



# Article Maize Disease Classification System Design Based on Improved ConvNeXt

Han Li <sup>1,2,†</sup>, Mingyang Qi <sup>1,†</sup>, Baoxia Du <sup>1</sup>, Qi Li <sup>1</sup>, Haozhang Gao <sup>1</sup>, Jun Yu <sup>2</sup>, Chunguang Bi <sup>3</sup>, Helong Yu <sup>3</sup>, Meijing Liang <sup>4</sup>, Guanshi Ye <sup>1,\*</sup> and You Tang <sup>1,2,\*</sup>

- <sup>1</sup> Electrical and Information Engineering College, Jilin Agricultural Science and Technology University, Jilin 132101, China
- <sup>2</sup> School of Information and Control Engineering, Jilin Institute of Chemical Technology, Jilin 132022, China
- <sup>3</sup> College of Information Technology, Jilin Agricultural University, Changchun 130118, China
- <sup>4</sup> Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164, USA
- Correspondence: yeguanshi@163.com (G.Y.); tangyou9000@163.com (Y.T.)
- <sup>+</sup> These authors contributed equally to this work.

Abstract: Maize diseases have a great impact on agricultural productivity, making the classification of maize diseases a popular research area. Despite notable advancements in maize disease classification achieved via deep learning techniques, challenges such as low accuracy and identification difficulties still persist. To address these issues, this study introduced a convolutional neural network model named Sim-ConvNeXt, which incorporated a parameter-free SimAM attention module. The integration of this attention mechanism enhanced the ability of the downsample module to extract essential features of maize diseases, thereby improving classification accuracy. Moreover, transfer learning was employed to expedite model training and improve the classification performance. To evaluate the efficacy of the proposed model, a publicly accessible dataset with eight different types of maize diseases was utilized. Through the application of data augmentation techniques, including image resizing, hue, cropping, rotation, and edge padding, the dataset was expanded to comprise 17,670 images. Subsequently, a comparative analysis was conducted between the improved model and other models, wherein the approach demonstrated an accuracy rate of 95.2%. Notably, this performance represented a 1.2% enhancement over the ConvNeXt model and a 1.5% improvement over the advanced Swin Transformer model. Furthermore, the precision, recall, and F1 scores of the improved model demonstrated respective increases of 1.5% in each metric compared to the ConvNeXt model. Notably, using the Flask framework, a website for maize disease classification was developed, enabling accurate prediction of uploaded maize disease images.

Keywords: CNN; SimAM attention; transfer learning; data augmentation; maize disease classification

## 1. Introduction

With the rapid advancement of artificial intelligence technology, current research is focused on addressing a critical issue: how to utilize deep learning techniques to rapidly and accurately identify and classify maize diseases. These diseases have a direct impact on maize yield and quality, making them crucial for agriculture and the food supply chain. This study aims to provide an efficient solution to promote the sustainable development of the maize industry.

Artificial intelligence is quickly becoming a major driving force shaping various facets of modern life. It is increasingly being used in a variety of fields, including agriculture, healthcare, and many other broad sectors, resulting in revolutionary advances [1]. In the realm of agriculture, maize cultivation encompasses a vast global expanse, with maize being recognized as one of the major crops and agricultural commodities. As a vital staple food and industrial raw material, the stable and sustainable advancement of the maize industry is crucial for ensuring food security, augmenting farmers' income, and



Citation: Li, H.; Qi, M.; Du, B.; Li, Q.; Gao, H.; Yu, J.; Bi, C.; Yu, H.; Liang, M.; Ye, G.; et al. Maize Disease Classification System Design Based on Improved ConvNeXt. *Sustainability* **2023**, *15*, 14858. https://doi.org/10.3390/su152014858

Academic Editors: Ali A. Alwan Al-juboori and Yonis Gulzar

Received: 26 August 2023 Revised: 7 October 2023 Accepted: 12 October 2023 Published: 13 October 2023



**Copyright:** © 2023 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/). fostering national economic progress [2]. Nevertheless, the quality and yield of maize are directly affected by various maize diseases, including dwarf leaf disease [3], gray leaf spot [4], rust [5], and leaf spot [6]. When maize is affected by these diseases, its overall physiological functions are significantly impaired, resulting in stunted plant growth and an inability to attain optimal growth conditions. Consequently, this adversely affects both yield and economic returns [7]. In this context, employing artificial intelligence and precise classification techniques, particularly methods based on image classification, can enhance the speedy diagnosis of maize diseases. This contributes to faster decisions in agricultural production, mitigating the impact of diseases on crop yield and quality and achieving a sustainable development of the maize industry [8].

Existing maize disease classification strategies suffer from certain limitations. Traditional classification methods exhibit low accuracy due to the diversity and large quantity of disease types. They are also influenced by subjective factors, consuming significant time and labor [9]. Additionally, many maize diseases share similar appearances, posing challenges for differentiation and exacerbating the difficulty of classification. Furthermore, the limitations of conventional image processing techniques make it difficult to identify images with complex texture structures, color variations, and shape changes [10]. These methods also require high-quality images to yield accurate results.

By optimizing feature extraction and classification algorithms, traditional machine learning algorithms, such as Support Vector Machine (SVM) [11], decision trees [12], K-Nearest Neighbors (KNN) [13], and random forests [14], have been utilized to achieve accurate recognition and classification of maize diseases. Noola et al. [15] proposed an enhanced KNN model aimed at distinguishing different categories of diseases. The model exhibited excellent performance in terms of precision, recall, and F1 score. However, it had not adequately considered its ability to generalize in complex environments. Kusumo et al. [16] investigated various maize disease features based on image processing and employed machine learning algorithms to assess the performance of these features. The findings indicated that RGB features exhibited the highest accuracy in classification among most classifiers. However, limitations persisted when confronted with the complex issue of multiple concurrent diseases affecting maize plants.

In recent years, advancements in deep learning techniques and maize disease detection have given rise to a prominent research area: deep learning-based maize disease classification. Convolutional neural networks (CNN) [17,18] are commonly used models in image classification tasks. Liu et al. [19] proposed a transfer learning-based fine-tuning approach adapted from the EfficientNet model. The last layer of the EfficientNet classifier was replaced with an 8-class softmax classifier, and validation was conducted using VGG16, InceptionV3, and Resnet50 architectures. The results demonstrated that the optimized model achieved significantly higher accuracy compared to other networks. It was worth noting that a small-sample dataset had a certain impact on the generalization ability of the model. Sun et al. [20] proposed a multi-scale feature fusion instance detection method for maize leaf blight based on CNN. This method established a connection between fine-tuning network and the detection module and replaced the loss function with Generalized Intersection over Union (GIoU) to achieve improved accuracy and detection speed. This method primarily focused on detecting small target diseases, but it encountered challenges when there were significant variations in target size and density. Haque et al. [21] employed rotation augmentation and brightness enhancement techniques to address the issue of class imbalance while utilizing a benchmark training approach to train the Inception-v3 network. The findings demonstrated that the optimized model outperformed several other models in performance. Due to the small dataset, the generalization capability of the model was limited.

Previous research encountered several challenges in the classification of maize diseases, including limitations in model generalization performance and insufficient adaptability to complex environments. To address these issues, this study designs and implements a maize disease classification system and evaluates its classification performance. The improved

ConvNeXt [22] model is employed for the classification of maize diseases, which incorporates a SimAM attention module after the downsample stage to enhance the model focus on crucial information and improve its performance. Additionally, the ConvNeXt model is implemented using the PyTorch framework, and model training is conducted using the AdamW [23] optimizer. Transfer learning is employed to accelerate the model training process. Furthermore, multiple data augmentation techniques are utilized to expand the dataset, thereby improving the model generalization ability. Comparative experiments are conducted on the ConvNeXt, ResNet34 [24], ResNeXt50 [25], DenseNet121 [26], MobileNetV2 [27], Vision Transformer [28], and Swin Transformer [29] models. The results demonstrate the superior classification performance of the proposed model. Accurate and timely maize disease classification is crucial for effective disease management and higher crop productivity. The proposed system has the potential to significantly reduce crop losses and enhance agricultural practices.

The primary contributions of this study are as follows:

- This study employs the ConvNeXt model, a pure convolutional neural network, for the extraction of features from maize disease images. To enhance the performance of the downsample module, a parameter-free SimAM [30] attention module is introduced, which improves the model's ability to extract crucial features and reduces the risk of overfitting, effectively enhancing maize diseases classification performance.
- 2. The incorporation of transfer learning into the ConvNeXt model enhances its applicability for maize disease classification tasks. This approach leads to improvements in accuracy, generalization performance, and training efficiency, thereby enhancing the usability of the model in practical applications.
- 3. In response to the limitations encountered in practical applications of maize disease classification, this study develops a website utilizing the Flask framework. The website allows users to conveniently upload relevant disease images for efficient classification, significantly lowering manual efforts and associated costs.

The structure of this study is as follows: Section 2 provides a detailed description of the methods employed. Section 3 presents experiments and results. Section 4 discusses the research findings, potential improvements, and future research directions. Section 5 summarizes the main contributions and their impact on the field of agriculture.

#### 2. Methods

## 2.1. The Overall Design of SimAM-ConvNeXt Model

In this study, the aim is to enhance the effectiveness of the ConvNeXt model by incorporating an attention mechanism. To identify the most suitable attention mechanism for this purpose, a comprehensive comparison of several popular attention mechanisms was conducted, including CBAM [31], NAM [32], SE, and SimAM. It found that the SimAM attention mechanism yielded the best results. Figure 1 illustrates the architecture of the proposed Sim-ConvNeXt model.

Next, a detailed explanation of different components of the model will be provided, primarily encompassing the ConvNeXt module, attention module, and downsampling module.

## 2.2. ConvNeXt-T Model

In this study, we employed ConvNeXt-T, a lightweight version derived from ConvNeXt and designed exclusively for image classification tasks. The model demonstrates outstanding feature extraction capabilities by capturing complex patterns and features within images through a series of convolutional layers. Furthermore, it exhibits robust generalization capabilities, allowing it to capture features at various image scales, thereby enhancing its adaptability to images of different sizes and shapes. Built upon the ResNet50 architecture, ConvNeXt-T incorporates structural elements inspired by models such as Swin Transformer, ResNeXt, and MobileNetV2. Table 1 presents the detailed architecture of the ConvNeXt model.



Figure 1. Design of Sim-ConvNeXt model.

Table 1. ConvNeXt-T structure information.

Layer	ConvNeXt-T	Input	Output
conv1	k4, s4, dim = 96	$224\times224\times3$	$56 \times 56 \times 96$
conv2_x	$\begin{bmatrix} d7 \times 7,96\\ 1 \times 1,384\\ 1 \times 1,96 \end{bmatrix} \times 3$	$56 \times 56 \times 96$	$56 \times 56 \times 96$
conv3_x	Downsample $\begin{bmatrix} d7 \times 7, 192 \\ 1 \times 1, 768 \\ 1 \times 1, 192 \end{bmatrix} \times 3$	$56 \times 56 \times 96$	$28\times 28\times 192$
conv4_x	Downsample $\begin{bmatrix} d7 \times 7,384 \\ 1 \times 1,1536 \\ 1 \times 1,384 \end{bmatrix} \times 9$	$28 \times 28 \times 192$	$14\times14\times384$
conv5_x	Downsample $\begin{bmatrix} d7 \times 7,768\\ 1 \times 1,3072\\ 1 \times 1,768 \end{bmatrix} \times 3$	$14 \times 14 \times 384$	$7 \times 7 \times 768$
	Global Avg Pooling Layer Norm Linear	$7 \times 7 \times 768$	1000

Blocks are stacked in the Swin Transformer's four stages in the ratio of 1:1:3:1, with Stage 3 having the most weight. To align with this structure, the block stacking ratio in ConvNeXt is adjusted from 3:4:6:3 to 3:3:9:3. The Swin Transformer employs convolutional layers with a  $4 \times 4$  kernel size and a stride of 4, resulting in a downsampling factor of 4. The downsampling module in ConvNeXt is accordingly modified to accommodate this structure.

In addition, ConvNeXt incorporates a reverse bottleneck module, which primarily draws inspiration from the MobileNetV2 model. Figure 2 shows the module construction, which uses two convolutional kernels of varying sizes to widen the receptive field and minimize the number of parameters.



Figure 2. Inverted bottleneck design in MobileNetV2 and ConvNeXt. (a) MobileNetV2; (b) ConvNeXt.

In Transformer models, the Gaussian Error Linear Unit (GELU) [33] activation function is commonly used (as shown in Figure 3a), while in convolutional neural networks, the Rectified Linear Unit (RELU) [34] activation function is often employed. The GELU activation function demonstrates outstanding performance in a variety of natural language processing tasks and image classification tasks. It possesses a smooth nonlinear characteristic that enables the neural network to learn more intricate patterns. Compared to RELU, GELU generates non-zero outputs for negative input values, allowing for better handling of negative inputs. To enhance the performance and efficiency of the ConvNeXt model, several improvements were implemented. These include the adoption of GELU as the activation function and a reduction in the number of activation functions used. Additionally, the Batch Normalization (BN) [35] layer was replaced with the Layer Normalization (LN) [36] layer (as shown in Figure 3b). Different from the BN layer, the LN layer normalizes each feature within every sample's data. By computing the mean and standard deviation of each feature, the LN layer applies linear transformation and scaling to adjust the distribution of features, rendering it better suited for normalizing sequence data and individual samples.



**Figure 3.** Micro-design of Swin Transformer and ConvNeXt. (**a**) Swin Transformer block; (**b**) ConvNeXt block.

## 2.3. SimAM Attention Mechanism

The attention mechanism is extensively employed in classification tasks as it effectively leverages the information within input data, improving the performance of classification models as well as providing interpretability and comprehension for the decision-making process of the model. SimAM is a conceptually simple yet highly effective attention mechanism module. It enhances feature extraction by focusing on crucial information in the downsampling module. This attention mechanism minimizes the risk of overfitting while effectively capturing discriminative features. The fundamental concept revolves around the integration of a straightforward attention module into a CNN architecture to replace additional learnable parameters. This approach enables seamless integration while maintaining lightweight and efficient computations and minimizing memory usage. Currently, most attention mechanisms primarily focus on channel attention and spatial attention modules. Channel attention can be understood as guiding the neural network to pay attention to specific positions, with SENet [37] being a representative method. It employs a modeling approach to assess the relative significance of various feature channels and subsequently applies channel enhancement or suppression techniques tailored to specific tasks. In a convolutional network, each convolutional kernel corresponds to a feature channel and focuses on allocating resources among them, as shown in Figure 4a. Spatial attention, on the other hand, focuses on the most important parts of the network without considering every part of the image. It can transform the spatial information of the original image to another space while preserving key information. Representative models include STN [38], as shown in Figure 4b. In contrast to the traditional spatial and channel attention mechanisms, the SimAM attention module used in this study can directly estimate 3D weights. It proposes an energy function based on neurons, where the weight of attention is calculated by estimating the importance of individual neurons, as illustrated in Figure 4c. The energy function for neurons is defined as follows:

$$e_t(w_t, b_t, y, x_i) = \frac{1}{M-1} \sum_{i=1}^{M-1} \left( -1 - (w_t x_i + b_t) \right)^2 + \left( 1 - (w_t t + b_t) \right)^2 + \lambda w_t^2$$
(1)

the  $w_t$  and  $b_t$  are the weights and biases of neuron transformation, y is a scalar quantity, t and  $x_i$  are the target neuron and other neurons of the input feature X,  $X \in \mathbb{R}^{C \times H \times W}$ , i is the index of a neuron in a specific channel, M is the number of neurons on that channel, and  $M = H \times W$ ,  $\lambda$  is the regularization coefficient.

#### 2.4. Improved Downsample Module

The structure of the downsample module in the ConvNeXt model is depicted in Figure 5a. It begins by normalizing the input using the LN technique, followed by feature extraction and transformation through a convolutional layer. The enhanced downsample module, as illustrated in Figure 5b, incorporated the Sigmoid activation function and the SimAM attention mechanism.

## 2.5. Design of Maize Disease Classification System

Currently, relevant research exists on the classification of maize diseases; however, practical applications in this domain remain relatively limited. In this study, we developed a web-based application for the classification of maize diseases, as depicted in Figure 6. Users could upload an image of a maize disease, and the system automatically recognized and categorized it into different disease types. The application was built using the Flask framework, allowing users to access the website through a web browser. The backend of the application utilized improved ConvNeXt model weights for training and generated the classification results.



**Figure 4.** Three types of attention modules. (**a**) Channel-wise attention; (**b**) Spatial-wise attention; (**c**) Full 3D weights for attention.



Figure 5. Improved downsample module. (a) Downsample module; (b) Improved downsample module.



Figure 6. Maize disease classification system.

## 3. Experiments and Results

To evaluate the effectiveness of the proposed model in classifying maize diseases, the model weights were first initialized by pretraining them on the ImageNet-1K dataset. Subsequently, training was conducted on the maize disease dataset, comparing the enhanced ConvNeXt model against ResNet34, ResNeXt50, MobileNetV2, DenseNet121, Vision Transformer, and Swin Transformer models. Comparative analysis of the experimental results was performed to assess the performance of the ConvNeXt model. The main steps of this experiment include data preprocessing, feature extraction, model training, and model evaluation.

## 3.1. Dataset and Augmentation

A publicly available maize disease dataset was utilized, and detailed information is presented in Table 2. There are eight different types of maize disease shown in the dataset, including dwarf leaf, health, gray, severe gray, rust, severe rust, leaf spot, and severe leaf spot, with a total of 3534 images.

Classes	Before	After	Train	Val	Test
Dwarf leaf	931	4655	2793	931	931
Healthy	430	2150	1290	430	430
Gray	191	955	573	191	191
Severe gray	218	1090	654	218	218
Rust	552	2760	1656	552	552
Severe rust	406	2030	1218	406	406
Leaf spot	237	1185	711	237	237
Severe leaf spot	569	2845	1707	569	569
Total	3534	17,670	10,602	3534	3534

Table 2. Dataset information.

The dataset was enriched using five data augmentation techniques, including resizing, hue, cropping, rotation, and edge padding, resulting in 17,670 images. Figure 7 demonstrates the effect of data augmentation. After data preprocessing, the dataset was partitioned into training, validation, and testing sets with a ratio of 6:2:2. Additionally, the image size was adjusted to  $224 \times 224$  to accommodate the structure of the model.



Figure 7. Data augmentation. (a) original; (b) resize; (c) hue; (d) crop; (e) rotate; and (f) padding.

#### 3.2. Experimental Environment

The experimental configuration used the PyTorch deep learning framework and was run on an Ubuntu computer. The experiments were carried out using an NVIDIA RTX3090 GPU. To ensure the validity of the experimental outcomes, each model was configured with identical hyperparameters, including a fixed number of epochs (200) and a batch size of 64.

## 3.3. Feature Extraction

Firstly, the ConvNeXt model extracted image features through a series of convolutional layers. These convolutional layers used different kernels to capture various features in the input image, including edges, textures, and higher-level semantic information. In this way, the model gradually built an abstract representation of the image. Next, during the stacking process of feature maps, the model progressively extracted and combined these features, making the feature maps more abstract and advanced. This meant that the model started focusing on more abstract image features, not just simple edges and textures. Then, through a downsampling module, the model reduced the spatial size of the feature maps while still retaining important feature information. This reduction in spatial dimensions contributed to reduced computational complexity, making the model more suitable for handling input images of varying sizes. Finally, the ConvNeXt model utilized the SimAM attention mechanism to enhance its focus on key information in images of corn diseases.

ConvNeXt extracted features through a series of convolutional layers, incorporating attention mechanisms like the SimAM attention module to enhance feature extraction. The network gradually processed the input image, captured features at different scales, and aggregated them to form high-level feature representations. These representations were then used to make predictions in various image classification tasks. This architecture allowed ConvNeXt to efficiently learn and represent complex patterns in images, making it suitable for tasks like maize disease classification.

#### 3.4. Model Training

After extracting features from maize disease images, the ConvNeXt model underwent training. In order to evaluate the performance and generalization capability of the models, a comparative analysis was conducted on the validation accuracies of different models. Figure 8 depicts the trend of validation accuracy of the model as the number of training epochs changes. According to the figure, the proposed model achieved an accuracy of 95.7%, followed by ConvNeXt with an accuracy of 95.0%, while MobileNetV2 exhibited the lowest accuracy, at only 82.5%. It is evident from the figure that the proposed model achieved a high validation accuracy, indicating its strong performance.



Figure 8. Validation accuracy performance in different model training.

## 3.5. Model Evaluation

By training the models to obtain optimal weights for each model, this study proceeded to conduct testing and analysis on the designated test set. In the context of this investigation on maize disease classification, common model evaluation metrics comprise accuracy, recall, precision, and F1 score. Additionally, the utilization of a confusion matrix allows for further analysis of the model's classification performance and misclassifications. The four elements in the confusion matrix are defined as follows in the context of the study: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN).

Accuracy, in the context of model evaluation, refers to the proportion of correctly classified samples out of the total number of samples. In the case of multi-classification problems, it is common to assess performance using macro-averaged accuracy and microaveraged accuracy. It helps us understand how the model performs in classification. Precision, also known as positive predictive value, quantifies the proportion of true positive samples identified by a model out of all samples identified as positive by the model. A higher precision indicates that the model identification of positive samples is more accurate, thereby exhibiting superior accuracy. It can minimize false-positive predictions, thus reducing unnecessary actions. The recall rate is a metric that quantifies the proportion of true positive samples identified by a model out of all the true samples. A higher recall rate indicates that the model is capable of accurately identifying true positive samples, thereby demonstrating superior discriminatory ability. It helps us understand the ability of the model to capture all maize disease cases, ensuring that the model doesn't miss any potential disease instances. The F1 score is a metric that comprehensively evaluates the balance between model accuracy and recall. A higher F1 score indicates that the model has achieved a better balance between accuracy and recall. It ensures that the model performs well in accurately identifying diseases and capturing all instances. The formulas for calculating these metrics are as follows:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$
(2)

$$Precision = \frac{TP}{(TP + FP)}$$
(3)

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{4}$$

$$F1-Score = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)}$$
(5)

#### 3.5.1. Utilization of Transfer Learning

Transfer learning can expedite model training and enhance prediction accuracy. It involves initially pretraining the model on a large-scale image classification task and then applying the learned features and weights to a specific task. This approach leverages existing knowledge, avoiding the need to train the model from scratch, thereby improving both efficiency and performance. To investigate the impact of transfer learning on classification outcomes, a classification experiment was conducted comparing the performance of the ConvNeXt model with and without the application of transfer learning. The experimental results, presented in Table 3, demonstrate that the employment of transfer learning yields a notable increase in accuracy, reaching 94.0%, compared to 87.0% achieved without transfer learning. Furthermore, precision, recall, and F1 score also exhibit significant improvements, indicating that the utilization of transfer learning enhances classification performance and effectiveness.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Non-Use	87.0	83.2	82.5	82.8
Use	94.0	92.4	91.8	92.1

Table 3. The impact of transfer learning on classification results.

3.5.2. Performance Comparison of Different Attention Modules

In order to investigate the impact of various attention modules on classification outcomes, CBAM, SE, NAM, and SimAM modules were incorporated into the ConvNeXt model, which underwent transfer learning. The experimental findings, as presented in Table 4, demonstrate that the inclusion of the SimAM module yielded the most favorable results, with an accuracy rate of 95.2%. Furthermore, this module exhibited superior performance across other metrics compared to the alternative modules.

Table 4. The impact of different attention modules on classification results.

Attention Module	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
+CBAM	93.4	91.4	90.6	91.0
+SE	93.9	92.5	91.3	91.9
+NAM	94.4	93.0	92.2	92.6
+SimAM(Ours)	95.2	93.9	93.3	93.6

## 3.5.3. Performance Comparison of Different Models

After training the improved ConvNeXt model, the evaluation of classification performance was conducted using a test set consisting of 3534 images consisting of 8 disease categories. A comparison was made between the enhanced ConvNeXt model and the ResNet34, ResNeXt50, MobileNetV2, DenseNet121, ViT, and Swin-T models. Table 5 presents the results, indicating that the model achieved an accuracy of 95.2%, a precision of 93.9%, a recall of 93.3%, and an F1 score of 93.6%.

Table 5. Classification performance of different models.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
ResNet34	92.0	89.7	89.0	89.3
ResNeXt50	92.8	90.9	90.2	90.5
MobileNetV2	81.4	76.0	72.9	74.4
DenseNet121	91.4	89.0	88.1	88.5
VIT	89.3	85.9	85.4	85.6
Swin-T	93.7	92.0	91.2	91.6
ConvNeXt-T	94.0	92.4	91.8	92.1
Sim-ConvNeXt	95.2	93.9	93.3	93.6

Figure 9 shows that the model exceled in all evaluation metrics, while MobileNetV2 exhibited a comparatively weaker performance in these metrics. These results demonstrate the superior performance of the proposed model compared to the other models and the unimproved ConvNeXt model.

Furthermore, a confusion matrix was employed to further evaluate the classification performance of the model. Figure 10 illustrates the confusion matrix results of the proposed model in comparison to seven other models. The proposed model exhibited better classification performance in five categories: healthy, gray, severe gray, severe rust, and severe leaf spot. However, the model demonstrated slightly lower effectiveness in classifying general symptoms of rust disease and leaf spot disease when compared to the Swin-T model. Although the model did not produce the best results across all categories, the overall classification performance was still impressive.



Figure 9. Visualizing the classification performance of different models.

## 3.6. Maize Disease Classification System

The model was installed on a web-based platform built with the Flask framework to make it more suitable for actual applications. This system enables rapid and accurate identification of maize diseases, thereby improving the efficiency and precision of disease recognition. This efficient recognition capability aids in the reduction of pesticide overuse, reducing negative impacts on the environment and ecosystems while also achieving the goals of sustainable agriculture. The website has been successfully deployed on a server, primarily implementing the functionality of uploading and classifying disease images. Users can access the website directly by visiting http://www.maize.love:4997/ (accessed on 5 August 2023). Figure 11 illustrates the user interface, displaying the predicted results of the maize disease system. Users upload the image requiring identification and then select the "predict" option to obtain the disease classification. The classification results can be obtained in around 260 ms. In Figure 11, the result "Rust" is displayed with a probability of 1.



Figure 10. Cont.





2

2

135

0

0

13

True Labels

0

175 1

0

11 0

(c)

0 1 0

0 0

31

dwarf leaf

severe gray

severe rust

severe leaf spot

leaf spot

dwarf leaf

severe gray

severe rust

leaf spot 1 0

healthy

gray

rust

Predicted Labels

Predicted Labels

healthy

grav

rust 0

928 9 1

٥

0

974 1

0 0 142 27

0 106 30 0

З

463

82

Figure 10. Confusion matrices of different models. (a) ResNet34; (b) ResNetX50; (c) MobileNetV2; (d) DenseNet121; (e) ViT; (f) Swin-T; (g) ConvNeXt-T; and (h) Sim-ConvNeXt.

This web-based platform allows farmers to identify diseases within seconds and receive classification results promptly. Compared to manual methods, the website enables faster classification and reduces the impact of subjective human judgments. It assists farmers in promptly addressing diseases. This real-time capability and convenience assist farmers in making faster agricultural production decisions, thus lowering the impact of diseases on crop yield and quality, achieving a more sustainable agricultural development.

upload image:	predict:	
1 9 Sec.	class:rust	probability:1.000
10 / 10 M	class:severe rust	probability:0.000
	class:leaf spot	probability:0.000
Contraction of the second	class:dwarf leaf	probability:0.000
Contraction of the second	class:severe leaf sp	ot probability:0.000
	class:severe gray	probability:0.000
	class:healthy	probability:0.000
	class:gray	probability:0.000

Figure 11. Maize disease prediction results.

#### 4. Discussion

Although the Sim-ConvNeXt model has achieved good performance in the classification of maize diseases by introducing the SimAM attention mechanism, there is still room for improvement. On one hand, optimizing data augmentation methods and model architecture will make the model more robust and adaptive to various data scenarios, thereby providing more accurate diagnostic tools for agricultural production. On the other hand, expanding the application scope of the model to other image classification tasks, such as plant disease classification [39] and medical image classification [40], will further elevate its technological proficiency in these domains, contributing to the sustainable development of agriculture and the medical industry.

Apart from the Sim-ConvNeXt model itself, the web-based classification system also needs improvements. Currently, it provides classification of several maize disease types, necessitating retraining and optimization for expanding its applicability to other disease types. This will assist farmers in better identifying and managing diverse disease types, decreasing pesticide use and minimizing resource waste. There is a need to improve the user experience and interaction modes, facilitating a more convenient and speedy usage of the system for classification and diagnosis. Extensive research also needs to be conducted to explore treatment methods, preventive strategies, and specific diseases.

This will not only provide farmers with more valuable information but also foster knowledge sharing and collaboration, offering more comprehensive support for the sustainable development of agriculture. These improvements and research efforts are expected to lead to a more sustainable growth in the agricultural sector, leading to increased economic, social, and environmental benefits.

## 5. Conclusions

This study focused on the classification system of maize diseases and proposed a convolutional neural network model called Sim-ConvNeXt. The SimAM attention modules were integrated after each downsampling module; transfer learning was employed to expedite the model training process and mitigate overfitting. By incorporating the SimAM attention mechanism and transfer learning, the Sim-ConvNeXt model demonstrated improved accuracy in maize disease identification, enhancing both classification precision and performance.

Due to the limited size of the original dataset, multiple data augmentation techniques were employed to expand the dataset, resulting in a total of 17,670 disease images. Subsequently, the enhanced ConvNeXt model was experimentally compared with seven other models. The results demonstrated the superior classification performance of the proposed model, with accuracy, precision, recall, and F1 score reaching 95.2%, 93.9%, 93.3%, and 93.6%, respectively. These values indicated improvements of 1.2%, 1.5%, 1.5%, and 1.5% compared to the original model. Additionally, the performance of the model across different disease categories was analyzed using a confusion matrix, confirming its superior efficacy. Furthermore, a user-friendly website for maize disease recognition was constructed using the Flask framework, enabling the classification of uploaded disease images. The proposed maize disease classification system holds significant importance because it has the potential to enhance disease management and increase crop yields. This system can rapidly and accurately identify diseases, reducing crop damage and improving agricultural efficiency, thereby reducing costs and enhancing agricultural decision-making.

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

**Funding:** This research was supported by the Jilin Province Science nd Technology Development Program Project (Project No. YDZJ202201ZYTS692) and the Project of the Doctoral Initial Scientific Research Fund Supported by Jilin Agricultural Science and Technology University (No.2023706).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

**Data Availability Statement:** The dataset analyzed during the current research was selected from open-source information available online: https://aistudio.baidu.com/aistudio/datasetdetail/11 1048 (accessed on 7 June 2023). The maize disease classification system website address: http://www.maize.love:4997/ (accessed on 5 August 2023).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Gulzar, Y. Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique. Sustainability 2023, 15, 1906. [CrossRef]
- Li, F.; Li, Y.; Novoselov, K.S.; Liang, F.; Meng, J.; Ho, S.-H.; Zhao, T.; Zhou, H.; Ahmad, A.; Zhu, Y.; et al. Bioresource Upgrade for Sustainable Energy, Environment, and Biomedicine. *Nano-Micro Lett.* 2023, 15, 35. [CrossRef] [PubMed]
- Kannan, M.; Ismail, I.; Bunawan, H. Maize Dwarf Mosaic Virus: From Genome to Disease Management. *Viruses* 2018, 10, 492. [CrossRef] [PubMed]
- 4. Dhami, N.B.; Kim, S.K.; Paudel, A.; Shrestha, J.; Rijal, T.R. A Review on Threat of Gray Leaf Spot Disease of Maize in Asia. *J. Maize Res. Dev.* 2015, 1, 71–85. [CrossRef]
- Olukolu, B.A.; Tracy, W.F.; Wisser, R.; De Vries, B.; Balint-Kurti, P.J. A genome-wide association study for partial resistance to maize common rust. *Phytopathology* 2016, 106, 745–751. [CrossRef] [PubMed]
- 6. Sun, X.; Qi, X.; Wang, W.; Liu, X.; Zhao, H.; Wu, C.; Chang, X.; Zhang, M.; Chen, H.; Gong, G. Etiology and Symptoms of Maize Leaf Spot Caused by *Bipolaris* spp. in Sichuan, China. *Pathogens* **2020**, *9*, 229. [CrossRef]
- Du, H.P.; Fang, C.; Li, Y.R.; Kong, F.J.; Liu, B.H. Understandings and future challenges in soybean functional genomics and molecular breeding. J. Integr. Plant Biol. 2023, 65, 468–495. [CrossRef]
- Spiertz, H. Challenges for Crop Production Research in Improving Land Use, Productivity and Sustainability. Sustainability 2013, 5, 1632–1644. [CrossRef]
- 9. Wu, Y.; Wang, D.H.; Lu, X.T.; Yang, F.; Yao, M.; Dong, W.S.; Shi, J.B.; Li, G.Q. Efficient Visual Recognition: A Survey on Recent Advances and Brain-inspired Methodologies. *Mach. Intell. Res.* **2022**, *19*, 366–411. [CrossRef]
- Li, X.F.; Liu, B.; Zheng, G.; Ren, Y.B.; Zhang, S.S.; Liu, Y.J.; Gao, L.; Liu, Y.H.; Zhang, B.; Wang, F. Deep-learning-based information mining from ocean remote-sensing imagery. *Natl. Sci. Rev.* 2020, *7*, 1584–1605. [CrossRef]
- Gangsar, P.; Tiwari, R. A support vector machine based fault diagnostics of Induction motors for practical situation of multi-sensor limited data case. *Measurement* 2019, 135, 694–711. [CrossRef]
- 12. Kotsiantis, S.B. Decision trees: A recent overview. Artif. Intell. Rev. 2013, 39, 261–283. [CrossRef]

- 13. Liang, Y.; Li, K.J.; Ma, Z.; Lee, W.J. Multilabel classification model for type recognition of single-phase-to-ground fault based on KNN-bayesian method. *IEEE Trans. Ind. Appl.* **2021**, *57*, 1294–1302. [CrossRef]
- Li, X.R.; Lin, Y.T.; Qiu, K.B. Stellar spectral classification and feature evaluation based on a random forest. *Res. Astron. Astrophys.* 2019, 19, 56–62. [CrossRef]
- 15. Noola, D.A.; Basavaraju, D.R. Corn leaf image classification based on machine learning techniques for accurate leaf disease detection. *Int. J. Electr. Comput. Eng.* 2022, *12*, 2509–2516. [CrossRef]
- Kusumo, B.S.; Heryana, A.; Mahendra, O.; Pardede, H.F. Machine learning-based for automatic detection of corn-plant diseases using image processing. In Proceedings of the 2018 International Conference on Computer, Control, Informatics and Its Applications (IC3INA), Tangerang, Indonesia, 1–2 November 2018; pp. 93–97.
- Alahmari, F.; Naim, A.; Alqahtani, H. E-Learning Modeling Technique and Convolution Neural Networks in Online Education. In *IoT-Enabled Convolutional Neural Networks: Techniques and Applications*, 1st ed.; Naved, M., Devi, V.A., Gaur, L., Elngar, A.A., Eds.; River Publishers: New York, NY, USA, 2023; pp. 261–295.
- 18. Krichen, M. Convolutional Neural Networks: A Survey. Computers 2023, 12, 151. [CrossRef]
- 19. Liu, J.C.; Wang, M.T.; Bao, L.; Li, X.F. EfficientNet based recognition of maize diseases by leaf image classification. *J. Phys. Conf. Ser.* **2020**, *1693*, 012148. [CrossRef]
- Sun, J.; Yang, Y.; He, X.; Wu, X. Northern Maize Leaf Blight Detection Under Complex Field Environment Based on Deep Learning. IEEE Access 2020, 8, 33679–33688. [CrossRef]
- 21. Haque, M.; Marwaha, S.; Deb, C.K.; Nigam, S.; Arora, A.; Hooda, K.S.; Soujanya, P.L.; Aggarwal, S.K.; Lall, B.; Kumar, M. Deep Learning-Based Approach for Identification of Diseases of Maize Crop. *Sci. Rep.* **2022**, *12*, 6334. [CrossRef]
- 22. Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A ConvNet for the 2020s. arXiv 2022, arXiv:2201.03545.
- 23. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. arXiv 2017, arXiv:1711.05101.
- 24. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
- Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1492–1500.
- Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 21–26 July 2017; pp. 4700–4708.
- Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
- 28. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* 2020, arXiv:2010.11929.
- 29. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. *arXiv* 2021, arXiv:2103.14030.
- Yang, L.; Zhang, R.Y.; Li, L.; Xie, X. Simam: A simple, parameter-free attention module for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021; pp. 11863–11874.
- 31. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. arXiv 2018, arXiv:1807.06521.
- 32. Liu, Y.; Shao, Z.; Teng, Y.; Hoffmann, N. NAM: Normalization-based Attention Module. arXiv 2021, arXiv:2111.12419.
- 33. Hendrycks, D.; Gimpel, K. Gaussian error linear units (gelus). arXiv 2016, arXiv:1606.08415.
- 34. Glorot, X.; Bordes, A.; Bengio, Y. Deep sparse rectifier neural net-works. J. Mach. Learn. Res. 2011, 15, 315–323.
- 35. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv* 2015, arXiv:1502.03167.
- 36. Ba, J.; Kiros, J.; Hinton, G. Layer normalization. arXiv 2016, arXiv:1607.06450.
- 37. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 18–21 June 2018; pp. 9423–9433.
- 38. Jaderberg, M.; Simonyan, K.; Zisserman, A.; Kavukcuoplu, K. Spatial transformer networks. arXiv 2015, arXiv:1506.02025.
- 39. Saleem, M.H.; Potgieter, J.; Arif, K. Plant Disease Detection and Classification by Deep Learning. *Plants* **2019**, *8*, 468. [CrossRef] [PubMed]
- Li, Q.; Cai, W.; Wang, X.; Zhou, Y.; Feng, D.D.; Chen, M. Medical image classification with convolutional neural network. In Proceedings of the 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV), Singapore, 10–12 December 2014; pp. 844–848.

**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.