agronomy

MDPI

Article

SC-ResNeXt: A Regression Prediction Model for Nitrogen Content in Sugarcane Leaves

Zihao Lu [1], Cuimin Sun [2,3,*], Junyang Dou [2], Biao He [2], Muchen Zhou [1] and Hui You [1]

[1] School of Mechanical Engineering, Guangxi University, Nanning 530004, China; 2211391038@st.gxu.edu.cn (Z.L.); 2111301085@st.gxu.edu.cn (M.Z.); usmlhy@iim.ac.cn (H.Y.)
[2] School of Computer, Electronics and Information, Guangxi University, Nanning 530004, China; 2413301004@st.gxu.edu.cn (J.D.); 2313393017@st.gxu.edu.cn (B.H.)
[3] Guangxi Colleges and Universities Key Laboratory of Multimedia Communications and Information Processing, Guangxi University, Nanning 530004, China
* Correspondence: cmsun@gxu.edu.cn

Abstract: In agricultural production, the nitrogen content of sugarcane is assessed with precision and the economy, which is crucial for balancing fertilizer application, reducing resource waste, and minimizing environmental pollution. As an important economic crop, the productivity of sugarcane is significantly influenced by various environmental factors, especially nitrogen supply. Traditional methods based on manually extracted image features are not only costly but are also limited in accuracy and generalization ability. To address these issues, a novel regression prediction model for estimating the nitrogen content of sugarcane, named SC-ResNeXt (Enhanced with Self-Attention, Spatial Attention, and Channel Attention for ResNeXt), has been proposed in this study. The Self-Attention (SA) mechanism and Convolutional Block Attention Module (CBAM) have been incorporated into the ResNeXt101 model to enhance the model's focus on key image features and its information extraction capability. It was demonstrated that the SC-ResNeXt model achieved a test R^2 value of 93.49% in predicting the nitrogen content of sugarcane leaves. After introducing the SA and CBAM attention mechanisms, the prediction accuracy of the model improved by 4.02%. Compared with four classical deep learning algorithms, SC-ResNeXt exhibited superior regression prediction performance. This study utilized images captured by smartphones combined with automatic feature extraction and deep learning technologies, achieving precise and economical predictions of the nitrogen content in sugarcane compared to traditional laboratory chemical analysis methods. This approach offers an affordable technical solution for small farmers to optimize nitrogen management for sugarcane plants, potentially leading to yield improvements. Additionally, it supports the development of more intelligent farming practices by providing precise nitrogen content predictions.

Keywords: sugarcane; nitrogen content; ResNeXt; prediction; self-attention; CBAM

check for updates

Academic Editors: Arjun Neupane, Tej Bahadur Shahi and Richard Koech

Received: 7 December 2024
Revised: 7 January 2025
Accepted: 10 January 2025
Published: 13 January 2025

Citation: Lu, Z.; Sun, C.; Dou, J.; He, B.; Zhou, M.; You, H. SC-ResNeXt: A Regression Prediction Model for Nitrogen Content in Sugarcane Leaves. *Agronomy* **2025**, *15*, 175. https://doi.org/10.3390/agronomy15010175

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

According to the Food and Agriculture Organization (FAO) of the United Nations' "2023 Statistical Yearbook" and statistical database, sugarcane accounts for 20% of global crop production, totaling 1.9 billion tons annually [1]. Growing and developing as an important economic crop, sugarcane is affected by external environmental factors, particularly nitrogen supply [2]. Nitrogen fertilizer application, as a critical environmental factor, can lead to poor growth and reduced yield when insufficient and may cause waste

and environmental pollution when excessive. Over the years, efforts have been made to address this balance [3–6]. Historically, determining whether sugarcane lacked nitrogen mainly relied on visual inspection and empirical summaries. Farmers can tell if sugarcane is deficient in nitrogen by looking at the color of the plant, the rate of growth, and the health of the leaves. For example, when sugarcane leaves turn yellow or the plant grows slowly, experienced farmers may infer that it is insufficient nitrogen and adjust their fertilization strategy accordingly. To precisely determine the nitrogen content of sugarcane, samples had to be destructively collected from the sugarcane plants and sent to laboratories. There, they underwent a series of complex procedures including multi-day drying, grinding, and digestion before instrumental analysis could be performed [7]. This elaborate process not only required significant time but also incurred substantial costs, with each sample analysis costing approximately USD 50–100. Neither method could fully meet the actual production needs of farmers.

With continuous breakthroughs in artificial intelligence technology in recent years, computer vision is being widely applied in agriculture, and comprehensive smart information systems for farmland are becoming cutting-edge research areas. These systems assist farmers in more accurately understanding the specific impacts of external environments (such as nitrogen fertilizer application) on plant growth, thereby optimizing management practices [8–10]. Current mainstream image acquisition methods often involve spectral equipment or drone cameras to photograph large areas of sugarcane fields [11–14]. Compared to traditional visual inspections, these technologies can more efficiently cover extensive agricultural areas. However, they provide only rough estimates of nitrogen content and cannot accurately reflect the nitrogen levels at the individual plant level. Additionally, both spectral cameras and drones come with significant costs, with prices ranging from several thousand to tens of thousands of dollars, making them less accessible for small-scale farmers. In contrast, smartphones, currently used by 4.3 billion users worldwide [15], can capture detailed agronomic traits and growth states of sugarcane through close-range photography with their cameras [14]. Such information, closely related to the absorption and utilization of various nutrients in sugarcane, is difficult to obtain through visual observation or drone photography [16,17].

The color, texture, and shape of plant leaves are affected by changes in the external environment (such as nitrogen absorption) and adjust their growth and development strategies through internal signaling transduction mechanisms [18]. Therefore, in previous studies, researchers typically focused on the relationship between leaf color [10,19–21], texture [20,22,23], and shape [24–26] characteristics and crop nutrient content. Existing methods rely on traditional machine learning models built on small sample data, such as SVM (Support Vector Machine), RF (Random Forest) based on statistical learning theory, and shallow ANN (Artificial Neural Network) models, to perform regression predictions of crop nutrient content. Due to constraints on human resources and time, researchers typically cannot collect thousands of leaf samples; a few hundred images are often the practical limit. Consequently, the data volume available for training these models is limited. Although these methods can achieve good accuracy for estimating nutrient content at specific stages, they exhibit poor generalization when dealing with data from different periods or varieties.

Moreover, in feature extraction, existing methods usually require researchers to manually select and define which features best aid the model in making accurate predictions based on experience. This easily overlooks non-defined traits or features, reducing prediction accuracy. For instance, Lei Sun et al. used a CNN (convolutional neural network) model for nitrogen content prediction in corn leaves but still required manual feature selection [21]. Sulistyo S.B. et al., after manually selecting and extracting 12 color features,

input them into a model combining Deep Sparse Extreme Learning Machines (DSELMs) and Genetic Algorithms (GAs) to estimate nitrogen content in wheat leaves [19]. Janani M. et al., after extracting color features from peanut leaves, also needed to annotate the data before inputting it into a CNN-based model [10]. These methods [20,27,28] evidently entail high labor and time costs.

Compared with other models, ResNeXt possesses strong feature extraction capabilities. It does this using multiple paths that work at the same time (parallel branches). Each path applies the same type of filters to the image and stacks these layers to learn different aspects of the image automatically. This setup helps ResNeXt understand complex patterns in images, allowing it to learn directly from raw data to final output in an end-to-end fashion. Consequently, ResNeXt can efficiently accomplish image regression prediction tasks, demonstrating stronger model performance and generalization ability. However, current research on the application of ResNeXt in regression prediction of crop nutrient content is not yet sufficient, and the effectiveness of regression prediction based on automatic feature extraction methods is relatively lacking [19,29].

In this study, in response to the need for precise assessment of nitrogen content in sugarcane in agricultural production, a deep learning-based regression prediction model called SC-ResNeXt was proposed. Based on ResNeXt101, this model integrates two attention mechanisms, Self-Attention (SA) and Convolutional Block Attention Module (CBAM), to enhance the recognition and extraction of key features from images of sugarcane leaves. Through images captured by smartphones, an end-to-end learning approach was adopted, complementing traditional trait observations with automatic feature extraction technology. The SC-ResNeXt model achieved a test R^2 value of 0.9349 in predicting the nitrogen content of sugarcane leaves, representing a significant improvement in prediction accuracy compared to previous studies. This research not only solved the problems of high cost and low efficiency associated with traditional methods but also helped optimize nitrogen management in sugarcane, increasing crop productivity and providing farmers with a precise and economically effective solution for nitrogen fertilizer application.

2. Materials and Methods

2.1. Experimental Description

Field trials were conducted in the Agricultural High-tech Industry Demonstration Zone of Quli Town, Fusui County, Chongzuo City, Guangxi Zhuang Autonomous Region, China (107.8° E, 22.5° N), an area characterized by a typical subtropical monsoon climate. In 2023, the average temperature was 23.1 °C, with 1513 h of sunshine and a total rainfall of 1183.5 mm. The soil texture of the experimental field was clay, and during the week of the experiment, the average soil moisture content at depths of 0–20 cm was 29.35%, while the average relative soil moisture content was 108.94%. Other properties of the experimental field's soil are shown in Table 1.

Table 1. Soil nutrient content of the experimental field.

	Total Nitrogen (g/kg)	Total Phosphorus (mg/kg)	Total Potassium (mg/kg)	Organic Carbon (g/kg)	pH
Values	0.97	63.92	102.72	11.46	5.12

The sugarcane variety ROC22 was selected for the trial, planted in January 2023, with the entire year's planned fertilizer amount applied as base fertilizer at planting time. New plantings were harvested in March 2024, and at the time of the experiment, the newly grown ratoon cane was transitioning from the seedling stage to the tillering phase. Adequate application of base fertilizer at planting can lead to the robust and rapid growth

of sugarcane, which has a critical need for nutrients during the seedling stage, especially nitrogen, essential for subsequent growth [30].

A total of 86 sugarcane plants were randomly selected from the experimental field, and their fully expanded leaves were collected. Two smartphones (Chinese Honor and Chinese Realme, both equipped with Sony IMX series CMOS sensors) were used to photograph different parts of the sugarcane leaves from fixed dual-camera positions at different angles, using A3 grid paper (grid size 1 cm × 1 cm) as a background. The camera heights were approximately 19.5 cm and 17.0 cm, respectively, as shown in Figure 1. All captured images were in jpg format with resolutions set at 3072 × 4096 and 6000 × 8000, and the focal lengths were fixed at 5.59 mm and 5 mm.



Figure 1. Field experiment.

2.2. Sample Measurement

The collected sugarcane leaf samples were washed with ultrapure water for 2 min and then placed in an oven at 105 °C for 120 min. Subsequently, the temperature was reduced to 75 °C where the samples were dried continuously for 36 h until their weight remained constant. The dried samples were then ground into a fine powder using a mill. An electronic balance was used to weigh out 0.2 g of the sample powder, which was placed into a digestion tube. After adding 5 mL of concentrated sulfuric acid and mixing thoroughly, the digestion tubes were placed in a digestion furnace and heated to 400 °C. Once the liquid inside the digestion tube had stabilized and turned brown, the tubes were removed. Hydrogen peroxide was slowly added while shaking until the digestion solution became clear and colorless. The tubes were then returned to the digestion furnace for approximately another 10 min to ensure all remaining hydrogen peroxide had evaporated. Finally, ultrapure water was added to dilute the digestion solution to a total volume of 500 mL, after which it was cooled to room temperature. This improved H_2SO_4-H_2O_2 digestion method is commonly used for the determination of total nitrogen in plants [7].

A 10 mL aliquot of the diluted digestion solution was taken for machine analysis. The total nitrogen content of the samples was measured using a discrete auto analyzer (Model: SMARTCHEM 200), manufactured by AMS Italy S.r.l., based in Pavia, Italy. The nitrogen content of the samples was calculated using Formula (1) as follows:

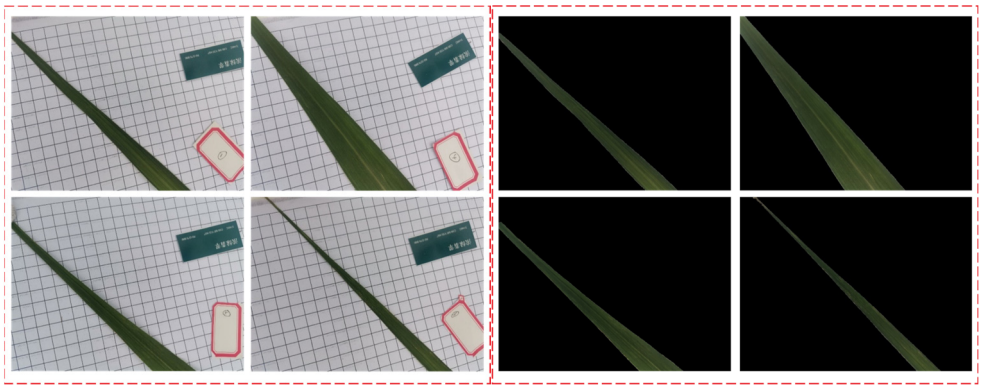
$$N = \frac{r \times 0.1}{m} \quad (1)$$

where N represents the total nitrogen content of the sample (g/kg), r is the instrument reading (mg/L), and m is the mass of the sample (g).

Before sample measurement, standard solutions were analyzed to create a calibration curve. The R^2 values for the calibration curve, determined over four measurements, were 0.9043, 0.9698, 0.9799, and 0.9955, respectively.

2.3. Dataset Preparation

A total of 86 sugarcane leaf samples were collected, from which 172 images were captured. All images were resized to a uniform dimension and converted into grayscale and HSV color space. Image backgrounds were removed using thresholding and contour extraction methods, with the removal effect illustrated in Figure 2. Due to the limited number of sample images, to enhance the accuracy of testing, the dataset was divided into training, validation, and test sets at a ratio of 7:1:2. The transforms function was imported from torchvision to resize the images to (800, 600) and then randomly augment the data threefold through transformations, such as rotation and cropping, thereby increasing data diversity and improving model generalization capability.



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

Figure 2. Comparison before and after image background removal: (**a**) before removal; (**b**) after removal.

2.4. Experimental Environment

The experimental environment parameters are shown in Table 2.

Table 2. Parameters of the experimental environment.

Items	Detail
Operating System	Linux
CPU	Intel Core i9-14900K
GPU	NVIDIA GeForce RTX 4090 (24G)
Acceleration Env	CUDA 12.6
Language	Python 3.10.4
Framework	Pytorch 2.3.0

2.5. SC-ResNeXt

This study opted for ResNeXt101 as the backbone network of the model, incorporating SA (Self-Attention) and CBAM (Convolutional Block Attention Module) attention mechanisms to construct an automatic feature extraction regression prediction model for sugarcane nitrogen content, named SC-ResNeXt. The architecture of the model is illustrated in Figure 3. The input images undergo preprocessing initially, followed by passage through a Conv1 convolutional layer and a MaxPool pooling layer to reduce the size of the

feature maps. Subsequently, 33 ResNeXt blocks extract features via deep learning processes that include 1 × 1 compression convolutions, 3 × 3 grouped convolutions, and 1 × 1 expansion convolutions, with residual connections ensuring information flow. To further enhance the model's feature selection capabilities, SA and CBAM attention mechanisms are introduced after the ResNeXt blocks. The SA mechanism generates attention maps through Query Conv, Key Conv, and Value Conv layers, thereby enhancing the model's focus on key features. The CBAM mechanism employs two modules, CAM (Channel Attention Module) and SAM (Spatial Attention Module), to weight features from channel and spatial dimensions, respectively, thus improving the model's feature selection ability. Finally, a Global Average Pooling (GAP) layer compresses the feature maps into a single vector, which is then passed through a fully connected layer (FC) to output the regression prediction result—the predicted value of sugarcane nitrogen content.

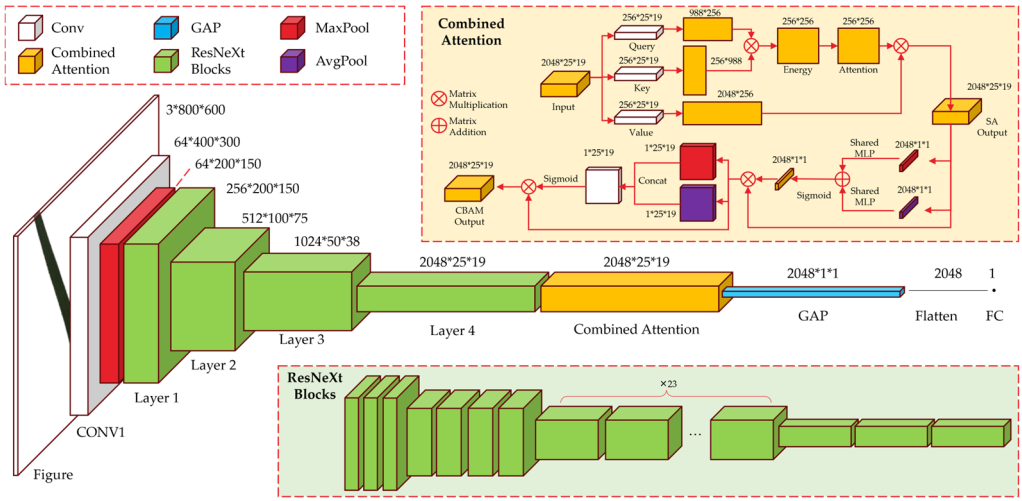


Figure 3. SC-ResNeXt model architecture.

2.5.1. ResNeXt

ResNeXt (Residual Networks with Next), proposed by the Facebook AI Research (FAIR) team in 2017 [31], is a deep convolutional neural network architecture that builds upon ResNet by introducing the concept of "cardinality" through parallel branch structures to further enhance the network's expressive power without significantly increasing the number of parameters or computational complexity. Initially designed for image recognition tasks, ResNeXt's robust feature extraction capability and efficient structure also render it suitable for image regression prediction tasks.

ResNeXt101, a classic configuration of the ResNeXt architecture, features a depth of 101 layers, enabling it to learn more complex feature representations. The symmetric group convolutions within ResNeXt101 allow for an excellent balance between performance and complexity, ensuring superior performance even with small sample datasets. The architecture of ResNeXt101 comprises an input layer, an initial convolution layer, improved residual blocks known as ResNeXt blocks, a global average pooling layer, and a fully connected layer, as illustrated in Figure 4.

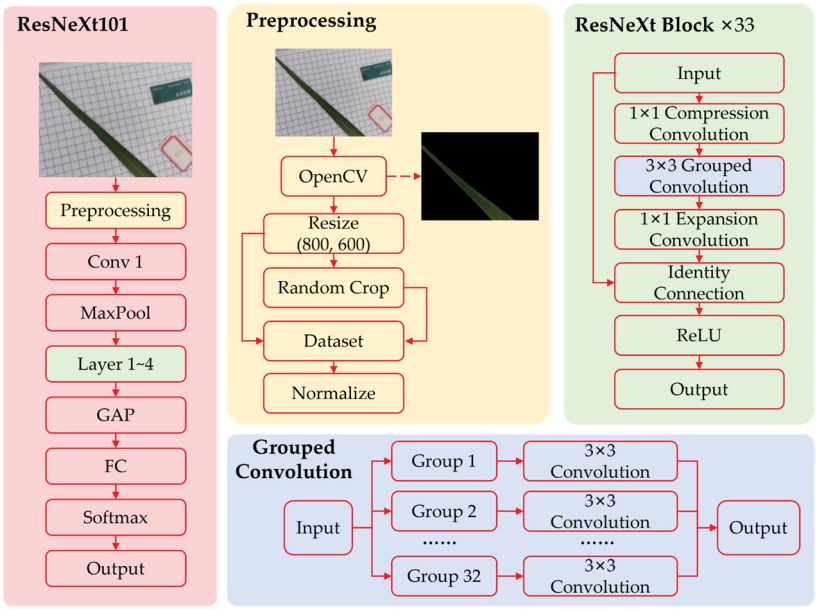


Figure 4. Structure of ResNeXt101.

Indeed, while ResNeXt has evolved from ResNet with notable advancements in network design and performance improvements, it still encounters limitations when applied to regression prediction tasks for crop nutrient content. Specifically, the challenges include limited feature extraction capability and suboptimal fitting performance. In regression tasks involving crop nutrition content, the available datasets tend to be quite small. Extracting sufficient information from these small sample datasets, particularly capturing global context information, is, therefore, critical. Traditional convolution operations primarily focus on extracting local features with a fixed receptive field, which represents a shortcoming of ResNeXt. The emergence and growing popularity of attention mechanisms have highlighted the need for models that can better handle the weighting of features, especially in scenarios where explicit attention mechanisms are lacking, as seen in ResNeXt. This deficiency makes it difficult for ResNeXt to properly address feature weight imbalances, thereby affecting its fitting performance. To address these issues, the introduction of CBAM (Convolutional Block Attention Module) and SA (Self-Attention) mechanisms aims to improve ResNeXt. CBAM highlights important features from both channel and spatial perspectives while suppressing less relevant ones, thereby strengthening model generalization and improving prediction accuracy. On the other hand, SA enables the model to capture relationships between distant pixels and global feature associations, such as potential shapes and textures, dynamically adjusting feature weights. This enriches feature representation and mitigates the impact of local noise.

By incorporating these two attention mechanisms, the structure effectively expands the receptive field of ResNeXt, enhancing the robustness of the model. It allows for more effective performance improvement, making it better suited for handling the complexities of regression prediction tasks related to crop nutrient content, even when working with smaller datasets. This enhancement ensures that the model can better extract and utilize the global context information necessary for accurate predictions.

2.5.2. CBAM

To further enhance the performance of the model, the CBAM (Convolutional Block Attention Module) attention mechanism was introduced [32]. CBAM incorporates channel attention (Channel Attention) and spatial attention (Spatial Attention) mechanisms, enabling the network to focus more on important features while suppressing irrelevant ones.

This enhances the feature representation of ResNeXt101 and improves model performance. The structure of CBAM is illustrated in Figure 5.

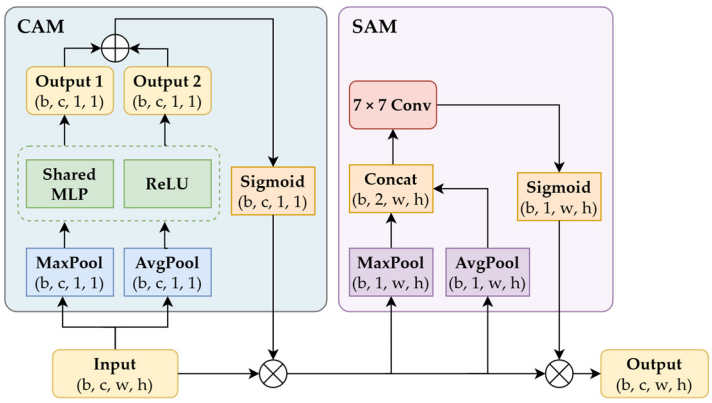


Figure 5. Structure diagram of CBAM.

As shown in Figure 5, the feature maps are first input into the CAM (Channel Attention Module) section. After computing MaxPool and AvgPool separately, the results pass through a shared MLP module. The number of channels is initially compressed to 1/R (reduction rate) of its original size and then expanded back to the original number of channels. Following ReLU activation, the outputs are added together and passed through the Sigmoid activation function to obtain the output of the CAM. The CAM output is then multiplied with the original feature map and fed into the SAM (Spatial Attention Module) section. Here, MaxPool and AvgPool are computed again, followed by concatenation via a Concat operation. The concatenated result undergoes a 7×7 convolution and passes through the Sigmoid activation function to produce the output of the SAM. Finally, multiplying the SAM output with the original feature map yields the CBAM output. The CBAM output is calculated using Formulas (2)–(5) as follows:

$$M_{cam}(F) = \sigma\{MLP[AvgPool(F)] + MLP[MaxPool(F)]\} \tag{2}$$

$$F' = M_{cam}(F) \otimes F \tag{3}$$

$$M_{sam}(F') = \sigma\left\{f^{7\times7}\left[AvgPool(F'); MaxPool(F')\right]\right\} \tag{4}$$

$$F'' = M_{sam}(F') \otimes F' \tag{5}$$

where F represents the input feature map, σ denotes the Sigmoid activation function, $M_{cam}(F)$ and $M_{sam}(F')$ represent the output results from the CAM and SAM sections, respectively, F' is the input to the SAM section, and F'' is the final output of CBAM.

2.5.3. Self-Attention

To further enhance the model's spatial perception capabilities, Self-Attention (SA), originally introduced to boost the performance of Transformer models [33], was incorporated. Self-Attention is a technique designed to capture dependencies between different positions within sequences or feature maps by calculating similarity scores between each position and all other positions, generating an attention weight matrix, and then re-weighting and aggregating features based on this matrix. The structure of SA is illustrated in Figure 6, where "w*h" represents the matrix multiplication of width and height dimensions in the output size.

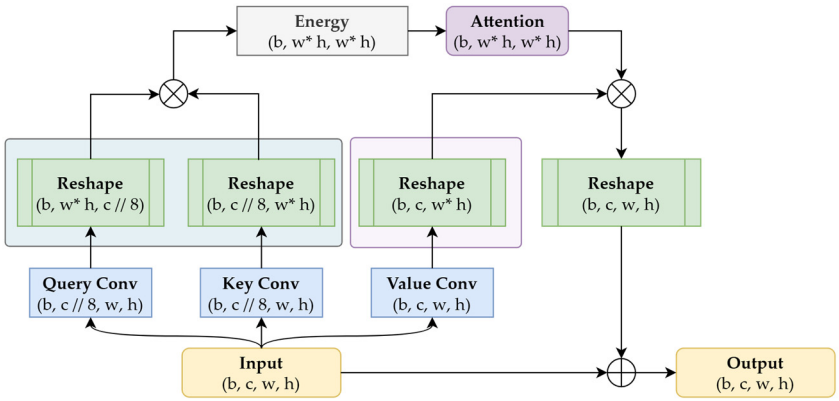


Figure 6. Structure diagram of self-attention.

According to Figure 6, the feature map is inputted and then goes through three convolutions to be transformed into Query, Key, and Value. Query and Key are matched to compute attention weights; Value contains the original feature information and will be weighted and summed in the attention mechanism. The energy matrix obtained by matrix multiplication is applied to the Softmax function and normalized into a weight matrix. Finally, Value is applied to the weight matrix via matrix multiplication to obtain the weighted feature map. After reshaping, it is linearly combined with the original feature map and outputted. The output of SA can be calculated using Formulas (6)–(9) as follows:

$$Q = reshape\left(X \cdot W_Q\right) \tag{6}$$

$$K = reshape\left(X \cdot W_K\right) \tag{7}$$

$$V = reshape\left(X \cdot W_V\right) \tag{8}$$

$$Y = X + \gamma \cdot reshape\left\{V \cdot [softmax(Q \cdot K)]^T\right\} \tag{9}$$

where Q, K, and V are the two-dimensional matrices of Query, Key, and Value after convolution and reshaping; W is the weight matrix; X is the input tensor; Y is the output tensor; and γ is a learnable scaling factor during training.

2.5.4. Evaluation Metrics

In this study, the performance of the model is primarily evaluated using the metrics outlined in Table 3.

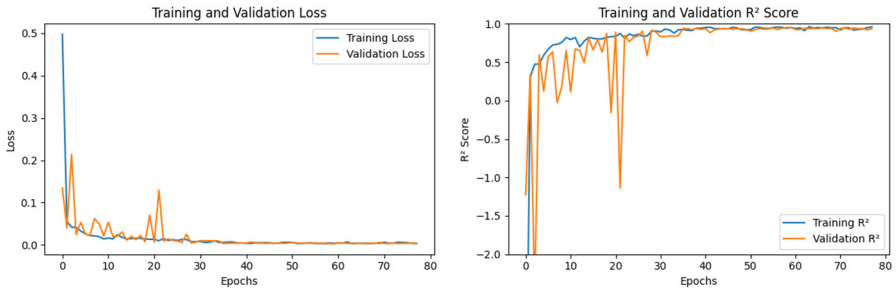
Table 3. Evaluation metrics for SC-ResNeXt model performance.

Metrics	Definition	Formula	Purpose
MAE (Mean Absolute Error)	The average of the absolute differences between predicted values and actual values.	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert\hat{y}_i - y_i\rvert$	MAE provides a clear indication of the magnitude of prediction errors.
MSE (Mean Square Error)	The average of the squared differences between true values and predicted values.	$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$	MSE is used to measure the deviation between the model's predictions and the actual values.
RMSE (Root Mean Square Error)	The square root of the mean square error.	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$	RMSE represents the sample standard deviation of the differences between predicted values and observed values.
R^2 (Coefficient of Determination)	Reflects the accuracy of the model in fitting the data.	$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\overline{y} - y_i)^2}$	An R^2 value closer to 1 indicates a better fit of the model to the data.

Where $\hat{y}_i$ represents the predicted value, y_i represents the true value, n represents the number of samples, and $\overline{y}$ represents the mean of the true values.

3. Results

In this study, the pre-trained weights of ResNeXt on the ImageNet dataset were loaded. During the model training, MSE was employed as the loss function. After performing a grid search for hyperparameter tuning, the AdamW optimizer was utilized with an initial learning rate of 0.005 and a weight decay coefficient of 0.01. A learning rate scheduler was configured such that if the validation loss did not continue to decrease for five consecutive epochs, the learning rate would be reduced to one-tenth of its current value. An early stopping mechanism was also established, whereby training would cease and the best epoch would be saved if no further improvements occurred over twenty consecutive epochs. The loss and R^2 during the model training process are illustrated in Figure 7.



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

Figure 7. The training process of SC-ResNeXt: (**a**) loss convergence; (**b**) R^2 convergence.

According to Figure 7, during the training process, SC-ResNeXt rapidly converged in the first three epochs. From the tenth epoch onwards, the rate of convergence slowed down, and early stopping was triggered at the seventy-eighth epoch, with the best result being saved from the fifty-eighth epoch. The validation and testing R^2 values for the optimal epoch reached 0.9440 and 0.9349, respectively, achieving highly accurate predictions of nitrogen content in sugarcane.

In deep learning models, explaining the decision-making process is crucial for understanding how the model works, enhancing model credibility, and identifying potential biases. Grad-CAM (Gradient-weighted Class Activation Mapping) not only highlights which parts of an image contribute most significantly to the final regression outcome but also maintains spatial granularity, ensuring a direct correspondence between the heatmap and the original input image. This aids in comprehending how the model utilizes different portions of the image for making regression decisions.

To gain an intuitive understanding of SC-ResNeXt's feature perception capability, the last convolutional layer of Layer4 was extracted, and heatmaps were generated using Grad-CAM. As shown in Figure 8, the color intensity in the heatmaps provides insight into SC-ResNeXt's attention mechanism; redder areas indicate higher focus, while bluer areas signify less importance. SC-ResNeXt effectively concentrates on the edges and internal textures of different leaves, particularly excelling in identifying critical features. Notably, the recognition effect on the last leaf in the bottom-right corner is exceptionally outstanding, with the model focusing almost exclusively on the leaf area, leaving regions outside the leaf virtually colorless. The results in Figure 8 confirm that SC-ResNeXt not only has an excellent feature selection mechanism for handling the regression task of sugarcane leaves but also demonstrates superior generalization across various samples. This robust performance underscores the model's ability to accurately identify key features consistently, even in diverse and complex leaf images.

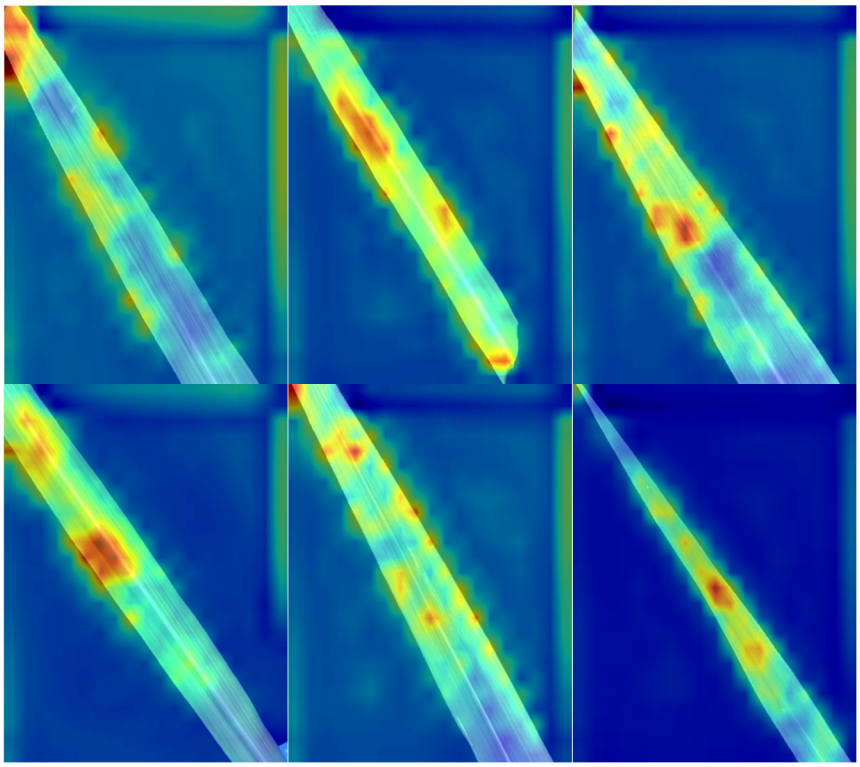


Figure 8. Grad-CAM heatmaps of SC-ResNeXt Layer4.

To observe how SC-ResNeXt gradually extracts more abstract and advanced features, the feature parameters before and after the application of the attention mechanism in the first convolutional layer of SC-ResNeXt were analyzed. For each image, the model extracted 1,920,000 features after the first convolutional layer, which were narrowed down to 972,800 features after the last ResNeXt block and further refined to 2048 features following the attention mechanism processing. The handling of high-dimensional data demonstrates the model's performance quality. Subsequently, the feature parameters were reduced to 100 dimensions via PCA (Principal Component Analysis) and then to two dimensions using t-SNE (t-distributed Stochastic Neighbor Embedding). Labels were divided into 5 categories, numbered from 0 to 4. In the resulting two-dimensional coordinate plot, each category is represented by points of a distinct color: category 0 through category 4, respectively. The visualized feature map is depicted in Figure 9.

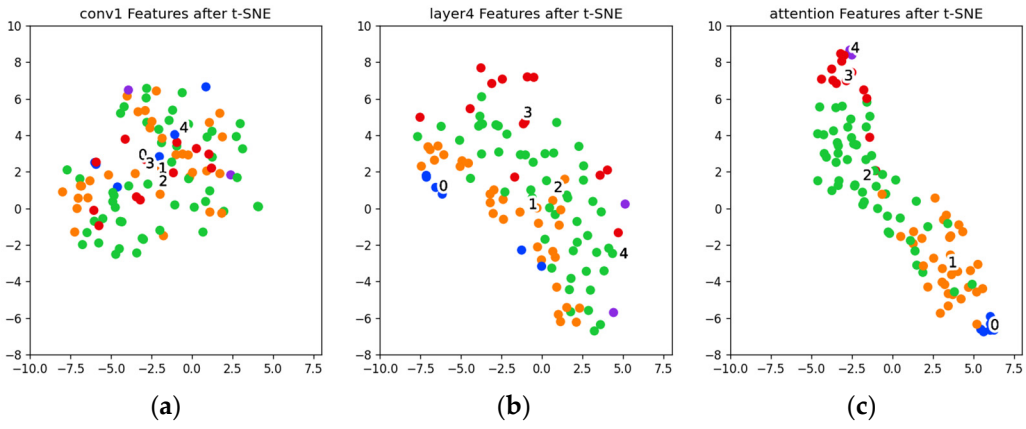


Figure 9. The t-SNE feature maps: (**a**) the output of the first convolutional layer; (**b**) before the attention mechanism; (**c**) after the attention mechanism.

As indicated by Figure 9, the features extracted by the first convolutional layer are intertwined and loosely distributed, showing high similarity and redundancy among them. Thanks to the powerful ResNeXt blocks, the output of Layer4 showed improvement with a clear trend of separation between different classes, demonstrating strong feature diversity. After further refinement through the attention mechanism, the distribution of feature points became more dispersed and orderly, with clearer boundaries between clusters. This transformation reflects the outstanding feature extraction and screening capabilities of SC-ResNeXt.

To verify the effectiveness of SC-ResNeXt in the regression prediction task of sugarcane nitrogen content, backbone comparison experiments and ablation studies were conducted to further confirm the model's performance.

3.1. Backbone Comparison Experiments

To determine whether ResNeXt was the optimal backbone, a comparison was made between ResNeXt50/101 and other backbones, including ResNet18/34/50/101 and WideResNet18-2/18-3. All backbones were loaded with pre-trained weights from the ImageNet dataset, and identical settings for learning rate adjustment and other parameters were applied. According to the experimental results presented in Figure 10, ResNeXt50/101 demonstrated the highest fitting degree, with test R^2 exceeding 0.8, effectively capturing data patterns and trends. While ResNet's performance did not match that of ResNeXt, it still showed good fitting, with test R^2 results ranging between 0.7 and 0.8. In contrast, Wide-ResNet exhibited poorer prediction effects, only slightly better than the mean prediction, and showed noticeable overfitting. Within the ResNeXt architecture, the performance difference between the 50-layer and 101-layer models was minimal, with a test R^2 difference of only 0.0107. Therefore, ResNeXt101 was identified as the best model backbone.

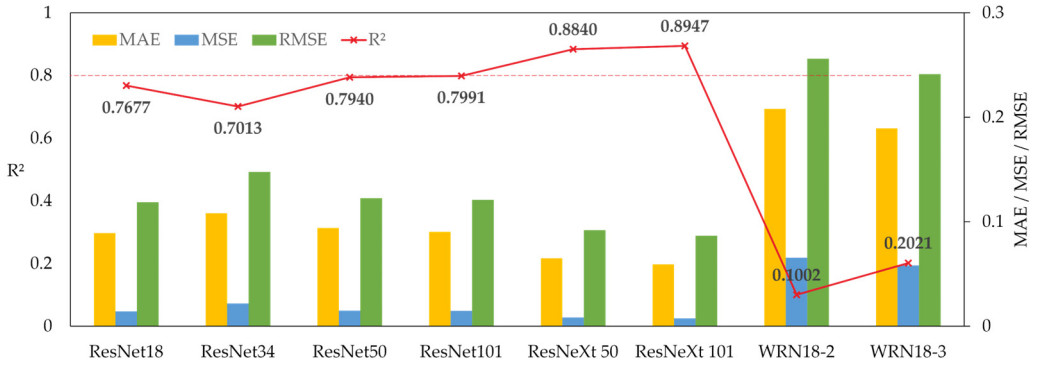


Figure 10. Regression prediction results of different backbones.

3.2. Ablation Study

To ascertain the enhancement effects of SA (Spatial Attention) and CBAM (Convolutional Block Attention Module) on the model, an ablation study was conducted. Given the close performance between ResNeXt50 and ResNeXt101, ResNeXt50 was included in the ablation study for comparison. The two attention mechanisms were sequentially integrated into the model, examining changes during validation and testing processes. To confirm the combined effect of the two attention mechanisms, their integration order was also switched and interfaces adjusted. The results of the ablation study are shown in Figure 11.

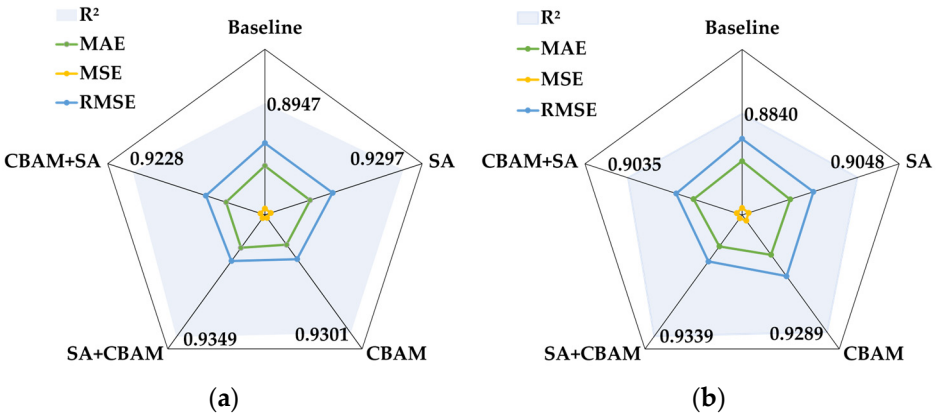


Figure 11. Ablation experiment results: (**a**) ResNeXt101; (**b**) ResNeXt50.

Based on Figure 11, when individually incorporated into ResNeXt101, SA and CBAM increased the test R^2 by 3.50% and 3.54%, respectively, whereas for ResNeXt50, the increases were 2.08% and 4.61%, respectively, with reduced bias in both cases. Both mechanisms validated the effectiveness of attention mechanisms in enhancing deep learning model performance. Among them, CBAM showed more pronounced improvement, considering both channel and spatial attention, while SA focused mainly on interactions between features, which might explain why the dual attention mechanism enhanced feature representation more effectively.

When combined, the attention mechanisms improved the test R^2 for ResNeXt101 by 4.02% and ResNeXt50 by 4.99%. The synergistic effect of SA and CBAM significantly enhances the model's performance. Specifically, SA captures global feature dependencies, while CBAM optimizes local details. This complementarity allows the model to better understand input features from both global and local perspectives. Following this, SA enriches the feature representation, which is then refined through finer feature selection and weighting performed by CBAM. Together, these mechanisms enhance the robustness and generalization of the model. SA improves the model's ability to recognize complex patterns, while CBAM mitigates the impact of noise and interference, leading to more stable and accurate performance across various input types. Therefore, introducing SA and CBAM sequentially into the model effectively reduces bias and achieves optimal prediction outcomes.

3.3. Comparison with Other Algorithms

To better compare the accuracy of this study in predicting nitrogen content in sugarcane, a selection of common deep learning regression algorithms—VIT (Vision Transformer), SqueezeNet, AlexNet, and DenseNet121—were compared with ResNeXt101. All these algorithms automatically extracted and selected features to complete the regression task. Similarly, all models were loaded with pre-trained weights from the ImageNet dataset, and parameters, such as learning rate adjustment settings, were optimized for the best performance.

During training, ResNeXt101 outperformed the other four commonly used models in terms of the loss convergence and R^2 validation results, demonstrating superior training effectiveness. It exhibited faster and more stable convergence, higher prediction accuracy, and insensitivity to fluctuations in the training data, maintaining consistent performance. Specifically, compared to VIT, SC-ResNeXt leverages the strengths of convolutional neural networks in capturing local spatial features while integrating attention mechanisms for enhanced global context understanding. Unlike SqueezeNet and AlexNet, which are relatively shallow and less capable of handling complex feature hierarchies, SC-ResNeXt's deep

architecture allows it to learn more intricate patterns from the data. Furthermore, unlike DenseNet121, which relies heavily on dense connections that can introduce redundancy, SC-ResNeXt's use of cardinality through grouped convolutions ensures efficient feature extraction without excessive parameter complexity. The incorporation of SA and CBAM further enhances its ability to focus on key features, leading to improved generalization and robustness.

According to the final test results presented in Figure 12, VIT's test R^2 approached zero, indicating that its predictive performance was equivalent to mean prediction. SqueezeNet, AlexNet, and DenseNet121 also performed poorly, with the former two achieving a test R^2 around 0.32 and the latter showing an even poorer result with a test R^2 of only 0.1163. All three models exhibited some degree of overfitting, suggesting they failed to capture effective features or relationships between features. In contrast, SC-ResNeXt showed the least bias and achieved the best predictive performance.

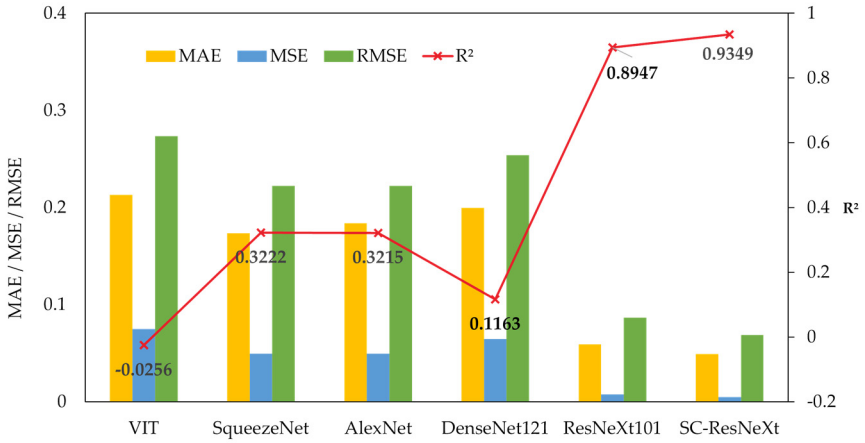


Figure 12. Test results of different algorithms.

4. Discussion

The SC-ResNeXt model has demonstrated significant success in predicting nitrogen content in sugarcane leaves, yet this study faces limitations due to a relatively small and homogeneous sample size, which restricts its generalization across diverse environmental conditions, varieties, and growth stages. To address these limitations, we are actively expanding our dataset by collecting additional samples. Additionally, variability in image quality from smartphone captures may impact prediction accuracy; standardizing equipment or introducing light correction preprocessing could mitigate this issue.

To further improve model performance, exploring additional attention mechanisms, like Spatial Transformer Networks (STNs) or Multi-Scale Attention Mechanisms, can aid in capturing global information and emphasizing local features. Addressing the diminishing returns from the large number of parameters in SC-ResNeXt, lightweight architectures might offer better performance for small datasets. Innovations such as Atrous/Dilated Convolution, Deformable Convolution, and Large Kernel Convolution, along with hybrid model designs, could strengthen feature extraction capabilities.

Beyond nitrogen prediction in sugarcane, SC-ResNeXt's reliance on visual features makes it applicable to other crops, like wheat, corn, and rice, optimizing fertilization strategies and soil management. Adjusting parameters and training sets can extend predictions to other nutrients, broadening agricultural applications. With advancing technology, integrating higher-definition smartphone cameras positions SC-ResNeXt as a key component in smart agriculture systems, supporting precision farming initiatives.

5. Conclusions

This study developed a novel regression prediction model, SC-ResNeXt, which integrates Self-Attention (SA) and Convolutional Block Attention Module (CBAM) into the ResNeXt101 architecture to estimate the nitrogen content in sugarcane leaves. Utilizing images captured by smartphones, SC-ResNeXt achieved a test R^2 value of 0.9349, with a 4.02% increase in accuracy due to the incorporated attention mechanisms. This model outperformed four classical deep learning algorithms, offering enhanced precision and economy over traditional methods that rely on manual feature selection. By providing small farmers with an affordable solution for precise nitrogen content assessment, SC-ResNeXt facilitates informed fertilizer application, optimizing nitrogen management and contributing to sustainable agricultural practices and intelligent farming systems.

Author Contributions: Conceptualization, methodology, software, formal analysis, visualization, and writing—original draft preparation, Z.L.; conceptualization, investigation, and data curation, M.Z.; resources and data curation, J.D. and B.H.; writing—review and editing, project administration, and funding acquisition, C.S. and H.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Guangxi Science and Technology Program (grant number AB24010048), the Guangxi Science and Technology Major Program (grant numbers AA22117005 and AA22117007), and the Guangxi University Innovation and Development Doubling Plan Project (grant numbers 202201343 and 202201369).

Data Availability Statement: The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

Acknowledgments: The relevant data of the field experiment were provided by Guangxi University Agricultural New Town and Fusui County Agriculture and Rural Bureau.

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

References

1. FAO. FAOSTAT. Available online: https://www.fao.org/faostat/en/#home (accessed on 24 October 2024).
2. Wu, K.; Wang, S.; Song, W.; Zhang, J.; Wang, Y.; Liu, Q.; Yu, J.; Ye, Y.; Li, S.; Chen, J.; et al. Enhanced sustainable green revolution yield via nitrogen-responsive chromatin modulation in rice. *Science* **2020**, *367*, eaaz2046. [CrossRef] [PubMed]
3. Liu, Y.; Zhuang, M.; Liang, X.; Lam, S.K.; Chen, D.; Malik, A.; Li, M.; Lenzen, M.; Zhang, L.; Zhang, R.; et al. Localized nitrogen management strategies can halve fertilizer use in Chinese staple crop production. *Nat. Food* **2024**, *5*, 825–835. [CrossRef]
4. Wang, C.; Shen, Y.; Fang, X.; Xiao, S.; Liu, G.; Wang, L.; Gu, B.; Zhou, F.; Chen, D.; Tian, H.; et al. Reducing soil nitrogen losses from fertilizer use in global maize and wheat production. *Nat. Geosci.* **2024**, *17*, 1008–1015. [CrossRef]
5. Liu, X.; Beusen, A.H.W.; van Grinsven, H.J.M.; Wang, J.; van Hoek, W.J.; Ran, X.; Mogollón, J.M.; Bouwman, A.F. Impact of groundwater nitrogen legacy on water quality. *Nat. Sustain.* **2024**, *7*, 891–900. [CrossRef]
6. Chen, X.; Cui, Z.; Fan, M.; Vitousek, P.; Zhao, M.; Ma, W.; Wang, Z.; Zhang, W.; Yan, X.; Yang, J.; et al. Producing more grain with lower environmental costs. *Nature* **2014**, *514*, 486–489. [CrossRef] [PubMed]
7. Li, Z.Y.; Zheng, L.; Lu, L.; Li, L. Improvement in the H_2SO_4-H_2O_2,Digestion Method for Determining Plant Total Nitrogen. *Chin. Agric. Sci. Bull.* **2014**, *30*, 159–162.
8. Iatrou, M.; Karydas, C.; Iatrou, G.; Pitsiorlas, I.; Aschonitis, V.; Raptis, I.; Mpetas, S.; Kravvas, K.; Mourelatos, S. Topdressing Nitrogen Demand Prediction in Rice Crop Using Machine Learning Systems. *Agriculture* **2021**, *11*, 312. [CrossRef]
9. Shankar, T.; Malik, G.C.; Banerjee, M.; Dutta, S.; Praharaj, S.; Lalichetti, S.; Mohanty, S.; Bhattacharyay, D.; Maitra, S.; Gaber, A.; et al. Prediction of the Effect of Nutrients on Plant Parameters of Rice by Artificial Neural Network. *Agronomy* **2022**, *12*, 2123. [CrossRef]
10. Janani, M.; Jebakumar, R. Detection and classification of groundnut leaf nutrient level extraction in RGB images. *Adv. Eng. Softw.* **2023**, *175*. [CrossRef]
11. Li, R.; Wang, D.; Zhu, B.; Liu, T.; Sun, C.; Zhang, Z. Estimation of nitrogen content in wheat using indices derived from RGB and thermal infrared imaging. *Field Crops Res.* **2022**, *289*, 108735. [CrossRef]

12. Cheng, Q.; Wu, B.; Ye, H.; Liang, Y.; Che, Y.; Guo, A.; Wang, Z.; Tao, Z.; Li, W.; Wang, J. Inversion of maize leaf nitrogen using UAV hyperspectral imagery in breeding fields. *Int. J. Agric. Biol. Eng.* **2024**, *17*, 144–155. [CrossRef]
13. Wang, D.; Li, R.; Liu, T.; Liu, S.; Sun, C.; Guo, W. Combining vegetation, color, and texture indices with hyperspectral parameters using machine-learning methods to estimate nitrogen concentration in rice stems and leaves. *Field Crops Res.* **2023**, *304*. [CrossRef]
14. Kolhar, S.; Jagtap, J. Plant trait estimation and classification studies in plant phenotyping using machine vision—A review. *Inf. Process. Agric.* **2023**, *10*, 114–135. [CrossRef]
15. Matthew Shanahan, K.B. *The State of Mobile Internet Connectivity 2023*; GSMA: London, UK, 2023.
16. Li, A.; Wu, Q.; Yang, S.; Liu, J.; Zhao, Y.; Zhao, P.; Wang, L.; Lu, W.; Huang, D.; Zhang, Y.; et al. Dissection of genetic architecture for desirable traits in sugarcane by integrated transcriptomics and metabolomics. *Int. J. Biol. Macromol.* **2024**, *280*, 136009. [CrossRef]
17. Meena, M.R.; Appunu, C.; Kumar, R.A.; Manimekalai, R.; Vasantha, S.; Krishnappa, G.; Kumar, R.; Pandey, S.K.; Hemaprabha, G. Recent Advances in Sugarcane Genomics, Physiology, and Phenomics for Superior Agronomic Traits. *Front. Genet.* **2022**, *13*, 854936. [CrossRef] [PubMed]
18. VanHook, A.M. Nitrogen assimilation gets a HY5. *Sci. Signal.* **2016**, *9*, ec59. [CrossRef]
19. Sulistyo, S.B.; Wu, D.; Woo, W.L.; Dlay, S.S.; Gao, B. Computational Deep Intelligence Vision Sensing for Nutrient Content Estimation in Agricultural Automation. *IEEE Trans. Autom. Sci. Eng.* **2018**, *15*, 1243–1257. [CrossRef]
20. You, H.; Zhou, M.; Zhang, J.; Peng, W.; Sun, C. Sugarcane nitrogen nutrition estimation with digital images and machine learning methods. *Sci. Rep.* **2023**, *13*, 14939. [CrossRef] [PubMed]
21. Sun, L.; Yang, C.; Wang, J.; Cui, X.; Suo, X.; Fan, X.; Ji, P.; Gao, L.; Zhang, Y. Automatic Modeling Prediction Method of Nitrogen Content in Maize Leaves Based on Machine Vision and CNN. *Agronomy* **2024**, *14*, 124. [CrossRef]
22. Xu, G.; Zhang, F.; Shah, S.G.; Ye, Y.; Mao, H. Use of leaf color images to identify nitrogen and potassium deficient tomatoes. *Pattern Recognit. Lett.* **2011**, *32*, 1584–1590. [CrossRef]
23. Xiong, X.; Zhang, J.; Guo, D.; Chang, L.; Huang, D. Non-Invasive Sensing of Nitrogen in Plant Using Digital Images and Machine Learning for *Brassica Campestris* ssp. *Chinensis* L.. *Sensors* **2019**, *19*, 2448. [CrossRef]
24. Ahmad, M.U.; Ashiq, S.; Badshah, G.; Khan, A.H.; Hussain, M.; Sarfraz, S. Feature Extraction of Plant Leaf Using Deep Learning. *Complexity* **2022**, *2022*, 6976112. [CrossRef]
25. Lee, K.-J.; Lee, B.-W. Estimation of rice growth and nitrogen nutrition status using color digital camera image analysis. *Eur. J. Agron.* **2013**, *48*, 57–65. [CrossRef]
26. Sun, Y.; Tong, C.; He, S.; Wang, K.; Chen, L. Identification of Nitrogen, Phosphorus, and Potassium Deficiencies Based on Temporal Dynamics of Leaf Morphology and Color. *Sustainability* **2018**, *10*, 762. [CrossRef]
27. Bo, H.; Ze, Z.; Qiang, Z.; Yiru, M.; Xiang, Y.; Xin, L. The Nitrogen Content in Cotton Leaves: Estimation Based on Digital Image. *Chin. Agric. Sci. Bull.* **2022**, *38*, 49–55. [CrossRef]
28. Yang, H.; Li, G.; Ma, J.; Wang, H.; Yang, J.; Yang, J. Diagnose Leaf Nutrition Level of Red Delicious Apple with Image Digital. *Gansu Agric. Sci. Technol.* **2022**, *53*, 59–63. [CrossRef]
29. Barman, U.; Saikia, M.J. Smartphone Contact Imaging and 1-D CNN for Leaf Chlorophyll Estimation in Agriculture. *Agriculture* **2024**, *14*, 1262. [CrossRef]
30. Kamboj, A.; Khokhar, K.K.; Chand, M.; Vikas; Kumar, S.; Singh, U.; Rani, M. Assessment of Method and Application Schedule of Fertilizer N and K on Growth and Productivity of Summer Planted Sugarcane Crop (*Saccharum officinarum* L.) under Wide Spacing. *Int. J. Plant Soil Sci.* **2023**, *35*, 34–46. [CrossRef]
31. Xie, S.N.; Girshick, R.; Dollár, P.; Tu, Z.W.; He, K.M. Aggregated Residual Transformations for Deep Neural Networks. In Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5987–5995.
32. Woo, S.H.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the 15th European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
33. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017.

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.