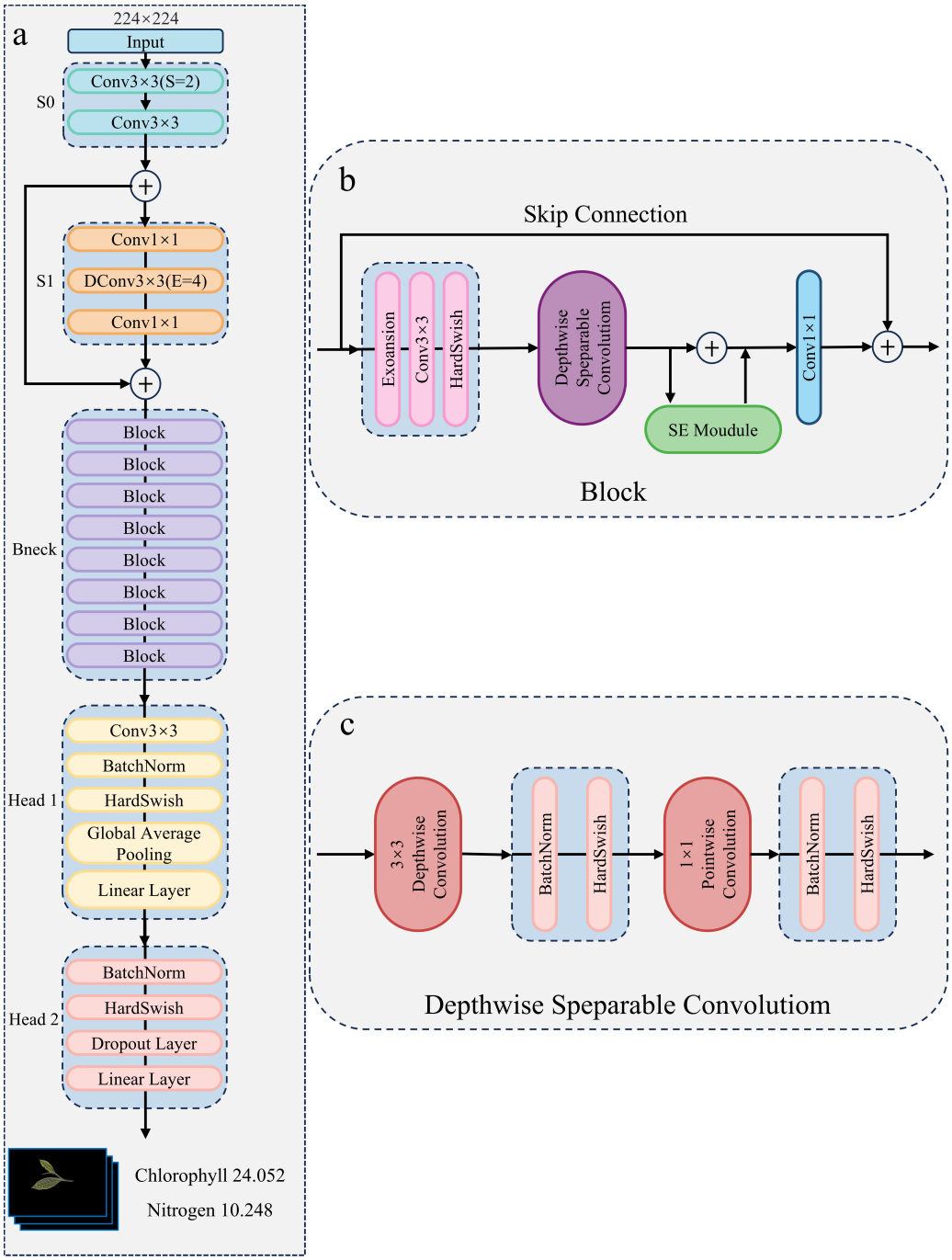


Figure 6. (**a**) Overall structure of the improved CoAtNet-based model. (**b**) Internal design of the MobileNetV3-inspired Bneck block. (**c**) Structure of the depthwise separable convolution used in Bneck blocks.

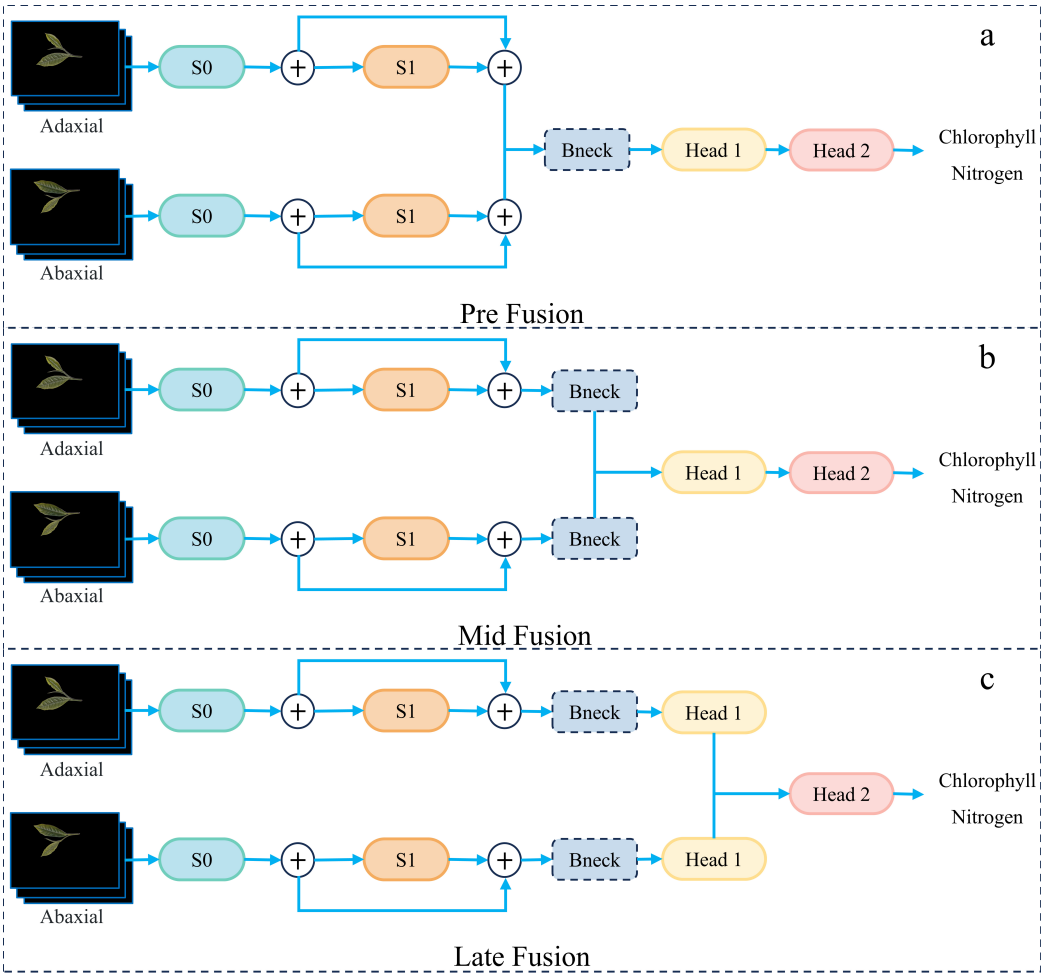


Figure 7. Comparison of the three fusion strategies: (**a**) Pre-fusion after S1, (**b**) Mid-fusion after Bneck, and (**c**) Late-fusion after the prediction heads.

Table 2. Feature map dimensions at key stages of Pre-fusion, Mid-fusion, and Late-fusion models.

Stage	Pre-Fusion	Mid-Fusion	Late-Fusion
Input	$3 \times 224 \times 224$	$3 \times 224 \times 224$	$3 \times 224 \times 224$
S0	$64 \times 112 \times 112$	$64 \times 112 \times 112$	$64 \times 112 \times 112$
S1	$96 \times 56 \times 56$ (after concat: $192 \times 56 \times 56$)	$96 \times 56 \times 56$	$96 \times 56 \times 56$
Bneck	$96 \times 14 \times 14$	$96 \times 14 \times 14$ (after concat: $192 \times 14 \times 14$)	$96 \times 14 \times 14$
Head 1	$576 \times 14 \times 14$	$576 \times 14 \times 14$	$576 \times 14 \times 14$ (after concat: $1152 \times 14 \times 14$)
Head 2	$1280 \rightarrow 1$	$1280 \rightarrow 1$	$1280 \rightarrow 1$

2.4. Experimental Configuration and Environment

All experiments were performed on a workstation equipped with an NVIDIA GeForce RTX 4090 GPU (24 GB), an Intel Xeon Silver 4214R CPU (2.40 GHz), and 128 GB of RAM, running PyTorch v2.3.1 with CUDA 12.6.

To ensure robust and unbiased model evaluation, we employed stratified ten-fold cross-validation. Stratification was performed according to three factors: tea-growing region (village), sampling area within each region, and leaf type (one bud with two leaves, young leaf, and older leaf). Each region was subdivided into multiple sampling areas, and all three leaf types were sampled within each area. During cross-validation, folds were constructed to maintain the proportional representation of all region–area–leaf-type strata, thereby minimizing the risk of distribution leakage. In each iteration, nine folds were used for training and one for testing, and the procedure was repeated ten times. Final performance metrics were reported as the mean and standard deviation across all folds.

Three fusion strategies for integrating adaxial and abaxial images were evaluated: Pre-fusion (integration after the stem module), Mid-fusion (integration after the per-branch Bneck blocks), and Late-fusion (integration at the prediction/output level).

Key experimental parameters are summarized in Table 3.

Table 3. Summary of experimental hyperparameters and settings.

Parameter	Description
Batch size	32
Number of epochs	100
Initial learning rate	0.0005
Optimizer	AdamW
Loss function	Mean Squared Error (MSE)
Cross-validation	ten-fold
Fusion strategies	Pre-fusion, Mid-fusion, Late-fusion

2.5. Evaluation Metrics

To quantitatively evaluate the predictive performance of the proposed dual-view fusion model, we employed three metrics: the coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error ($RMSE$). Together, these metrics provide complementary perspectives on model accuracy and robustness.

The coefficient of determination (R^2) quantifies the proportion of variance in the observed data that can be explained by the model and is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{1}$$

where y_i denotes the observed value, $\hat{y}_i$ the predicted value, $\bar{y}$ the mean of the observed values, and n the total number of samples. An R^2 value closer to 1 indicates stronger predictive performance.

The MAE quantifies the average magnitude of prediction errors, disregarding their direction, and is defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|. \tag{2}$$

MAE is robust to outliers and is theoretically optimal under a Laplacian error distribution, offering a measure of typical prediction error that is less sensitive to extreme deviations.

The $RMSE$ represents the square root of the mean of the squared differences between predicted and observed values and is defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}. \tag{3}$$

$RMSE$ penalizes larger errors more heavily than MAE and is statistically optimal when prediction errors follow a normal distribution, reflecting the model's sensitivity to extreme deviations.

By reporting R^2, MAE, and $RMSE$ together, this study provides a comprehensive and theoretically grounded evaluation of model performance. Specifically, R^2 captures the proportion of variance explained, MAE provides a robust measure of typical errors, and $RMSE$ emphasizes larger deviations. This combination ensures transparency and mitigates the limitations of relying on a single metric, in accordance with best practices in the statistical literature.

3. Results

3.1. Ablation Study

3.1.1. Effect of Fusion Strategies: Pre-, Mid-, and Late-Fusion

The Pre-fusion strategy integrates adaxial and abaxial image features immediately after the stem module, allowing the network to learn joint representations from an early stage. Experimental results for Pre-fusion across ten folds, including $RMSE$, MAE, and R^2, are summarized in Table 4.

As shown in Table 4, the Pre-fusion strategy achieved an average $RMSE$ of 4.51, MAE of 3.52, and R^2 of 91.84% across all folds, with moderate standard deviations (±0.86 for $RMSE$ and ±3.58 for R^2). These results indicate generally stable performance; however, the observed variability suggests that early fusion can occasionally introduce interference between modalities. Although most folds demonstrated strong predictive ability (with R^2 consistently above 88%), fusing features too early may compromise modality-specific information. This limitation motivated the investigation of alternative fusion strategies designed to better preserve and integrate complementary cues from adaxial and abaxial images.

Table 4. Ten-fold cross-validation results of the Pre-fusion strategy.

Fold	*RMSE*	*MAE*	R^2 (%)
1	4.16	3.24	91.72
2	5.94	4.55	85.39
3	5.43	4.16	88.46
4	4.37	3.54	91.30
5	5.87	4.54	88.63
6	4.52	3.54	92.25
7	3.41	2.49	95.82
8	3.76	2.86	95.22
9	4.68	3.80	93.25
10	2.98	2.43	96.32
Mean ± Std	4.51 ± 0.86	3.52 ± 0.60	91.84 ± 3.58

In the Mid-fusion strategy, adaxial and abaxial views are processed independently through their respective Bneck blocks following the stem module, with fusion performed at an intermediate stage. This design preserves view-specific features before integration, enabling a more balanced and stable combination.

Table 5 presents the ten-fold cross-validation results for the Mid-fusion configuration. Compared to Pre-fusion, Mid-fusion demonstrated markedly improved consistency and predictive accuracy: the average $RMSE$ and MAE decreased to 3.84 and 3.00, respectively, while the mean R^2 increased to 94.19%, with a smaller standard deviation of ±1.75. These results indicate enhanced stability across folds and stronger generalization. The findings suggest that performing fusion at a deeper, feature-rich stage allows the model to first capture view-specific structural and phenotypic patterns independently, and then integrate them synergistically, thereby improving overall biochemical prediction performance.

In the Late-fusion strategy, adaxial and abaxial images are processed independently through the entire network pipeline, including separate Bneck modules and prediction heads, with fusion performed only at the final output stage. This approach allows each view to be deeply and independently modeled, enabling the extraction of highly specialized features from each surface.

Table 5. Ten-fold cross-validation results of the Mid-fusion strategy.

Fold	*RMSE*	*MAE*	R^2 (%)
1	3.61	2.86	93.79
2	4.06	3.09	93.17
3	4.57	3.68	91.81
4	3.66	2.86	93.90
5	4.21	3.09	94.14
6	3.64	2.70	94.96
7	2.99	2.58	96.79
8	2.68	2.19	97.58
9	5.12	3.91	91.90
10	3.84	3.03	93.89
Mean ± Std	3.84 ± 0.65	3.00 ± 0.45	94.19 ± 1.75

The ten-fold cross-validation results for Late-fusion are summarized in Table 6. The model achieved an average *RMSE* of 4.00, *MAE* of 3.08, and R^2 of 93.75%. While Late-fusion outperformed Pre-fusion in both accuracy and stability, it performed slightly worse than Mid-fusion, which attained lower errors and higher R^2. One possible explanation is that, although Late-fusion enables fully independent feature extraction, the lack of interaction between adaxial and abaxial features until the final stage may limit the model's ability to leverage their complementarity during intermediate feature learning. Additionally, Late-fusion has the highest parameter count (2.60 M), resulting in increased computational cost, which is less favorable for deployment in resource-constrained agricultural environments.

Table 6. Ten-fold cross-validation results of the Late-fusion strategy.

Fold	*RMSE*	*MAE*	R^2 (%)
1	3.91	3.26	92.72
2	3.68	2.51	94.40
3	4.81	3.98	90.94
4	3.17	2.64	95.42
5	5.29	3.66	90.77
6	4.24	3.21	93.18
7	3.86	3.17	94.64
8	2.87	2.23	97.21
9	4.42	3.27	93.96
10	3.72	2.82	94.26
Mean ± Std	4.00 ± 0.64	3.08 ± 0.52	93.75 ± 1.88

Despite its slightly lower predictive accuracy compared to Mid-fusion, the consistent performance of Late-fusion across folds (indicated by low standard deviations) demonstrates that it is a robust strategy. This approach may be preferred in scenarios where maximizing per-view feature specialization is important or where decision-level fusion is desired for modular system design.

To comprehensively assess the impact of different fusion strategies, Table 7 summarizes the average metrics across ten-fold cross-validation for each approach, including *RMSE*, *MAE*, R^2, and total model parameters.

As shown, Mid-fusion achieved the best overall performance across all key metrics, with the lowest average *RMSE* (3.84), lowest *MAE* (3.00), and highest R^2 (94.19%), all accompanied by relatively small standard deviations. These results highlight the superior accuracy, robustness, and generalizability of the Mid-fusion model. Its effectiveness can be attributed to the architectural design: each modality (adaxial and abaxial views) is

encoded independently before fusion, preserving view-specific structural and biochemical information while enabling meaningful, synergistic integration at an intermediate semantic level.

In contrast, Pre-fusion showed weaker performance, with higher prediction errors and lower R^2 compared to the other strategies. This indicates that fusing the two views at such an early stage can lead to information interference or feature dilution, particularly when the visual characteristics of the adaxial and abaxial surfaces differ substantially.

Late-fusion offered a compromise between accuracy and specialization: it outperformed Pre-fusion while approaching the effectiveness of Mid-fusion. By processing adaxial and abaxial images through fully independent branches and combining outputs only at the decision level, it allows each pathway to fully specialize, which is beneficial for model interpretability and modularity. However, Late-fusion also incurs the highest computational cost (2.60M parameters), which may limit its suitability for deployment scenarios that require compact and efficient models.

To further illustrate variability and agreement across folds, additional visual analyses were conducted. Figure 8 presents fold-wise *RMSE*, *MAE*, and R^2 values along with their corresponding interquartile ranges (IQR). The Mid-fusion strategy achieves the lowest mean errors and the narrowest IQR, demonstrating consistent performance across folds. In contrast, Pre-fusion exhibits the widest IQR and larger fluctuations, while Late-fusion falls in between, with errors close to Mid-fusion but slightly less stable.

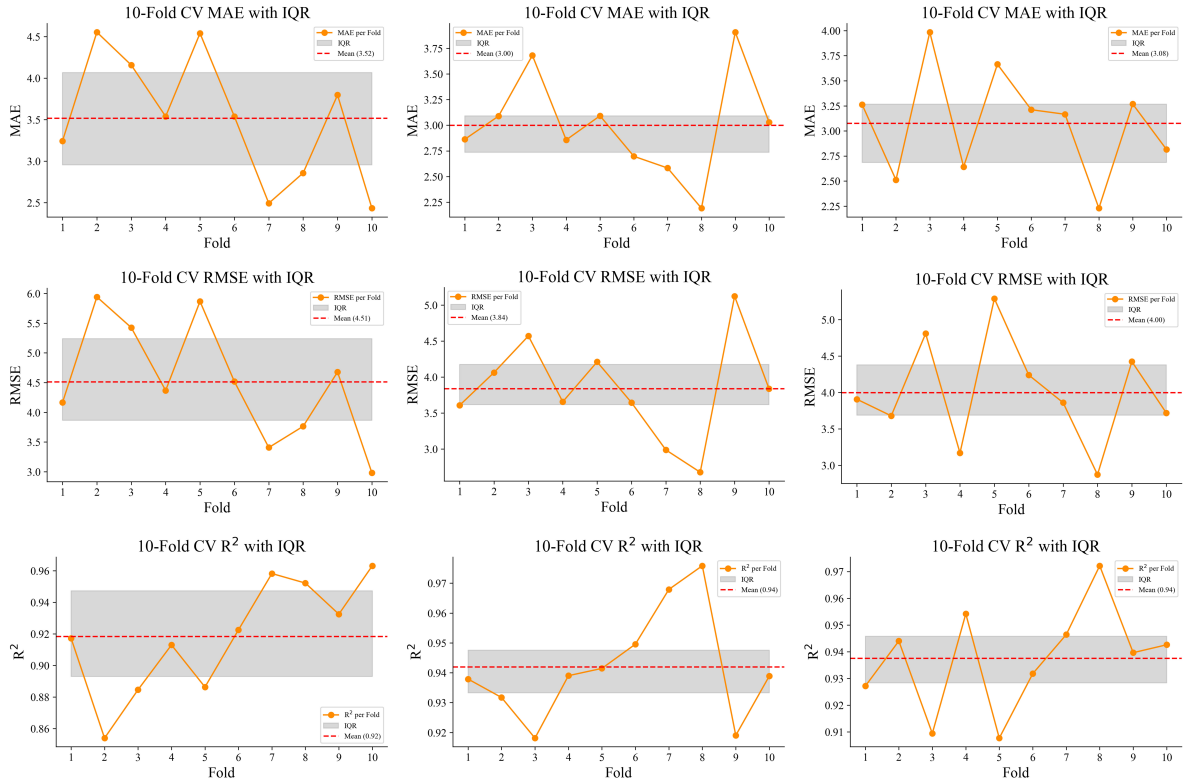
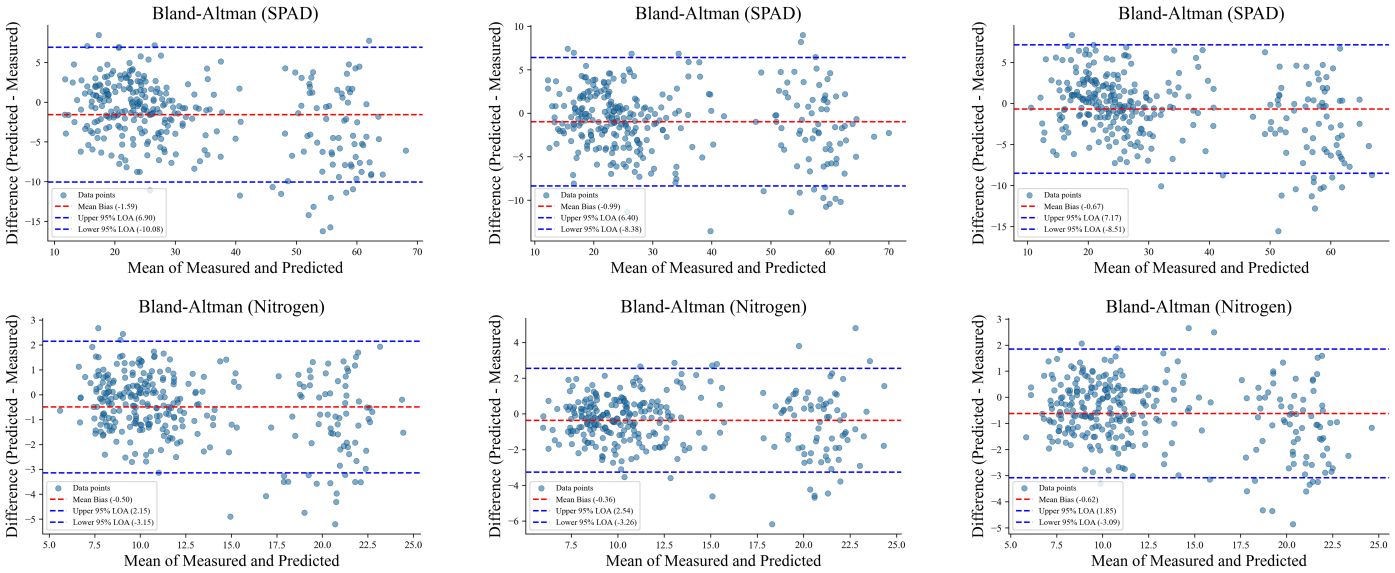


Figure 8. Distribution of ten-fold cross-validation metrics (*RMSE*, *MAE*, and R^2) for Pre-, Mid-, and Late-fusion strategies. The orange lines denote fold-wise metric values, the gray shaded areas represent the IQR, and the red dashed lines indicate mean values across folds. Mid-fusion demonstrates the narrowest IQR and the lowest error metrics, underscoring its stability and robustness.

Table 7. Performance comparison of different fusion strategies.

Strategy	*RMSE*	*MAE*	R^2 (%)	Parameters (M)
Pre-fusion	4.51 ± 0.86	3.52 ± 0.60	91.84 ± 3.58	1.88
Mid-fusion	3.84 ± 0.65	3.00 ± 0.45	94.19 ± 1.75	1.92
Late-fusion	4.00 ± 0.64	3.08 ± 0.52	93.75 ± 1.88	2.60

Complementary insights are provided by the Bland–Altman plots in Figure 9, which assess agreement between predicted and measured values for SPAD and nitrogen content. Mid-fusion shows the smallest mean bias and the tightest limits of agreement (LOA), indicating strong alignment with ground truth. Pre-fusion displays larger systematic deviations and wider LOA, whereas Late-fusion performs intermediately, reducing bias relative to Pre-fusion but showing slightly lower consistency than Mid-fusion.



(a) Pre-Fusion (b) Mid-Fusion (c) Late-Fusion

Figure 9. Bland–Altman analysis of predicted versus measured values for SPAD and nitrogen content under different fusion strategies. The red dashed line denotes the mean bias, while the blue dashed lines mark the 95% LOA. Mid-fusion shows the smallest bias and narrowest LOA, confirming its superior predictive agreement.

Taken together, the quantitative results (Tables 4–7) and visual analyses (Figures 8 and 9) consistently demonstrate the superiority of the Mid-fusion strategy. By preserving view-specific representations prior to integration, Mid-fusion delivers accurate, robust, and generalizable predictions of chlorophyll and nitrogen content.

3.1.2. Impact of Single-View vs. Dual-View Input

To evaluate the effectiveness of multi-view image inputs, ten-fold cross-validation experiments were conducted under three configurations: (1) adaxial images only, (2) abaxial images only, and (3) both views combined via the Mid-fusion strategy. Fold-wise evaluation results for each configuration are reported in Tables 8, 9 and 5, respectively.

Table 8. Ten-fold cross-validation results using only adaxial images.

Fold	*RMSE*	*MAE*	R^2 (%)
1	4.77	3.89	89.16
2	5.06	3.97	89.40
3	6.63	4.83	82.79
4	3.20	2.54	95.32
5	4.36	3.12	93.71
6	3.91	3.24	94.21
7	4.81	3.90	91.69
8	3.05	2.32	96.85
9	4.88	3.40	92.65
10	3.61	2.68	94.60
Mean ± Std	4.43 ± 0.96	3.39 ± 0.72	92.04 ± 3.88

The adaxial-only configuration achieved strong predictive performance, with a mean $RMSE$ of 4.43 (±0.96), MAE of 3.39 (±0.72), and R^2 of 92.04% (±3.8%). The relatively low standard deviations indicate consistent behavior across folds, underscoring the robustness of the adaxial view. This suggests that the adaxial surface provides reliable phenotypic cues, such as distinct coloration, venation, and textural features, that are particularly informative for chlorophyll and nitrogen estimation.

Table 9. Ten-fold cross-validation results using only abaxial images.

Fold	*RMSE*	*MAE*	R^2 (%)
1	4.72	3.85	89.36
2	3.29	2.52	95.52
3	6.45	5.34	83.68
4	4.32	3.39	91.49
5	4.83	3.88	92.29
6	5.51	4.59	88.47
7	4.70	4.00	92.05
8	4.16	3.13	94.15
9	4.94	3.99	92.46
10	5.00	4.26	89.62
Mean ± Std	4.79 ± 0.83	3.89 ± 0.78	90.91 ± 3.78

In contrast, the abaxial-only configuration demonstrated reduced predictive reliability compared to the adaxial-only setup, yielding a mean $RMSE$ of 4.79 (±0.83), MAE of 3.89 (±0.78), and R^2 of 90.91% (±3.78%). While the average performance remains reasonable, the greater variability across folds indicates less stable generalization. This suggests that the abaxial surface provides comparatively less consistent phenotypic cues, potentially due to variations in venation patterns, texture, or sensitivity to lighting, making it less robust when used in isolation.

As shown in Table 5, the Mid-fusion configuration, which integrates both adaxial and abaxial views at an intermediate stage, delivers the best overall performance among the three setups. The dual-view Mid-fusion approach achieves the lowest mean $RMSE$ (3.84 ± 0.65) and MAE (3.00 ± 0.45), as well as the highest mean R^2 (94.19% ± 1.75%), demonstrating strong consistency across folds and highlighting its robustness.

In comparison, the adaxial-only configuration performs reasonably well ($RMSE$ = 4.43 ± 0.96, MAE = 3.39 ± 0.72, R^2 = 92.04% ± 3.88%), confirming that the adaxial surface provides reliable visual cues for biochemical trait estimation. The abaxial-only configuration yields the weakest results ($RMSE$ = 4.79 ± 0.83, MAE = 3.89 ± 0.78, R^2 = 91.00% ± 0.03%), indi-

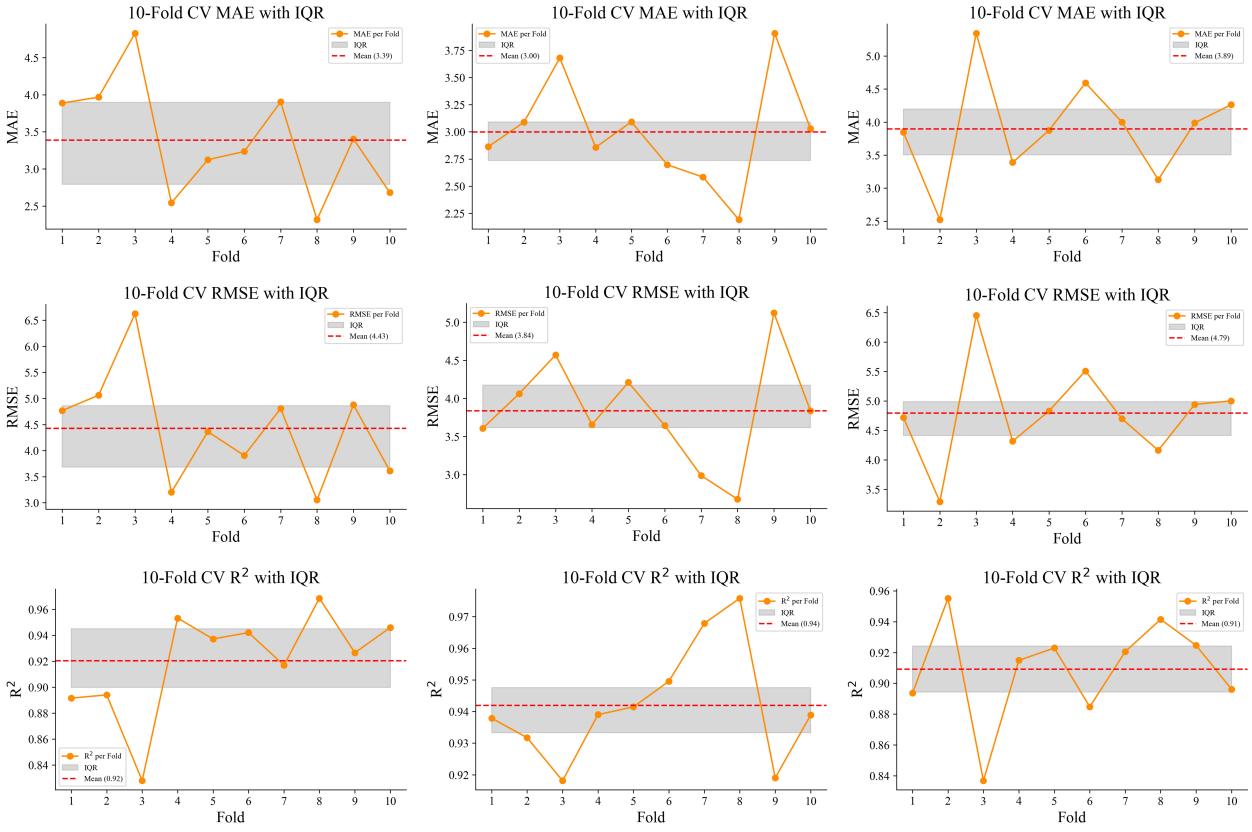
cating less consistent visual patterns. Collectively, these findings suggest that integrating both adaxial and abaxial views through mid-level fusion enables the model to capture complementary information not fully represented by either single view, while preserving view-specific features that contribute to robust and accurate predictions.

A comparative summary of the three configurations is provided in Table 10.

Table 10. Comparison of prediction performance under different input configurations.

Input Setting	*RMSE*	*MAE*	R^2 (%)
Adaxial Only	4.43 ± 0.96	3.39 ± 0.72	92.04 ± 3.88
Abaxial Only	4.79 ± 0.83	3.89 ± 0.78	90.91 ± 3.78
Dual-View	3.84 ± 0.65	3.00 ± 0.45	94.19 ± 1.75

To complement the quantitative evaluation, we conducted a visual assessment of model performance using fold-wise error distributions (*MAE* with IQR) and Bland–Altman plots for both biochemical traits (SPAD and nitrogen content). These analyses are summarized in Figures 10 and 11.

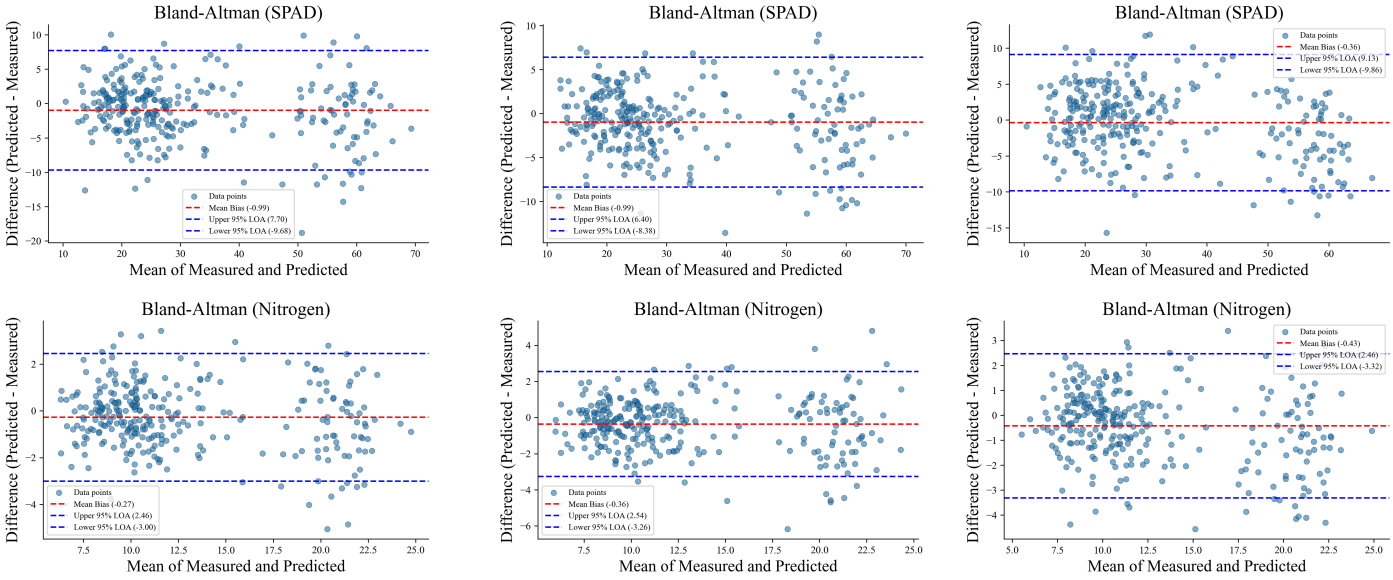


(a) Adaxial (b) Mid-Fusion (c) Abaxial

Figure 10. Ten-fold cross-validation *MAE* per fold with IQR for adaxial, Mid-fusion, and abaxial inputs. The orange lines indicate *MAE* per fold, gray shaded areas represent the IQR, and the red dashed lines denote the mean *MAE*.

As shown in Figure 10, the Mid-fusion configuration exhibits the narrowest IQR bands across folds, indicating more stable predictive performance compared with single-view inputs. Predictions based on adaxial-only images show moderate fold-to-fold variability, whereas abaxial-only results display the largest spread, consistent with the quantitative

metrics. Notably, the mean *MAE* for Mid-fusion is consistently lower than that of either single-view configuration, reinforcing the complementary value of integrating both leaf surfaces.



(a) Adaxial (b) Mid-Fusion (c) Abaxial

Figure 11. Bland-Altman plots comparing predicted versus measured values for SPAD (**top row**) and nitrogen content (**bottom row**) under adaxial, Mid-fusion, and abaxial inputs. The red dashed lines denote mean bias, and the blue dashed lines represent 95% LOA.

Bland–Altman analysis (Figure 11) further corroborates these findings. For SPAD prediction, the Mid-fusion model exhibits the smallest mean bias (≈ -0.36) and the tightest 95% LOA, indicating reduced systematic error and narrower prediction deviations. In contrast, adaxial- and abaxial-only predictions display larger biases and wider LOA, particularly for the abaxial input, reflecting increased variability. Similar patterns are observed for nitrogen content, where the Mid-fusion approach again demonstrates minimal bias and improved agreement with measured values. These results visually confirm that mid-level fusion effectively integrates complementary information from both leaf surfaces, enhancing both accuracy and consistency.

Overall, the visual analyses using fold-wise IQR and Bland–Altman plots reinforce the quantitative findings: dual-view Mid-fusion consistently outperforms single-view predictions, achieving both lower error rates and greater reliability across folds.

In summary, this ablation study highlights the value of dual-view input and confirms Mid-fusion as the optimal configuration. While the adaxial view alone provides reasonably accurate predictions, incorporating abaxial information further enhances accuracy and consistency. By effectively leveraging complementary spatial and phenotypic cues from both leaf surfaces, the Mid-fusion model achieves superior performance in estimating tea leaf chlorophyll and nitrogen contents.

3.1.3. Comparison with the Baseline CoAtNet Model

To assess the effectiveness of the proposed lightweight architecture, we conducted a comparative study against the baseline CoAtNet model, with both models employing the same mid-level fusion strategy for dual-view input. The CoAtNet baseline maintains its original five-stage design, using convolutional operations in the early layers (S0–S1) and transformer-based attention mechanisms in the later stages (S2–S4). In con-

trast, the proposed model replaces the attention-heavy blocks from stage S2 onward with MobileNetV3-style Bneck modules, resulting in a significantly more compact architecture.

Table 11 presents the ten-fold cross-validation results for the CoAtNet baseline, while Table 12 provides an overall comparison of performance metrics and parameter counts between the two models.

Table 11. Ten-fold cross-validation results of the baseline CoAtNet with Dual-View.

Fold	*RMSE*	*MAE*	R^2 (%)
1	4.03	3.27	92.27
2	5.46	3.99	87.67
3	4.04	3.14	93.60
4	4.91	3.46	89.01
5	6.27	5.12	87.03
6	4.27	3.53	93.06
7	3.73	2.91	94.98
8	3.72	3.09	95.32
9	4.78	3.97	92.96
10	3.83	3.05	93.92
Mean ± Std	4.50 ± 0.77	3.55 ± 0.57	91.98 ± 2.74

Table 12. Comparison between the baseline CoAtNet and our model.

Model	*RMSE*	*MAE*	R^2 (%)	Parameters (M)
CoAtNet	4.50 ± 0.77	3.55 ± 0.57	91.98 ± 2.74	17.02
Our model	3.84 ± 0.65	3.00 ± 0.45	94.19 ± 1.75	1.92

As shown in Table 12, the proposed lightweight model consistently outperforms the CoAtNet baseline across all evaluation metrics. It reduces *RMSE* from 4.50 to 3.84, decreases *MAE* from 3.55 to 3.00, and improves R^2 from 91.98% to 94.19%. Remarkably, these performance gains are achieved with only 1.92 million parameters, an 89% reduction compared to the baseline's 17.02 million parameters.

The substantial reduction in parameters is primarily due to architectural simplifications. Specifically, replacing the transformer-based attention blocks in stages S2–S4 with lightweight MobileNetV3-style Bneck modules significantly decreases computational overhead. The use of depthwise separable convolutions, combined with channel attention via SE modules [29], preserves essential representational capacity while eliminating redundant parameters. Furthermore, the dual-branch stem design allows adaxial and abaxial views to share early-stage processing (S0 and S1), enhancing parameter efficiency without compromising predictive performance.

Despite its compact design, the proposed model not only maintains competitive performance but also surpasses the CoAtNet baseline. This improvement is attributable to two key factors: (1) the task-specific architecture employs a dual-input design that separately encodes adaxial and abaxial views, capturing complementary phenotypic cues relevant to chlorophyll and nitrogen estimation while avoiding unnecessary complexity; and (2) the mid-level fusion strategy allows independent representation learning for each view prior to integration, preserving modality-specific features and enabling more effective cross-view information exchange than early fusion schemes. The enhanced R^2 and reduced fold-to-fold variance further indicate improved generalizability and robustness across data splits.

In summary, these results demonstrate that the proposed model achieves a favorable trade-off between predictive accuracy and computational efficiency, effectively ad-

dressing the limitations of the original CoAtNet backbone and making it well-suited for deployment in resource-constrained environments, such as edge devices or mobile phenotyping systems.

3.2. Performance Comparison with Advanced Models

The detailed quantitative results for all models under adaxial-only, abaxial-only, and dual-view Mid-fusion configurations are summarized in Table 13. This table reports mean ± standard deviation values for $RMSE$, MAE, R^2, and parameter counts, providing a comprehensive basis for cross-model comparison.

To contextualize the effectiveness of the proposed model, its performance was compared against a diverse set of advanced deep learning and classical regression models, including CoAtNet [27], GhostNet [31], MobileNetV3 [28], ResMLP [32], ResNet34 [33], StartNet [34], TinyViT [35], as well as KNN [36], MLR [37], and SVR [38]. These models encompass lightweight convolutional architectures, transformer-based designs, and classical methods, enabling a thorough evaluation of our model's robustness and efficiency across different input configurations (adaxial-only, abaxial-only, and dual-view Mid-fusion).

Across all configurations, the proposed model consistently achieves superior performance. In the abaxial-only setting, it attains $RMSE$ = 4.79 (±0.83), MAE = 3.89 (±0.78), and R^2 = 90.91% (±3.78%), outperforming most competitors including CoAtNet ($RMSE$ = 5.39, R^2 = 88.22%), ResMLP ($RMSE$ = 6.91, R^2 = 81.12%), and classical methods such as KNN, MLR, and SVR. Lightweight networks like MobileNetV3 and GhostNet exhibit higher $RMSE$ (>7.3%) and lower R^2 (<77%), highlighting that efficiency alone does not ensure accuracy without architectural adaptation to leaf phenotypic characteristics.

In the adaxial-only configuration, the proposed model achieves $RMSE$ = 4.43 (±0.96), MAE = 3.39 (±0.72), and R^2 = 92.04% (±3.88%), surpassing CoAtNet in MAE and demonstrating competitive overall accuracy (CoAtNet: $RMSE$ = 4.20, R^2 = 93.03%). TinyViT also performs well ($RMSE$ = 4.50, R^2 = 91.91%), whereas MobileNetV3, GhostNet, and StartNet show considerably poorer results ($RMSE > 7.3$, $R^2 < 70\%$). Notably, deeper conventional networks such as ResNet34 display substantial variability ($RMSE$ = 6.01, R^2 = 74.31%) under adaxial-only input, reflecting challenges in generalization for standard deep networks when processing single-view leaf images.

The most pronounced improvements are observed in the dual-view Mid-fusion configuration. Our model achieves $RMSE$ = 3.84 (±0.65), MAE = 3.00 (±0.45), and R^2 = 94.19% (±1.75%), outperforming CoAtNet ($RMSE$ = 4.50, R^2 = 91.98%) by 6.7% in $RMSE$ while reducing the parameter count from 17.02M to 1.92M. Other strong backbones, such as TinyViT and ResNet34, also benefit from Mid-fusion, but their best R^2 scores (91.72% and 80.82%, respectively) remain 2–13 percentage points lower than our model. The consistent advantage of Mid-fusion across all architectures underscores the importance of integrating adaxial and abaxial views at an intermediate stage, preserving view-specific information while capturing complementary features.

Analysis of fusion strategies across all models confirms that Mid-fusion consistently outperforms both Pre-fusion and Late-fusion in terms of accuracy and stability. Pre-fusion often suffers from feature entanglement, whereas Late-fusion may overlook cross-view interactions. Intermediate integration in Mid-fusion effectively balances these trade-offs, resulting in robust performance even for smaller backbones.

Parameter efficiency is another key strength of our model. With only 1.92M parameters, it uses less than one-ninth of CoAtNet and one-eighth of ResMLP, yet surpasses both in predictive accuracy. This efficiency is achieved through MobileNetV3-style Bneck modules, depthwise separable convolutions, and SE blocks, which preserve representational capacity without increasing complexity. The dual-branch stem design allows independent pro-

cessing of adaxial and abaxial views, further reducing computational cost while retaining critical phenotypic cues.

Table 13. Performance comparison of different models under adaxial, abaxial, and dual-view configurations.

Model	*RMSE*	*MAE*	R^2 (%)	Parameters (M)
		Abaxial		
ResMLP	6.91 ± 1.16	5.43 ± 0.84	81.12 ± 7.21	15.35
MobileNetV3	7.82 ± 2.37	6.23 ± 2.02	73.76 ± 18.54	1.68
TinyViT	4.59 ± 0.59	3.71 ± 0.54	**91.74 ± 2.40**	5.07
StartNet	6.81 ± 1.98	5.88 ± 2.02	80.16 ± 13.41	2.68
GhostNet	7.38 ± 3.27	5.80 ± 2.76	76.03 ± 22.13	6.86
ResNet34	6.04 ± 4.46	5.05 ± 4.13	80.82 ± 32.50	21.29
CoAtNet	5.39 ± 1.24	4.16 ± 0.99	88.22 ± 4.82	17.03
KNN	7.811 ± 1.140	5.969 ± 0.768	76.1 ± 4.1	-
MLR	8.896 ± 1.540	6.635 ± 1.027	68.0 ± 9.6	-
SVR	10.032 ± 2.054	7.491 ± 1.489	59.4 ± 13.7	-
Ours	4.79 ± 0.83	3.89 ± 0.78	90.91 ± 3.78	2.79
		Adaxial		
ResMLP	5.35 ± 0.92	4.15 ± 0.71	88.61 ± 4.57	15.35
MobileNetV3	9.28 ± 5.86	7.35 ± 5.08	58.12 ± 53.33	1.68
TinyViT	4.50 ± 0.56	3.54 ± 0.54	91.91 ± 3.02	5.07
StartNet	6.42 ± 2.15	5.54 ± 2.02	82.75 ± 11.46	2.68
GhostNet	8.78 ± 3.12	7.34 ± 2.98	66.58 ± 22.96	6.86
ResNet34	6.01 ± 5.73	4.99 ± 5.28	74.31 ± 57.81	21.29
CoAtNet	4.20 ± 0.71	3.12 ± 0.43	93.03 ± 2.55	17.03
KNN	8.458 ± 0.850	6.495 ± 0.912	71.2 ± 6.7	-
MLR	6.408 ± 1.229	4.821 ± 0.973	82.9 ± 7.1	-
SVR	9.875 ± 1.215	7.403 ± 1.079	60.4 ± 11.1	-
Ours	4.43 ± 0.96	3.39 ± 0.72	92.04 ± 3.88	2.79
		Fusion / Dual-View		
ResMLP	6.60 ± 1.67	5.04 ± 1.08	81.64 ± 11.87	15.26
MobileNetV3	6.47 ± 1.88	5.01 ± 1.26	82.10 ± 12.60	2.41
TinyViT	4.55 ± 0.71	3.54 ± 0.50	91.72 ± 2.13	5.73
StartNet	6.81 ± 1.98	5.88 ± 2.02	80.16 ± 13.41	2.68
GhostNet	5.94 ± 1.21	4.85 ± 1.16	85.70 ± 6.75	3.16
ResNet34	6.04 ± 4.46	5.05 ± 4.13	80.82 ± 32.50	21.29
CoAtNet	4.50 ± 0.84	3.55 ± 0.66	91.98 ± 2.99	17.02
KNN	8.118 ± 1.509	6.225 ± 1.094	73.2 ± 9.1	-
MLR	6.916 ± 0.842	5.184 ± 0.576	80.8 ± 4.3	-
SVR	9.988 ± 1.557	7.628 ± 1.315	59.7 ± 12.0	-
Ours	3.84 ± 0.65	3.00 ± 0.45	94.19 ± 1.75	1.92

3.3. Validation of Transferability to Mixed Acquisition Images

To assess the transferability of the proposed dual-view RGB framework beyond the original scanner-acquired dataset, a proof-of-concept experiment was conducted using a combined dataset of 638 adaxial and abaxial images. Half of the images were captured using a smartphone under natural illumination, while the other half were obtained from a conventional flatbed scanner. These images, collected from tea samples in Dali and Kunming tea gardens, were paired into 319 dual-view samples. During smartphone acquisition, a white sheet was placed beneath the leaves and the samples were enclosed in a box to minimize external interference. Automatic white balance with white preference was applied to ensure consistent color representation across images.

Figure 12 presents representative examples of smartphone-acquired images alongside their background-removed versions, demonstrating the effectiveness of the contour-based preprocessing pipeline. Scanner-acquired images were processed using the standard background removal procedure described previously.

The paired images were processed using the proposed dual-input model with the Mid-fusion strategy. On the held-out test set, the model achieved $RMSE$ = 4.48, MAE = 3.49, and R^2 = 82.42%. These results demonstrate that, despite variations in acquisition devices and lighting conditions, the model maintains reasonable predictive performance. While there is an expected decrease compared to the controlled scanner-only measurements, the findings support the framework's transferability across both smartphone- and scanner-acquired images, highlighting its potential applicability in more practical, field-based scenarios.



(a) Adaxial



(b) Abaxial

Figure 12. Example of smartphone-acquired tea leaf images. A white paper was placed beneath the leaves to reduce background interference, and images were processed via contour-based background removal.

Figure 13 illustrates the model's predictive performance on the mixed-device test set. Most points cluster closely around the red reference line, indicating that the dual-view Mid-fusion model effectively captures the underlying relationships between leaf color features and biochemical contents across both smartphone and scanner images. Minor deviations are observed at higher SPAD and nitrogen values, likely due to variations in illumination and imaging conditions. Overall, these results demonstrate robust generalization of the model to images acquired from different devices, confirming its potential for non-destructive estimation of tea leaf chlorophyll and nitrogen content in practical, field-based applications.

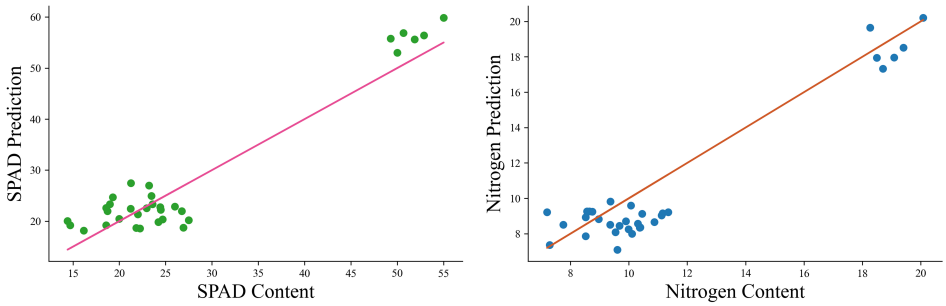


Figure 13. Prediction results of the dual-view RGB model on the mixed-device test set. The red line represents the perfect fit (true values), while scatter points correspond to model predictions for SPAD and nitrogen contents.

4. Discussion

This study presents a dual-view image–based deep learning framework for predicting chlorophyll and nitrogen content in tea leaves. By incorporating both adaxial and abaxial leaf surfaces, the method captures complementary phenotypic cues that are inaccessible through single-view observations. The model combines a lightweight hybrid architecture with an effective Mid-fusion strategy, achieving enhanced prediction accuracy while maintaining low computational cost. Several key insights emerged from the experimental results:

First, the dual-view input strategy consistently enhanced prediction performance compared with single-view inputs across all tested models. In particular, the Mid-fusion approach, where intermediate features from each view are independently encoded before fusion, achieved the highest accuracy, demonstrating that deeper cross-view interactions enable the model to extract semantically rich, task-relevant representations. These results underscore the complementarity of adaxial and abaxial surfaces in reflecting chlorophyll and nitrogen traits, highlighting the value of multi-view modeling in plant phenotyping.

Second, our improved CoAtNet-based model with Mid-fusion outperformed all compared models under ten-fold cross-validation. This superior performance can be attributed to its hybrid design, which combines convolutional inductive biases with transformer-style attention, along with MobileNetV3-inspired bottleneck blocks that reduce parameter count without sacrificing representational capacity. With only 1.92 million parameters, the model delivers highly competitive results.

Third, the choice of fusion strategy plays a critical role in multi-view modeling. Comparison of Pre-fusion, Mid-fusion, and Late-fusion schemes revealed that early concatenation (Pre-fusion) or late merging at the decision level (Late-fusion) compromises prediction accuracy. Pre-fusion may introduce noise and hinder learning of distinct spatial features from each view, while Late-fusion does not fully capture shared latent patterns. In contrast, Mid-fusion aligns and integrates higher-level features, facilitating more expressive and generalizable cross-view representations.

Furthermore, the model demonstrated robustness across leaf categories (young, mature, and old) and samples from multiple tea-producing regions. Consistent performance despite morphological and regional variability confirms its strong generalization ability. It should be noted, however, that the RGB data were collected over a short temporal window (16–19 April 2025) to minimize variability in leaf physiological maturity and environmental factors such as light, temperature, and humidity. While this controlled sampling enhances internal consistency and reduces confounding effects, it may limit generalization across different seasonal stages.

Compared with existing studies, our method advances leaf trait prediction. Table 14 summarizes representative works using single-view RGB, hyperspectral, UAV, and Vis/NIR

modalities alongside our dual-view Mid-fusion model. Relative to single-view RGB approaches (e.g., Tang et al., 2023 [39], R^2 = 66%, $RMSE$ = 2.36), the dual-view strategy substantially improves R^2 (94.19%) and reduces prediction errors. While hyperspectral or Vis/NIR methods (e.g., Azadnia et al., 2023 [40], R^2 = 97.80%) may achieve slightly higher accuracy under controlled conditions, they require specialized equipment and preprocessing. In contrast, our dual-view RGB approach achieves competitive accuracy while capturing complementary structural and biochemical information, highlighting the measurable benefits of multi-view modeling. At the same time, the results suggest room for further improvement, particularly in closing the gap with spectroscopy-based methods and extending predictions to more variable growth stages.

Despite these strengths, certain limitations should be acknowledged. A key challenge is the potential impact of concept drift, where dynamic environmental factors such as soil conditions, lighting, or climate variability may influence model accuracy over time. Agricultural systems are inherently heterogeneous and evolving, and reliance on RGB imagery alone may limit robustness and scalability. To address this, future research could explore multi-source data fusion strategies. For example, deep learning features extracted from RGB images could be combined with environmental information, such as soil moisture, temperature, or meteorological data, to enhance prediction consistency across diverse conditions.

In summary, the proposed dual-view deep learning framework, particularly the Mid-fusion variant of the improved CoAtNet model, provides an effective, interpretable, and computationally efficient solution for estimating leaf biochemical traits. By enabling fine-grained visual representation learning across both adaxial and abaxial surfaces, our approach advances plant phenotyping research and offers a generalizable method applicable to other crops and traits requiring multi-view analysis.

Table 14. Representative studies on leaf trait prediction with single- or dual-view inputs and different imaging modalities. $RMSE$, MAE, and R^2 values are provided to situate the performance of the proposed dual-view Mid-fusion model.

Study	Modality	$RMSE$	MAE	R^2 (%)
Tang et al., 2023 [39]	RGB single-view (microalgae)	2.36	-	66.00
Zhang et al., 2023 [41]	Hyperspectral UAV (Chinese cabbage)	-	-	71.0
Hu et al., 2024 [42]	Hyperspectral (rice N content)	0.08	3.89	-
Azadnia et al., 2023 [40]	Vis/NIR spectroscopy (apple)	0.01	-	97.80
Barman and Choudhury, 2022 [43]	Smartphone RGB (citrus)	0.26	0.07	94.00
Ours	RGB dual-view (tea)	3.84	3.00	94.19

5. Conclusions

In this study, we proposed a dual-view RGB imaging framework combined with a lightweight deep learning model for non-destructive estimation of chlorophyll and nitrogen contents in tea leaves. A paired adaxial–abaxial dataset of 299 image pairs was collected from four tea-producing villages in Lincang, Yunnan, under controlled scanner-based illumination, with corresponding biochemical measurements obtained via a rapid detector.

Ten-fold cross-validation demonstrated that the Mid-fusion strategy achieved the best predictive performance, with R^2 = 94.19% ± 1.75%, $RMSE$ = 3.84 ± 0.65, and MAE = 3.00 ± 0.45, outperforming the single-view CoAtNet baseline (R^2 = 93.03% ± 2.55%, $RMSE$ = 4.20 ± 0.71, MAE = 3.12 ± 0.43) while reducing the parameter count from 17.03 M to 1.92 M. Ablation analyses confirmed that incorporating both adaxial and abaxial views significantly enhances predictive reliability and robustness.

The main contributions of this work are threefold: (i) the construction of a controlled-condition, paired-view dataset for tea leaf biochemical estimation; (ii) the design of a

lightweight dual-input CoAtNet variant that balances predictive accuracy with model compactness; and (iii) the empirical validation of Mid-fusion as an effective multi-view integration strategy. It should be noted that the current findings are limited to a single cultivar, four villages, and a short acquisition period under scanner-based conditions. Future work will focus on domain adaptation and generalization to variable field environments, as well as extending the framework to other crops and traits requiring multi-view analysis.

Author Contributions: Funding acquisition, W.W.; Methodology, W.W., G.P. and B.Z.; Validation, Z.L.; Visualization, X.Q.; Writing—original draft, G.P.; Writing—review & editing, B.W. and L.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Major Science and Technology Special Programs in Yunnan Province (grant number 202302AE09002001), and the Basic Research Special Program in Yunnan Province (grant number 202401AS070006).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

References

1. Elango, T.; Jeyaraj, A.; Dayalan, H.; Arul, S.; Govindasamy, R.; Prathap, K.; Li, X. Influence of shading intensity on chlorophyll, carotenoid and metabolites biosynthesis to improve the quality of green tea: A review. *Energy Nexus* **2023**, *12*, 100241. [CrossRef]
2. Wang, Y.; Hong, L.; Wang, Y.; Yang, Y.; Lin, L.; Ye, J.; Jia, X.; Wang, H. Effects of soil nitrogen and pH in tea plantation soil on yield and quality of tea leaves. *Allelopath. J.* **2022**, *55*, 51–60. [CrossRef]
3. Wang, P.; Grimm, B. Connecting chlorophyll metabolism with accumulation of the photosynthetic apparatus. *Trends Plant Sci.* **2021**, *26*, 484–495. [CrossRef]
4. Huang, Y.; Wang, H.; Zhu, Y.; Huang, X.; Li, S.; Wu, X.; Zhao, Y.; Bao, Z.; Qin, L.; Jin, Y.; et al. THP9 enhances seed protein content and nitrogen-use efficiency in maize. *Nature* **2022**, *612*, 292–300. [CrossRef] [PubMed]
5. Cai, J.; Qiu, Z.; Liao, J.; Li, A.; Chen, J.; Wu, Z.; Khan, W.; Sun, B.; Liu, S.; Zheng, P. Comprehensive Analysis of the Yield and Leaf Quality of Fresh Tea (*Camellia sinensis* cv. Jin Xuan) under Different Nitrogen Fertilization Levels. *Foods* **2024**, *13*, 2091. [CrossRef] [PubMed]
6. Li, A.; Qiu, Z.; Liao, J.; Chen, J.; Huang, W.; Yao, J.; Lin, X.; Huang, Y.; Sun, B.; Liu, S.; et al. The Effects of Nitrogen Fertilizer on the Aroma of Fresh Tea Leaves from *Camellia sinensis* cv. Jin Xuan in Summer and Autumn. *Foods* **2024**, *13*, 1776. [CrossRef] [PubMed]
7. Tang, S.; Fu, H.; Pan, W.; Zhou, J.; Xu, M.; Han, K.; Chen, K.; Ma, Q.; Wu, L. Improving tea (*Camellia sinensis*) quality, economic income, and environmental benefits by optimizing agronomic nitrogen efficiency: A synergistic strategy. *Eur. J. Agron.* **2023**, *142*, 126673. [CrossRef]
8. Czarniecka-Skubina, E.; Korzeniowska-Ginter, R.; Pielak, M.; Sałek, P.; Owczarek, T.; Kozak, A. Consumer choices and habits related to tea consumption by poles. *Foods* **2022**, *11*, 2873. [CrossRef]
9. Sonobe, R.; Hirono, Y. Applying variable selection methods and preprocessing techniques to hyperspectral reflectance data to estimate tea cultivar chlorophyll content. *Remote Sens.* **2022**, *15*, 19. [CrossRef]
10. Hodson, T.O. Root mean square error (*RMSE*) or mean absolute error (*MAE*): When to use them or not. *Geosci. Model Dev. Discuss.* **2022**, *2022*, 1–10. [CrossRef]
11. Barman, U.; Sarmah, A.; Sahu, D.; Barman, G.G. Estimation of Tea Leaf Chlorophyll Using MLR, ANN, SVR, and KNN in Natural Light Condition. In Proceedings of the International Conference on Computing and Communication Systems: I3CS 2020, NEHU, Shillong, India, 28–30 April 2020; Springer: Singapore, 2021; pp. 287–295.
12. Janiesch, C.; Zschech, P.; Heinrich, K. Machine learning and deep learning. *Electron. Mark.* **2021**, *31*, 685–695. [CrossRef]
13. Tranmer, M.; Elliot, M. Multiple linear regression. *Cathie Marsh Cent. Census Surv. Res. (CCSR)* **2008**, *5*, 1–5.
14. Zou, J.; Han, Y.; So, S.S. Overview of artificial neural networks. In *Artificial Neural Networks: Methods and Applications*; Humana Press: Totowa, NJ, USA, 2009; pp. 14–22.
15. Awad, M.; Khanna, R.; Awad, M.; Khanna, R. Support vector regression. In *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*; Apress: Berkeley, CA, USA, 2015; pp. 67–80.

16. Kramer, O. K-nearest neighbors. In *Dimensionality Reduction with Unsupervised Nearest Neighbors*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 13–23.
17. Zeng, X.; Tang, S.; Peng, S.; Xia, W.; Chen, Y.; Wang, C. Estimation of Chlorophyll Content of Tea Tree Canopy using multispectral UAV images. In Proceedings of the 2023 11th IEEE International Conference on Agro-Geoinformatics (Agro-Geoinformatics), Wuhan, China, 25–28 July 2023; pp. 1–6.
18. Wang, W.; Cheng, Y.; Ren, Y.; Zhang, Z.; Geng, H. Prediction of chlorophyll content in multi-temporal winter wheat based on multispectral and machine learning. *Front. Plant Sci.* **2022**, *13*, 896408. [CrossRef]
19. Rasooli Sharabiani, V.; Soltani Nazarloo, A.; Taghinezhad, E.; Veza, I.; Szumny, A.; Figiel, A. Prediction of winter wheat leaf chlorophyll content based on VIS/NIR spectroscopy using ANN and PLSR. *Food Sci. Nutr.* **2023**, *11*, 2166–2175. [CrossRef]
20. Zahir, S.A.D.M.; Omar, A.F.; Jamlos, M.F.; Azmi, M.A.M.; Muncan, J. A review of visible and near-infrared (Vis-NIR) spectroscopy application in plant stress detection. *Sens. Actuators A Phys.* **2022**, *338*, 113468. [CrossRef]
21. Brewer, K.; Clulow, A.; Sibanda, M.; Gokool, S.; Naiken, V.; Mabhaudhi, T. Predicting the chlorophyll content of maize over phenotyping as a proxy for crop health in smallholder farming systems. *Remote Sens.* **2022**, *14*, 518. [CrossRef]
22. Wang, S.M.; Ma, J.H.; Zhao, Z.M.; Yang, H.Z.Y.; Xuan, Y.M.; Ouyang, J.X.; Fan, D.M.; Yu, J.F.; Wang, X.C. Pixel-class prediction for nitrogen content of tea plants based on unmanned aerial vehicle images using machine learning and deep learning. *Expert Syst. Appl.* **2023**, *227*, 120351. [CrossRef]
23. Wu, J.; Zareef, M.; Chen, Q.; Ouyang, Q. Application of visible-near infrared spectroscopy in tandem with multivariate analysis for the rapid evaluation of matcha physicochemical indicators. *Food Chem.* **2023**, *421*, 136185. [CrossRef]
24. Al-Selwi, S.M.; Hassan, M.F.; Abdulkadir, S.J.; Muneer, A.; Sumiea, E.H.; Alqushaibi, A.; Ragab, M.G. RNN-LSTM: From applications to modeling techniques and beyond—Systematic review. *J. King Saud Univ.-Comput. Inf. Sci.* **2024**, *36*, 102068. [CrossRef]
25. Barzin, R.; Lotfi, H.; Varco, J.J.; Bora, G.C. Machine learning in evaluating multispectral active canopy sensor for prediction of corn leaf nitrogen concentration and yield. *Remote Sens.* **2021**, *14*, 120. [CrossRef]
26. Zhang, R.; Yang, P.; Liu, S.; Wang, C.; Liu, J. Evaluation of the methods for estimating leaf chlorophyll content with SPAD chlorophyll meters. *Remote Sens.* **2022**, *14*, 5144. [CrossRef]
27. Dai, Z.; Liu, H.; Le, Q.V.; Tan, M. Coatnet: Marrying convolution and attention for all data sizes. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 3965–3977.
28. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.
29. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
30. Ji, Q.; Jiang, L.; Zhang, W. Dual-view noise correction for crowdsourcing. *IEEE Internet Things J.* **2023**, *10*, 11804–11812. [CrossRef]
31. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. Ghostnet: More features from cheap operations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 1580–1589.
32. Touvron, H.; Bojanowski, P.; Caron, M.; Cord, M.; El-Nouby, A.; Grave, E.; Izacard, G.; Joulin, A.; Synnaeve, G.; Verbeek, J.; et al. Resmlp: Feedforward networks for image classification with data-efficient training. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 5314–5321. [CrossRef] [PubMed]
33. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
34. Gao, M.; Xu, M.; Davis, L.S.; Socher, R.; Xiong, C. Startnet: Online detection of action start in untrimmed videos. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 5542–5551.
35. Wu, K.; Zhang, J.; Peng, H.; Liu, M.; Xiao, B.; Fu, J.; Yuan, L. Tinyvit: Fast pretraining distillation for small vision transformers. In *Proceedings of the European Conference on Computer Vision*; Springer: Cham, Switzerland, 2022; pp. 68–85.
36. Zhang, S. Challenges in KNN classification. *IEEE Trans. Knowl. Data Eng.* **2021**, *34*, 4663–4675. [CrossRef]
37. Shams, S.R.; Jahani, A.; Kalantary, S.; Moeinaddini, M.; Khorasani, N. The evaluation on artificial neural networks (ANN) and multiple linear regressions (MLR) models for predicting SO_2 concentration. *Urban Clim.* **2021**, *37*, 100837. [CrossRef]
38. Sun, Y.; Ding, S.; Zhang, Z.; Jia, W. An improved grid search algorithm to optimize SVR for prediction. *Soft Comput.-Fusion Found. Methodol. Appl.* **2021**, *25*, 5633–5644. [CrossRef]
39. Tang, D.Y.Y.; Chew, K.W.; Ting, H.Y.; Sia, Y.H.; Gentili, F.G.; Park, Y.K.; Banat, F.; Culaba, A.B.; Ma, Z.; Show, P.L. Application of regression and artificial neural network analysis of Red-Green-Blue image components in prediction of chlorophyll content in microalgae. *Bioresour. Technol.* **2023**, *370*, 128503.

40. Azadnia, R.; Rajabipour, A.; Jamshidi, B.; Omid, M. New approach for rapid estimation of leaf nitrogen, phosphorus, and potassium contents in apple-trees using Vis/NIR spectroscopy based on wavelength selection coupled with machine learning. *Comput. Electron. Agric.* **2023**, *207*, 107746. [CrossRef]
41. Zhang, D.; Zhang, J.; Peng, B.; Wu, T.; Jiao, Z.; Lu, Y.; Li, G.; Fan, X.; Shen, S.; Gu, A.; et al. Hyperspectral model based on genetic algorithm and SA-1DCNN for predicting Chinese cabbage chlorophyll content. *Sci. Hortic.* **2023**, *321*, 112334. [CrossRef]
42. Hu, T.; Liu, Z.; Hu, R.; Tian, M.; Wang, Z.; Li, M.; Chen, G. Convolutional neural network-based estimation of nitrogen content in regenerating rice leaves. *Agronomy* **2024**, *14*, 1422. [CrossRef]
43. Barman, U.; Choudhury, R.D. Smartphone image based digital chlorophyll meter to estimate the value of citrus leaves chlorophyll using Linear Regression, LMBP-ANN and SCGBP-ANN. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 2938–2950. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.