

Article

Using a Resnet50 with a Kernel Attention Mechanism for Rice Disease Diagnosis

Mehdhar S. A. M. Al-Gaashani ¹, Nagwan Abdel Samee ², Rana Alnashwan ^{2,*}, Mashaal Khayyat ³
and Mohammed Saleh Ali Muthanna ⁴

- ¹ College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; mr.mehdhar@gmail.com
- ² Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia; nmabdelsamee@pnu.edu.sa
- ³ Department of Information Systems and Technology, College of Computer Science and Engineering, University of Jeddah, Jeddah 23218, Saudi Arabia; mkhayyat@uj.edu.sa
- ⁴ Institute of Computer Technologies and Information Security, Southern Federal University, 347922 Taganrog, Russia; muthanna@sfedu.ru
- * Correspondence: roalnashwan@pnu.edu.sa

Abstract: The domestication of animals and the cultivation of crops have been essential to human development throughout history, with the agricultural sector playing a pivotal role. Insufficient nutrition often leads to plant diseases, such as those affecting rice crops, resulting in yield losses of 20–40% of total production. These losses carry significant global economic consequences. Timely disease diagnosis is critical for implementing effective treatments and mitigating financial losses. However, despite technological advancements, rice disease diagnosis primarily depends on manual methods. In this study, we present a novel self-attention network (SANET) based on the ResNet50 architecture, incorporating a kernel attention mechanism for accurate AI-assisted rice disease classification. We employ attention modules to extract contextual dependencies within images, focusing on essential features for disease identification. Using a publicly available rice disease dataset comprising four classes (three disease types and healthy leaves), we conducted cross-validated classification experiments to evaluate our proposed model. The results reveal that the attention-based mechanism effectively guides the convolutional neural network (CNN) in learning valuable features, resulting in accurate image classification and reduced performance variation compared to state-of-the-art methods. Our SANET model achieved a test set accuracy of 98.71%, surpassing that of current leading models. These findings highlight the potential for widespread AI adoption in agricultural disease diagnosis and management, ultimately enhancing efficiency and effectiveness within the sector.

Keywords: rice disease classification; self-attention mechanism; agriculture imaging



Citation: Al-Gaashani, M.S.A.M.; Samee, N.A.; Alnashwan, R.; Khayyat, M.; Muthanna, M.S.A. Using a Resnet50 with a Kernel Attention Mechanism for Rice Disease Diagnosis. *Life* **2023**, *13*, 1277. <https://doi.org/10.3390/life13061277>

Academic Editors: Tao Huang and Xin Jin

Received: 25 April 2023

Revised: 20 May 2023

Accepted: 24 May 2023

Published: 29 May 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/>).

1. Introduction

The cultivation of rice, a fundamental food crop, is confronted with a multitude of obstacles presented by various diseases, which have a considerable impact on its development and productivity. According to the Food and Agriculture Organization of the United Nations (FAO), the annual global crop yield losses resulting from diseases and pests are estimated to amount to approximately USD 220 billion [1]. The prompt detection and precise diagnosis of plant diseases are of utmost importance in order to mitigate associated financial damages. Farmers frequently overlook initial indications of disease or postpone seeking medical attention owing to misinterpretation, primarily as a result of insufficient specialized expertise and resources.

Increased accessibility to smartphones and advanced digital cameras has facilitated the process of capturing visual representations of afflicted agricultural crops. Simultaneously, progressions in computer vision technologies have facilitated the processing of images and

the diagnosis of diseases in diverse research endeavors [2–4]. Pesticides are commonly employed as the principal approach for the prevention and management of diseases in the field of agriculture. Precise diagnostic results are essential in directing the utilization of pesticides, as their overuse is a significant contributor to the deterioration of the environment [5]. Therefore, it is imperative to accurately and promptly diagnose diseases.

The conventional method of diagnosing rice leaf diseases is known to be arduous and time-intensive. Consequently, there has been an increasing inclination towards utilizing computer-assisted diagnostics as a pivotal instrument for the detection and classification of rice ailments. The exceptional generalization capabilities of convolutional neural networks (CNNs), which belong to the category of deep neural networks (DNNs), have been demonstrated in various image-processing studies. For instance, a deep residual network based on attention mechanisms was proposed to identify viruses in tomato leaves [6]. Several classifiers based on deep learning have demonstrated their efficacy in identifying images of rice diseases [7,8]. Attention modules have been found to improve a model's effectiveness in capturing relevant dependencies [9–11]. The selection of suitable loss and activation functions is a critical factor in attaining favorable outcomes with deep networks, as indicated by sources [12,13]. The utilization of deep neural network (DNN) methodologies in the field of rice disease diagnosis has not been extensively explored, resulting in a dearth of models that are tailored to the classification of rice diseases, despite the swift advancements in this area.

The objective of this investigation was to perform a comprehensive examination of the disease patterns present in rice leaves by utilizing a deep learning algorithm. Our proposal involves the implementation of a linear-kernel attention-based mechanism that aims to enhance the capacity of deep neural networks to selectively attend to crucial features. In contrast to prior research, our model takes into account distant connections within feature maps, which constitutes a notable progression in the domain. In order to effectively train deep learning models, a significant quantity of rice leaf images is typically necessary to extract diagnostically relevant features. The kernel attention method that we propose aims to improve the performance of the deep neural network model's process of learning by directing the model's attention toward the extraction of features that contain more relevant information. The kernel attention mechanism is a relatively new concept that was recently presented in the work that was conducted by Li et al. [14] for the segmentation of remote-sensing images. Imagery obtained by remote sensing can be used to keep track of and locate new urban areas as they emerge as a result of ongoing urbanization. The work presented here applies the same principle of extracting the most useful feature, utilizing kernel attention for a different application, namely, the classification of rice diseases rather than their segmentation. Comprehending the learning mechanisms of neural networks and effectively visualizing this process are of utmost importance. The utilization of feature visualization aids in illuminating the viewpoint of the model regarding the visual surroundings. It also provides insights into the manner in which a pre-existing convolutional network extracts features as well as desirable characteristics, such as feature composition and class discrimination, which become more pronounced as the network layers increase. The utilization of visualization techniques has been demonstrated to facilitate the process of model debugging, ultimately resulting in improved outcomes [15]. The contributions of the present study are as follows:

- I. A neural network is developed that utilizes self-attention based on kernel attention linear complexity (SANET) for the purpose of classifying various types of rice diseases.
- II. The SANET model is designed to hierarchically aggregate contextual data using multiscale kernel attention, thereby enabling the inference of global contextual dependencies.
- III. A novel self-attention mechanism is proposed that incorporates kernel attention to reduce high computational demand with linear complexity.

This paper is organized as follows: Section 2 presents an overview of deep learning models and their application in rice disease studies. Section 3 provides an in-depth review of the proposed model and the impact of kernel attention in feature extraction and long-range dependencies. Section 4 discusses the experimental results and evaluates the deep learning model's performance. Finally, Section 5 concludes the research findings and outlines avenues for future work.

2. Related Work

In recent years, many unique deep-learning algorithms have been introduced and put into use for the purpose of identifying rice diseases. A new method for detecting diseases in rice was proposed by Yang Lu et al. [16] using deep CNN techniques. The researchers used a dataset with 500 images of both healthy and diseased rice leaves and stems. These images were captured in an experimental rice field that was infected with 10 different types of rice infection. When put through a 10-fold cross-validation method, the accuracy of the proposed CNNs-based model was confirmed to be 95.48%.

State-of-the-art architectures, such as Inceptionv3 and VGG16, were adapted by Rahman et al. [17] for the classification and detection of rice diseases. The results from their experiments proved the models' value when used with real-world data. The proposed two-stage small CNN architecture was compared to memory-efficient solutions such as MobileNet [18], NasNet Mobile [19], and SqueezeNet [20]. As large-scale architectures are incompatible with mobile devices, this change was made. After making significant reductions in the size of the model, they were still able to achieve the desired level of accuracy (93.3%).

With the assistance of several experts, Liang et al. [21] have published a dataset for the classification of rice leaf diseases. In addition to this, they suggested using a CNN as the basis for an approach to feature extraction and disease classification. The results of their experiments demonstrated that the high-level features derived from convolutional neural networks possessed superior discriminative capabilities compared to those derived from Haar-WT (wavelet transform) and local binary pattern histograms (LBPH). According to the results of their research, the hybrid CNN and SVM that they proposed, which they called support vector machines, had greater accuracy and a larger value of the AUC, area under the curve, than conventional methods such as Haar-WT + SVM and LBP + SVM.

Ramesh and Vydeki made use of several image-processing techniques in order to reduce the amount of reliance placed on farmers to ensure the safety of agricultural products [22]. They proposed an algorithm for the classification of paddy leaf diseases utilizing an improved deep neural network in conjunction with the Jaya algorithm. Their photographs of rice plant leaves included healthy plants as well as those with bacterial brown spots, blights, blast diseases, and sheath rot. These images were taken directly from the agricultural field. During the preprocessing phase, the RGB images were converted to HSV images to eliminate the background, and then, based on the hue and saturation components of the images, binary images were extracted to distinguish sick tissue from healthy tissue. In order to categorize the infected leaves, they developed a deep neural network using the Jaya optimization algorithm (DNN-JOA). They were able to classify blast-afflicted leaves, bacterial blight leaves, sheath rot leaves, brown spot leaves, and healthy leaf images with accuracies of 98.9%, 95.78%, 92%, 94%, and 90.57%, respectively. RiceTalk [23] is a project that was developed by Chen et al. that makes use of non-image Internet of Things sensors to detect rice blights. RiceTalk was based on an Internet of Things platform for soil agriculture and was able to achieve an accuracy of 89.4% on rice blast disease.

Until the advent of deep learning, machine vision systems were often implemented using statistical machine learning techniques, notably, those based on the fields of supervised and unsupervised learning. Naive Bayes (NB), discriminant analysis (DA), support vector machines (SVMs), and k-nearest neighbors (kNN) are some of the modern methods studied by Rehman et al. [24]. The study provided an in-depth examination of how these methods are currently being used in a range of agricultural domains, drawing the conclusion that

different methods should be used for different purposes while acknowledging their limitations. On top of that, Duong-Trung et al. [25] used transfer learning to categorize rice discoloration disease, which was previously thought to be a major threat to rice production. To improve the efficiency of deep models that are already familiar with low-level characteristics, transfer learning employs weights that are pre-trained on data from different domains. The results of this study used an InceptionV3 model pre-trained on ImageNet to achieve an accuracy of 88.18% in classification, similar to how Shrivastava et al. [26] used transfer learning for a convolutional neural network to categorize rice plant diseases (CNN). Using a training/testing split of 80/20, the suggested model achieved a classification accuracy of 91.368%.

Chung et al. [27] developed a non-destructive method using machine vision to distinguish between infected and healthy seedlings after three weeks of growth. Their work centered on identifying bakanae disease, a seed-borne threat to rice. Infected plants either fail to thrive or produce fruitless panicles. In order to quantify the morphological and color characteristics of the infected and control seedlings, pictures were collected with the aid of flatbed scanners. Support vector machine (SVM)-based classifiers were utilized. Additionally, it was suggested that a genetic algorithm may be employed to find the optimal combination of required and optional model parameters. Their method had an accuracy of 87.89% in distinguishing between healthy and diseased seedlings. To properly evaluate the efficacy of a deep learning model, a thorough analysis of the features it employs is required. This was the motivation behind the study's implementation of a method for visualizing the feature maps of rice disease images [16]. In a subsequent study [17], Rahman and colleagues expanded on this method by extracting and retaining information from the deep learning model's early and intermediate layers to classify various forms of rice disease.

Tai et al. [28] used two ViT models in tandem to handle images of varying resolutions, and techniques [29,30] have recently implemented the ViT (ViT-B16 with 16 attention blocks and ViT-B32 with 32 attention blocks) without altering the images in any way. Some of these studies examined the specific disease that was affecting a plant [28,30], whereas others focused on the categorization of plants rather than the diseases that were affecting them. However, as additional plant species and disease strains emerge, it becomes increasingly challenging to address the issue. However, using such a deep neural network for the categorization of plant diseases may be excessive, and simpler models may be able to perform adequately well in some situations. Despite the benefits of the ViT's performance, it is highly impractical to use transformer-based models for leaf classification. Thus, in this paper, we propose combining the self-attention power of transformers with Resnet50 [31]. We used linearly complex kernel attention to capture long-term dependencies across many resolutions, which directly reduced the model's memory footprint.

The study conducted by Wang et al. [32] was another effort that is extremely comparable to ours and also very important to our approach. They devised the attention-based depth-wise separable neural network with Bayesian optimization, which is abbreviated as ADSNN-BO, in order to detect rice diseases in a timely and accurate manner. The foundation of their model is a MobileNet pre-trained CNN that incorporates an attention mechanism. In addition, the Bayesian optimization method is utilized in order to fine-tune the hyper-parameters. It has been determined that their model is 94.65% accurate. Additionally, their technique helps improve interpretability by offering feature analysis through the utilization of an activation map and filters visualization. In order to highlight the contrasts between our research and Wang et al.'s model, we would like to emphasize that although Wang et al. employed a MobileNet pre-trained CNN model in the classification of input images, we have classified our images using Resnet50 in our system. In addition, our model is intended to hierarchically aggregate contextual input through the use of multiscale kernel attention; yet, Wang et al.'s model only added an attention-augmented layer to the MobileNet pre-trained model. In addition to this, we are including a novel self-attention mechanism in our model, which also makes use of kernel attention, with the goal of lowering the high computing demand while maintaining linear complexity.

3. Proposed Method

3.1. Dot Product Attention

The standard dot product attention architecture is depicted in Figure 1. Given the features $[x_1, x_2, x_3 \dots, x_n] \in \mathbb{R}^N \times C$, dot-product attention generates three projected matrices, i.e., query matrix Q , key matrix K , and value matrix V , using $W_q \in \mathbb{R}^N \times C$, where N indicates the size of the input and C denotes the input channels.

$$Q = XW_q \in \mathbb{R}^N \times D_k \tag{1}$$

$$K = XW_k \in \mathbb{R}^N \times D_k \tag{2}$$

$$V = XW_v \in \mathbb{R}^N \times D_k \tag{3}$$

$D(\cdot)$ Indicates the dimensions of dot product. We represent both Q and K with the same symbol because they are supposed to have the same shape.

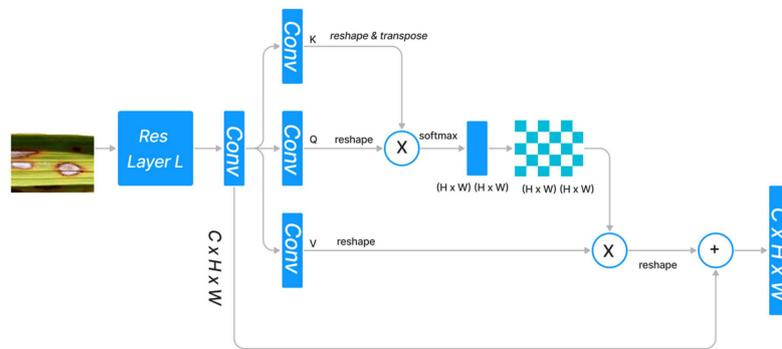


Figure 1. Standard Dot Product Attention Architecture.

In order to calculate the similarity between the i th query feature vector $(q_i^T) \in \mathbb{R}^{D_k}$ and the j th key feature vector $k_j \in \mathbb{R}^{D_k}$, we used the normalization function $p(q_i^T k_j) \in \mathbb{R}^1$. Since the query feature and the key feature are frequently generated by distinct layers, the similarities among $p(q_i^T)$ and $(p(q_j^T))$ are typically asymmetrical. The dot product attention module determines the value at the position i by performing a weighted sum across all positions, where each position's value feature is assigned a weight based on its similarity to all other positions.

$$D(Q, K, V) = p(QK^T)V \tag{4}$$

$$p(QK^T) = softmax(QK^T) \tag{5}$$

softmax indicates that the softmax operation is carried out along each column of the matrix QK^T . $p(QK^T)$ denotes the similarity between all pairs of locations. The memory complexity and computational complexity are both $O(N^2)$ because $Q \in \mathbb{R}^{D_k \times N}$ and $K^T \in \mathbb{R}^{D_k \times N}$, respectively, and because the product $Q \in \mathbb{R}^N \times N$ $K^T \in \mathbb{R}^N \times N$. Therefore, the dot-product's resource requirement severely restricts its applicability to high-dimension inputs. Altering the softmax is a strategy, while reframing the attention via the lens of the kernel is another approach. Figure 1 depicts the design of the dot-product attention mechanism, which integrates the refined features with the original input through a skip link after capturing the long-range context information from feature maps produced by CNN.

3.2. Dot Product Based on Kernel Attention

As demonstrated in Equation (2), the i th row of the result matrix obtained by dot-product attention may be written as follows:

$$D(Q, K, V)_i = \frac{\sum_{j=1}^N e^{q_i q_j k_j} v_j}{\sum_{j=1}^N e^{q_i q_j k_j}} \tag{6}$$

where i is an iteration identifier and softmax is the softmax normalization function.

From Equation (4), we can deduce that the dot-product attention mechanism works by averaging the weights assigned to the value matrix V using $e^{q_i q_j k_j}$, using the similarity measure $sim(q_i, k_j) = e^{q_i q_j k_j}$ between query matrix Q and key matrix K . Therefore, we can generalize Equation (4) by substituting a generic form for the softmax function, as follows:

$$D(Q, K, V)_i = \frac{\sum_{j=1}^N sim(q_i, k_j) v_j}{\sum_{j=1}^N sim(q_i, k_j)}, \quad sim(q_i, k_j) \geq 0 \tag{7}$$

Specifically, $sim(q_i, k_j)$ is the function that assesses the degree of similarity between q_i and k_j . Under the assumption that $sim(q_i, k_j) = e^{q_i q_j k_j}$, we have Equation (5), which is equivalent to Equation (4). Simultaneously, we can write $sim(q_i, k_j) = \varphi(q_i)^T \varnothing(k_j)$, where $\varphi(\cdot)$ and $\varnothing(\cdot)$ are kernel smoothers [33]. Therefore, the inner product space can be described as $\langle \varphi(q_j) \varnothing(k_j) \rangle$.

Equation (4) can then be further rewritten as

$$D(Q, K, V)_i = \frac{\sum_{j=1}^N \varphi(q_i)^T \varnothing(k_j) v_j}{\sum_{j=1}^N \varphi(q_i)^T \varnothing(k_j)} \tag{8}$$

which can be further simplified as

$$D(Q, K, V)_i = \frac{\varphi(q_i)^T \sum_{j=1}^N \varnothing(k_j) v_j}{\varphi(q_i)^T \sum_{j=1}^N \varnothing(k_j)} \tag{9}$$

$K \in \mathbb{R}^{D_k \times N}$ and $V^T \in \mathbb{R}^{D_k \times N}$, which significantly lowers the level of complexity with the dot-product attention method.

3.3. Kernel Attention Mechanism

We took $\varphi(q_j) = \varnothing(k_j) = swish(\cdot)$ where

$$swish(\cdot) = x * sigmoid(x) \tag{10}$$

The reason we chose Swish(.) over ReLU(.) is because when the input is somewhat negative, then the nonzero property of Swish can allow the attention mechanism to avoid zero gradients. The function can be implemented as

$$sim(q_i, k_j) = swish(q_i)^T swish(k_j) \tag{11}$$

Consequently, Equation (5) may be written as

$$D(Q, K, V)_i = \frac{swish(q_i)^T \sum_{j=1}^N swish(k_j) v_j^T}{swish(q_i)^T \sum_{j=1}^N swish(k_j)} \tag{12}$$

which can be further simplified as

$$D(Q, K, V) = \frac{\text{swish}(Q) \text{swish}(K^T) V}{\text{swish}(Q) \sum_j \text{swish} K_{i,j}^T} \tag{13}$$

Given $\sum_{j=1}^N \text{swish}(k_j) v_j^T$ and $\sum_{j=1}^N \text{swish}(k_j)$, the time and memory complexity of the proposed kernel attention technique based on (11) is merely $O(N)$, as each query can be calculated and reused.

3.4. Attention Network

With increasing input size $N = H \times W$, the computational complexity of the dot-product attention mechanism increases exponentially. We provided a self-attention module for the spatial dimension based on kernel attention to address this problem (SAKM). In the vast majority of instances, the number of input channels C was significantly lower than the number of pixels N in the feature maps along the channel dimension. Therefore, channel softmax function complexity was reasonable at $O(C^2)$ (3). The channel attention mechanism (CAM) [34] based on the dot-product was thus implemented (as shown in Figure 2). Comparable residual relationships existed between the SAKM, CAM, and dot-product attention mechanism, which was the direct sum of output and input features. SAKM and CAM were applied to produce an attention block that improved the discriminative capabilities of the generated feature maps for each layer. Both the SAKM and the CAM utilized the ResBlock’s created characteristics to hone in on the data’s location and channel, respectively. Figure 2 shows that the output of the attention block was generated by concatenating the revised feature maps. Figure 2 presents a representation of the proposed SANET architecture (a). These feature maps were generated using an ImageNet-pre-trained ResNet-50. The feature maps generated by Resnet50’s convolutional layers were further augmented using attention blocks. The characteristics from the attention blocks were combined, and the resulting output was then processed through a classification head to forecast the class.

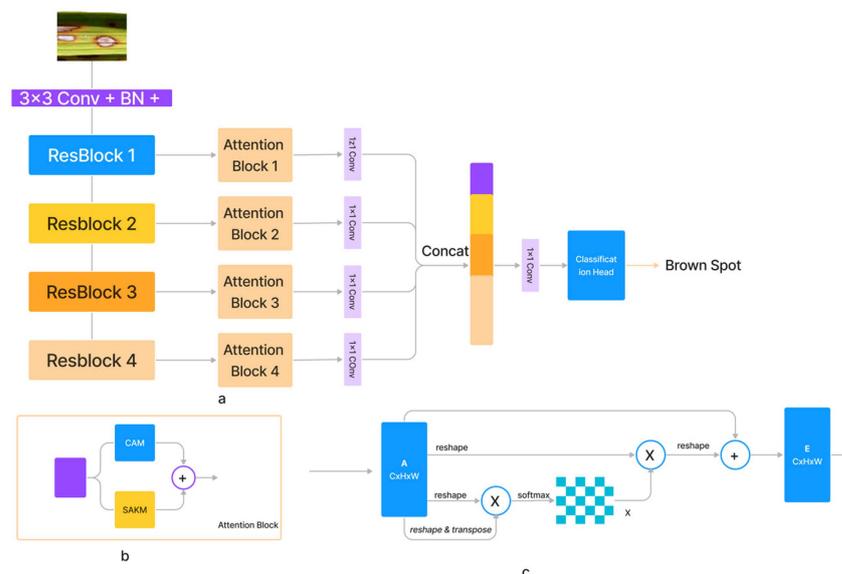


Figure 2. The architecture of (a) proposed SANET, (b) SAKM, (c) CAM Block.

4. Dataset and Experimental Settings

4.1. Dataset Description

For this study, we focused on the three most common types of rice disease: brown spot, rice hispa damage, and rice leaf blast. The foundation of the common manual diagnostic process is the visual representation of symptoms. Brown spots, also known as age spots,

are flat, dark brown lesions that are typically spherical or oval in shape and surrounded by a yellow halo. A lesion's round shape is maintained regardless of its size, and it always has a gray, necrotic center and a reddish-brown to dark-brown border. Rice hispa destroys the epidermis on the upper surface of the leaf blades. The disease eats away at the leaf tissue. Plants lose health and vitality when subjected to extreme stress. The damage caused by hispa can be easily determined if the bug is spotted on a rice leaf. Rice leaf blast can cause lesions ranging from tiny black dots to larger oval patches with a reddish rim and a gray or white center. Spots grow longer and more diamond or baroque in shape, with sharply pointed ends and gray, lifeless centers surrounded by thinner rings of reddish brown. There were some completely healthy instances of rice among the 2370 leaf samples we analyzed [35]. Table 1 illustrates the number of samples for each disease category. From each disease category, 100 samples were selected for training and testing. In Figure 3, Some samples from the rice dataset are shown.

Table 1. Rice disease dataset information by classes.

Category	Number of Samples
Rice Hispa Damage	565
Leaf Blast	779
Brown Spot	523
Healthy	501
Total	2368

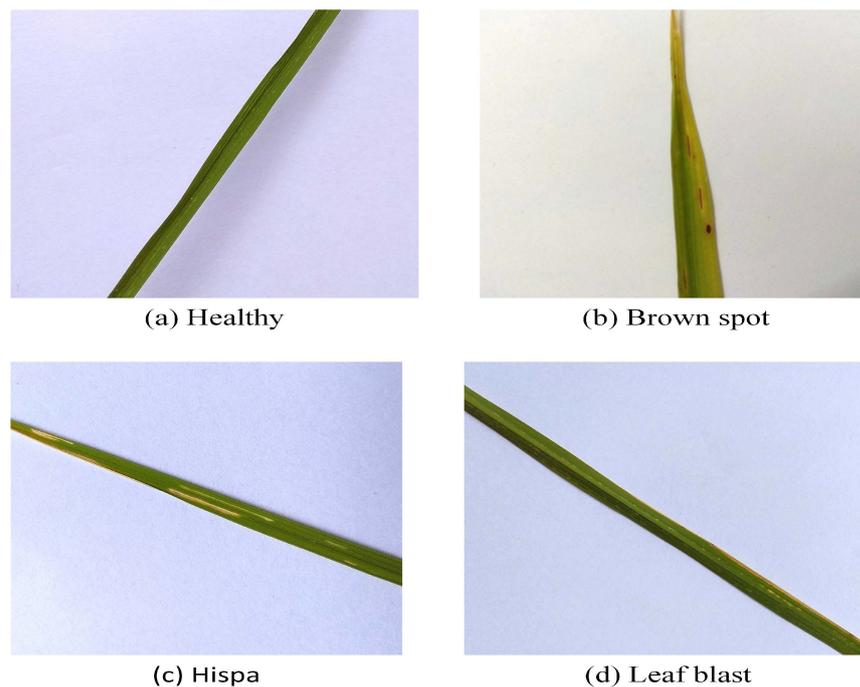


Figure 3. Some samples from the rice disease dataset. Healthy (a) and diseased leaf samples from the dataset. Three disease types in this study: (b) brown spot, (c) rice hispa damage, and (d) leaf blast.

4.2. Experimental Setting

As the backbone, we chose ResNet-50, which had already been trained on ImageNet. The Adam optimizer was configured with a learning rate of 0.0001 with 32 batch sizes. All experiments were carried out on a single NVIDIA 3090 GPU with 24 gigabytes of VRAM. For quantitative evaluation, cross-entropy loss combined with backpropagation was used to measure the gap between predicted classes and labels.

$$(p, y) = -y \log(p) - (1 - y) \log(1 - p) \quad (14)$$

p represents the prediction and y indicates the label.

4.3. Evaluation Metrics

The performance of SANET on the rice disease classification dataset was measured using the classification accuracy (14) and F1 score (15), which were computed on the cumulative confusion matrix. An F1 score is used to evaluate models in machine learning by measuring how well they perform. Essentially, it averages a model's precision and recall.

$$\text{Classification Accuracy} = \frac{\sum_{k=1}^N TP}{\sum_{k=1}^N TP + FP + TN + FN} \quad (15)$$

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (16)$$

TP_k , FP_k , TN_k , and FN_k represent the true positives, false positives, true negatives, and false negatives, respectively. Classification accuracy was calculated for all classes.

4.4. Experimental Results and Analysis

We conducted experiments using a dataset of 400 images labeled with four disease categories. The proposed model was compared against a number of other deep learning models, including VGG16 [36], ResNet50 [31], Dense-Net121 [37], MobileNetV1 [18], InceptionV3 [38], Xception [39], and ViT [33], as shown in Table 2. Each CNN model with pre-trained weights was tested on ImageNet using five-fold cross-validation and the same architecture as the original. Based on the performance metrics discussed in the evaluation criteria, Table 3 displays the assessed results. Each measurement has both its mean and standard deviation listed in the table. The five-fold cross-validation results underpinned these values. Characters in bold represent the deep learning models with the highest ratings from the evaluations (see Table 4 for details).

Table 2. Comparison with other deep learning algorithms.

Model	Precision \pm SD	Recall \pm SD	F-1 Score \pm SD	Test Accuracy \pm SD
Xception	88.2 \pm 9.68	80.4 \pm 11.06	84.2 \pm 10.76	90.38 \pm 2.74
ResNet50	62.8 \pm 25.24	76.8 \pm 14.04	64.0 \pm 12.98	79.32 \pm 6.12
MobileNet	85.6 \pm 12.95	90.2 \pm 9.28	86.8 \pm 6.22	91.36 \pm 2.82
VGG16	22.3 \pm 17.08	62.5 \pm 26.30	35.5 \pm 2.38	67.12 \pm 5.84
Inception V3	70.8 \pm 18.58	64.8 \pm 18.19	65.2 \pm 13.77	83.20 \pm 5.08
DenseNet121	84.0 \pm 12.67	76.0 \pm 16.31	78.2 \pm 7.98	88.27 \pm 4.77
ViT	97.6 \pm 0.007	97.1 \pm 0.26	97.6 \pm 0.12	97.15 \pm 2.07
SANET	99.2 \pm 0.006	98.8 \pm 0.002	98.9 \pm 0.06	98.7 \pm 0.16

Table 3. Classification performance comparison with deep learning algorithms based on three categories.

Model	Classification Category			
	Brown Spot	Healthy	Rice Hispa Damage	Leaf Blast
Xception	0.856 \pm 0.110	0.954 \pm 0.026	0.824 \pm 0.110	0.702 \pm 0.294
ResNet50	0.764 \pm 0.173	0.632 \pm 0.314	0.533 \pm 0.232	0.359 \pm 0.278
MobileNet	0.825 \pm 0.197	0.937 \pm 0.048	0.931 \pm 0.062	0.704 \pm 0.243
VGG16	0.712 \pm 0.414	0.629 \pm 0.488	0.600 \pm 0.548	0.563 \pm 0.519
InceptionV3	0.647 \pm 0.185	0.887 \pm 0.097	0.563 \pm 0.272	0.540 \pm 0.306
DenseNet121	0.684 \pm 0.282	0.770 \pm 0.275	0.743 \pm 0.167	0.784 \pm 0.190
VIT	0.980 \pm 0.009	0.977 \pm 0.005	0.961 \pm 0.017	0.967 \pm 0.021
SANET	0.995 \pm 0.004	0.981 \pm 0.002	0.985 \pm 0.0011	0.987 \pm 0.007

Table 4. Comparison of our method with state-of-the-art methods in terms of complexity.

Model	Parameters (M)	Training Time (min)/Epoch
VGG16	138	6.6
ResNet50	25.6	3.4
DenseNet121	8.062	9.2
MobileNet	13.58	3.5
Inception V3	24.8	5.55
Xception	23.82	7.4
ViT	307	5.7
SANET	28.62	3.8

The proposed SANET model had the lowest variability in its performance while still achieving the highest test accuracy, recall, precision, and F-1 score. As shown in Table 4, the SANET improved upon the ViT by 1.6% in precision, 1.2% in F-1 score, and 1.55% in accuracy while using the same 28.62 million parameters. We also evaluated the relative merits of the three most accurate models (ViT, MobileNet, and Xception) in comparison to SANET’s overall performance. Figures 4 and 5 show box plots representing the model’s accuracy and F-1 score. To classify images consistently with a lower error rate, the SANET outperformed other state-of-the-art models across the board for each type of rice disease. According to our analysis of multiple classes of rice diseases, the leaf blast disease is the most difficult to categorize. The leaf blast disease is the most difficult to classify when compared to other types of diseases discussed in this study, such as brown spot and rice hispa damage. This is because the images in each class were organized in a specific way. Brown spots, as described in Section 4, are flat, dark brown lesions that are often spherical or oval in shape and encircled by a yellow halo. The occurrence of brown spots is not constrained to specific regions; they have the propensity to manifest on any part of the plant’s anatomy. Rice hispa is responsible for the death of the epidermis that is found on the upper surface of leaf blades. The leaf tissue is consumed by the illness as it spreads. When under severe stress, plants experience a decline in their overall health and vitality. If a hispa is found on a rice leaf, the extent of the damage that it has inflicted can be ascertained with relative ease. Rice leaf blasts, on the other hand, can create lesions that range from very small black spots to larger oval patches with a ruddy ring and a gray or white center.

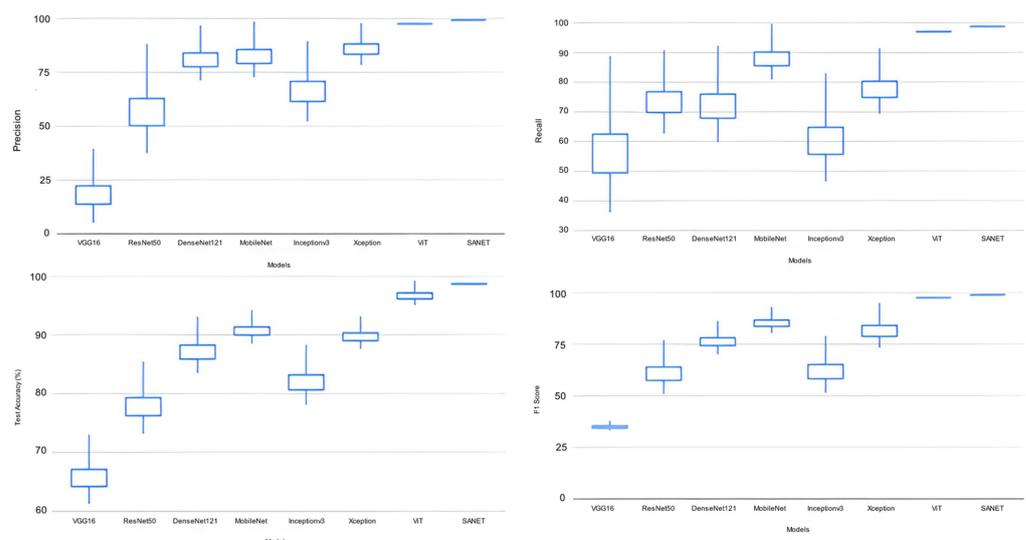


Figure 4. Classification Performance Variation.

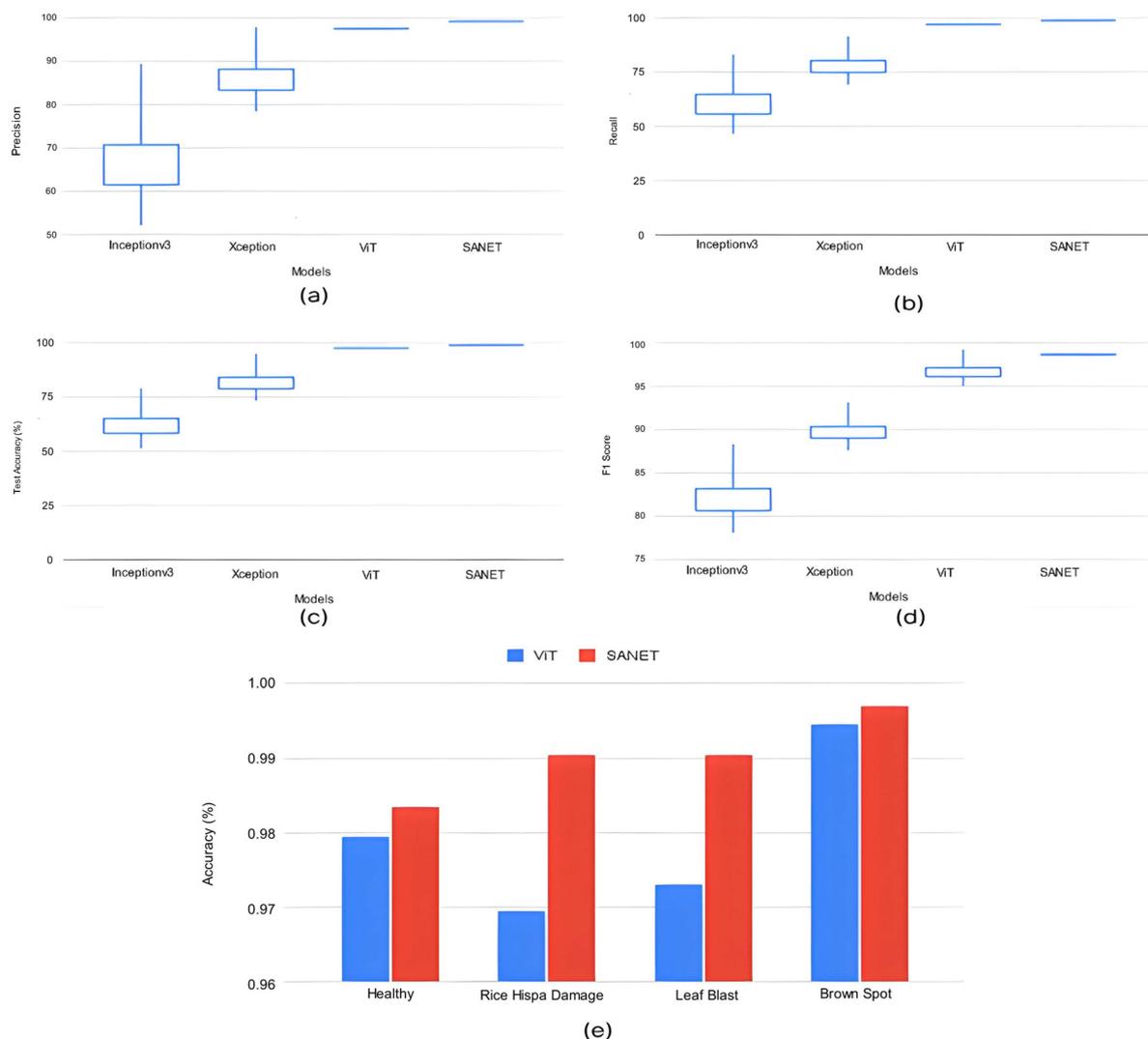


Figure 5. (a–d) Classification Performance Variation in Top Deep Learning Models. (e) Accuracy variation in top performance models for all classes.

Two methods are used to assess the quality of the characteristics extracted by the proposed model.

- (1) Activation maps
- (2) Filter visualization.

An activation map is widely regarded as a technique for visualizing data and producing a diagram that shows the activation levels. To generate the activation maps, we needed to load the weights of the best-performing network. The test sample of rice images was passed as an input to the network. Following the execution of each layer, the output indicates the kind of input that activates the layer to its fullest extent. A number of carefully chosen activation maps were constructed, as shown in Figure 6a. The figure illustrates how the model treated the various diseases that might affect rice. When attempting to diagnose brown spot disease, for instance, the network was able to successfully capture and extract the spot pattern. In addition, activation maps for each of the filters that were contained inside a single layer are shown in Figure 6b. According to the observations, it is evident that the majority of filters were activated in relation to the pattern of each disease, and it is also evident that different types of patterns were activated for different disease categories. If the generated feature maps were blank, this indicated that the network was not able to

localize the disease in the given test image. These patterns were almost certainly made up of intricate shapes that were not in the original image.

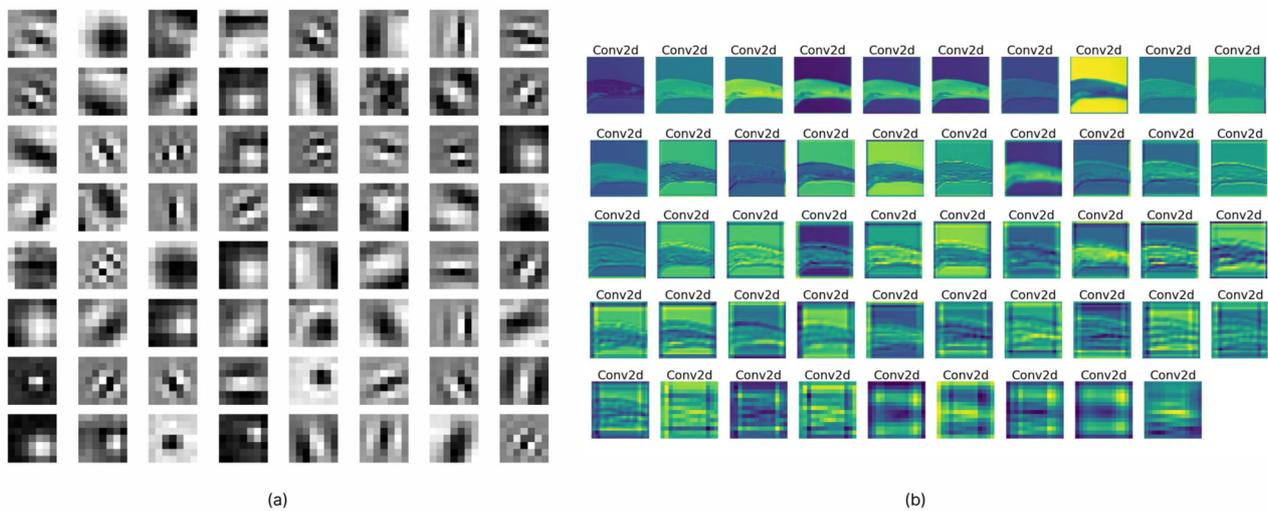


Figure 6. (a) Filter visualizations. (b) Feature maps activations.

Filter visualization of a deep learning model differs from activation maps in that it may depict filters, which are the network's weights obtained from training. Figure 7 illustrates the filters of SANET's several convolutional layers that treat rice diseases. As shown in the figure, the uppermost levels of a model's neural network often provide evidence of its ability for learning. Since the bottom half of the model tended to amass unnecessary visual patterns and information, top-layer filters were frequently intelligible and capable of evaluating a model's performance. In addition, one may claim that SANET is sensitive to the pattern of brown spot spots. The filters were more identifiable as a result of the attention process. In addition, the attention process enabled SANET to locate the illness in the provided picture, as illustrated in Figure 7. This localization was entirely unsupervised and required no annotations.

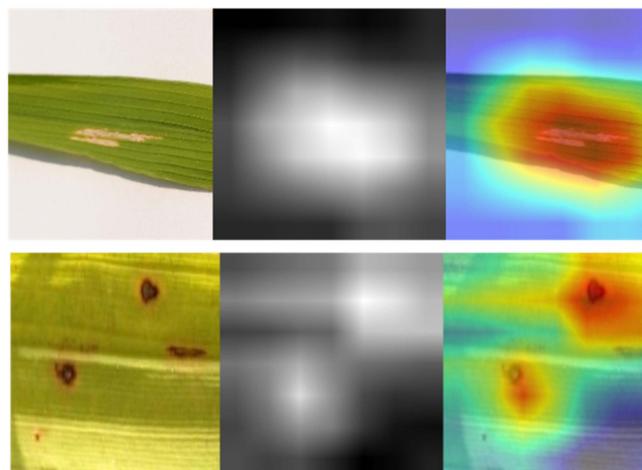


Figure 7. Class Activation Maps from SANET.

5. Conclusions

An approach to the classification of rice leaf diseases that is simple and straightforward is proposed in this article. We devised a method of aggregating contextual features at various levels of the encoder by integrating kernel attention within self-attention modules that had linear complexity. This allowed us to extract semantic information from several layers. In a variety of experiments, the performance of our proposed SANET model was

superior to that of previously developed deep learning models. In addition, a feature assessment, activation maps, and filter operations were carried out in order to highlight the performance of the model. On the rice disease dataset, our proposed model outperformed other state-of-the-art evaluated models by a significant margin, with only a small amount of variation in classification performance.

Author Contributions: Conceptualization, M.S.A.M.A.-G. and N.A.S.; methodology, M.K. and N.A.S.; software, M.S.A.M.A.-G.; validation, M.S.A.M.A.-G., N.A.S., M.K. and R.A.; formal analysis, M.S.A.M.A.-G., N.A.S., M.K. and R.A.; investigation, M.S.A.M.A., N.A.S., M.K. and R.A.; resources, N.A.S.; data curation, M.S.A.M.; writing—original draft preparation, M.S.A.M.A.-G., N.A.S. and M.K.; writing—review and editing, M.S.A.M., N.A.S., M.K. and R.A.; visualization, M.S.A.M.; supervision, N.A.S.; project administration, N.A.S. and R.A. All authors have read and agreed to the published version of the manuscript.

Funding: Princess Nourah bint Abdulrahman University Researchers Supporting Project Number (PNURSP2023R408), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are contained within the article and/or available from the corresponding author upon reasonable request.

Acknowledgments: The authors express their gratitude to Princess Nourah bint Abdulrahman University Researchers Supporting Project Number (PNURSP2023R408), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

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

References

1. Agrios, G. *Plant Pathology*, 5th ed; Elsevier Academic Press: Burlington, MA, USA, 2005; pp. 79–103.
2. Wang, H.; Won, D.; Yoon, S.W. A deep separable neural network for human tissue identification in three-dimensional optical coherence tomography images. *IJSE Trans. Health Syst. Eng.* **2019**, *9*, 250–271. [[CrossRef](#)]
3. Lu, H.; Wang, H.; Zhang, Q.; Won, D.; Yoon, S.W. A dual-tree complex wavelet transform based convolutional neural network for human thyroid medical image segmentation. In Proceedings of the 2018 IEEE International Conference on Healthcare Informatics (ICHI), New York, NY, USA, 4–7 June 2018; pp. 191–198.
4. Zhang, Q.; Wang, H.; Lu, H.; Won, D.; Yoon, S.W. Medical image synthesis with generative adversarial networks for tissue recognition. In Proceedings of the 2018 IEEE International Conference on Healthcare Informatics (ICHI), New York, NY, USA, 4–7 June 2018; pp. 199–207.
5. Wu, Y.; Yu, L.; Xiao, N.; Dai, Z.; Li, Y.; Pan, C.; Zhang, X.; Liu, G.; Li, A. Characterization and evaluation of rice blast resistance of Chinese indica hybrid rice parental lines. *Crop J.* **2017**, *5*, 509–517. [[CrossRef](#)]
6. Karthik, R.; Manikandan, H.; Anand, S.; Mathikshara, P.; Johnson, A.; Menaka, R. Attention embedded residual CNN for disease detection in tomato leaves. *Appl. Soft Comput.* **2020**, *86*, 105933.
7. Stephen, A.; Punitha, A.; Chandrasekar, A. Designing self attention-based ResNet architecture for rice leaf disease classification. *Neural Comput. Appl.* **2022**, *35*, 6737–6751. [[CrossRef](#)]
8. Reddy, S.R.; Varma, G.S.; Davuluri, R.L. Resnet-based modified red deer optimization with DLCNN classifier for plant disease identification and classification. *Comput. Electr. Eng.* **2023**, *105*, 108492. [[CrossRef](#)]
9. Yang, L.; Yu, X.; Zhang, S.; Long, H.; Zhang, H.; Xu, S.; Liao, Y. GoogLeNet based on residual network and attention mechanism identification of rice leaf diseases. *Comput. Electron. Agric.* **2023**, *204*, 107543. [[CrossRef](#)]
10. Jiang, M.; Feng, C.; Fang, X.; Huang, Q.; Zhang, C.; Shi, X. Rice Disease Identification Method Based on Attention Mechanism and Deep Dense Network. *Electronics* **2023**, *12*, 508. [[CrossRef](#)]
11. Zhang, H. Attention-based feature enhancement for rice leaf disease recognition. In Proceedings of the 2nd International Conference on Artificial Intelligence, Automation, and High-Performance Computing (AIAHPC 2022), Zhuhai, China, 25–27 February 2022; Volume 12348, pp. 29–37.
12. Goceri, E.; Gooya, A. On the importance of batch size for deep learning. In Proceedings of the International Conference on Mathematics (ICOMATH2018), An Istanbul Meeting for World Mathematicians, Istanbul, Turkey, 3–6 July 2018.
13. Zheng, Q.; Yang, M.; Yang, J.; Zhang, Q.; Zhang, X. Improvement of Generalization Ability of Deep CNN via Implicit Regularization in Two-Stage Training Process. *IEEE Access* **2018**, *6*, 15844–15869. [[CrossRef](#)]
14. Li, R.; Zheng, S.; Zhang, C.; Duan, C.; Su, J.; Wang, L.; Atkinson, P.M. Multiattention Network for Semantic Segmentation of Fine-Resolution Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5607713. [[CrossRef](#)]

15. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Cham, Switzerland, 2014; pp. 818–833.
16. Lu, Y.; Yi, S.; Zeng, N.; Liu, Y.; Zhang, Y. Identification of rice diseases using deep convolutional neural networks. *Neurocomputing* **2017**, *267*, 378–384. [[CrossRef](#)]
17. Rafeed Rahman, C.; Saha Arko, P.; Eunus Ali, M.; Khan MA, I.; Hasan Apon, S.; Nowrin, F.; Wasif, A. Identification and recognition of rice diseases and pests using convolutional neural networks. *arXiv* **2018**, arXiv:1812.01043. [[CrossRef](#)]
18. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
19. Zoph, B.; Vasudevan, V.; Shlens, J.; Le, Q.V. Learning transferable architectures for scalable image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8697–8710.
20. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. *arXiv* **2016**, arXiv:1602.07360.
21. Liang, W.-J.; Zhang, H.; Zhang, G.-F.; Cao, H.-X. Rice Blast Disease Recognition Using a Deep Convolutional Neural Network. *Sci. Rep.* **2019**, *9*, 2869. [[CrossRef](#)] [[PubMed](#)]
22. Ramesh, S.; Vydeki, D. Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm. *Inf. Process. Agric.* **2019**, *7*, 249–260. [[CrossRef](#)]
23. Chen, W.-L.; Lin, Y.-B.; Ng, F.-L.; Liu, C.-Y.; Lin, Y.-W. RiceTalk: Rice Blast Detection Using Internet of Things and Artificial Intelligence Technologies. *IEEE Internet Things J.* **2019**, *7*, 1001–1010. [[CrossRef](#)]
24. Rehman, T.U.; Mahmud, M.S.; Chang, Y.K.; Jin, J.; Shin, J. Current and future applications of statistical machine learning algorithms for agricultural machine vision systems. *Comput. Electron. Agric.* **2019**, *156*, 585–605. [[CrossRef](#)]
25. Duong-Trung, N.; Quach, L.-D.; Nguyen, M.-H.; Nguyen, C.-N. Classification of grain discoloration via transfer learning and convolutional neural networks. In Proceedings of the 3rd International Conference on Machine Learning and Soft Computing, Da Lat, Vietnam, 25–28 January 2019; pp. 27–32.
26. Shrivastava, V.K.; Pradhan, M.K.; Minz, S.; Thakur, M.P. Rice plant disease classification using transfer learning of deep convolution neural network. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *3*, 631–635. [[CrossRef](#)]
27. Chung, C.-L.; Huang, K.-J.; Chen, S.-Y.; Lai, M.-H.; Chen, Y.-C.; Kuo, Y.-F. Detecting Bakanae disease in rice seedlings by machine vision. *Comput. Electron. Agric.* **2016**, *121*, 404–411. [[CrossRef](#)]
28. Tai, H.-T.; Tran-Van, N.-Y.; Le, K.-H. Artificial cognition for early leaf disease detection using vision transformers. In Proceedings of the 2021 International Conference on Advanced Technologies for Communications (ATC), Virtual, 14–16 October 2021; pp. 33–38.
29. Reedha, R.; Dericquebourg, E.; Canals, R.; Hafane, A. Transformer neural network for weed and crop classification of high resolution UAV images. *Remote Sens.* **2022**, *14*, 592. [[CrossRef](#)]
30. Wu, S.; Sun, Y.; Huang, H. Multi-granularity feature extraction based on vision transformer for tomato leaf disease recognition. In Proceedings of the 2021 3rd International Academic Exchange Conference on Science and Technology Innovation (IAECST), Guangzhou, China, 10–12 December 2021; pp. 387–390.
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
32. Wang, Y.; Wang, H.; Peng, Z. Rice diseases detection and classification using attention based neural network and bayesian optimization. *Expert Syst. Appl.* **2021**, *178*, 114770. [[CrossRef](#)]
33. 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.
34. Tsai YH, H.; Bai, S.; Yamada, M.; Morency, L.P.; Salakhutdinov, R. Transformer Dissection: A Unified Understanding of Transformer’s Attention via the Lens of Kernel. *arXiv* **2019**, arXiv:1908.11775.
35. Do, H.M. Rice Diseases Image Dataset: An Image Dataset for Rice and Its Diseases. 2019. Available online: <https://www.kaggle.com/datasets/minhhuy2810/rice-diseases-image-dataset> (accessed on 19 May 2023).
36. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
37. 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, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
38. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
39. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.

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.