

Article

A Deep Learning-Based Crop Disease Diagnosis Method Using Multimodal Mixup Augmentation

Hyunseok Lee ¹, Young-Sang Park ¹, Songho Yang ¹, Hoyul Lee ¹, Tae-Jin Park ² and Doyeob Yeo ^{2,*}

¹ Daegu-Gyeongbuk Medical Innovation Foundation, Daegu 427724, Republic of Korea; hs.lee@kmedihub.re.kr (H.L.); yspark@kmedihub.re.kr (Y.-S.P.); sh@kmedihub.re.kr (S.Y.); hoyul@kmedihub.re.kr (H.L.)

² Korea Atomic Energy Research Institute, Daejeon 34057, Republic of Korea; etjpark@kaeri.re.kr

* Correspondence: yeody@kaeri.re.kr

Abstract: With the widespread adoption of smart farms and continuous advancements in IoT (Internet of Things) technology, acquiring diverse additional data has become increasingly convenient. Consequently, studies relevant to deep learning models that leverage multimodal data for crop disease diagnosis and associated data augmentation methods are significantly growing. We propose a comprehensive deep learning model that predicts crop type, detects disease presence, and assesses disease severity at the same time. We utilize multimodal data comprising crop images and environmental variables such as temperature, humidity, and dew points. We confirmed that the results of diagnosing crop diseases using multimodal data improved 2.58%p performance compared to using crop images only. We also propose a multimodal-based mixup augmentation method capable of utilizing both image and environmental data. In this study, multimodal data refer to data from multiple sources, and multimodal mixup is a data augmentation technique that combines multimodal data for training. This expands the conventional mixup technique that was originally applied solely to image data. Our multimodal mixup augmentation method showcases a performance improvement of 1.33%p compared to the original mixup method.

Keywords: deep learning; multimodal; mixup augmentation; crop disease; diagnosis



Citation: Lee, H.; Park, Y.-S.; Yang, S.; Lee, H.; Park, T.-J.; Yeo, D. A Deep Learning-Based Crop Disease Diagnosis Method Using Multimodal Mixup Augmentation. *Appl. Sci.* **2024**, *14*, 4322. <https://doi.org/10.3390/app14104322>

Academic Editors: Luis E Rodriguez-Saona and Didem Aykas

Received: 17 April 2024

Revised: 14 May 2024

Accepted: 17 May 2024

Published: 20 May 2024



Copyright: © 2024 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

Crop diseases pose significant challenges, as they contribute directly to reduction in both crop yield and quality. Hence, the early detection, treatment, and prevention of crop diseases play a vital role in safeguarding farmers' interests. There are several methods for diagnosing crop diseases, including diagnostic kits [1–3], genetic analysis [4,5], and the visual observation of symptoms [6]. Symptoms manifested on leaves, stems, and fruits serve as visual indicators for identifying diseased crops. However, the direct identification and diagnosis of crop diseases by human eyes face limitations due to the diverse types and varying progression of the diseases observed in different crops. Consequently, numerous automatic crop disease diagnosis technologies employing computer vision and deep learning have been developed.

Deep learning has demonstrated excellent performance compared to other image processing methods in automated disease diagnosis. Ferentinos et al. achieved a 99.53% accuracy rate in classifying 58 crop diseases using VGG networks [7], and Picon et al. employed the Resnet for classifying crop diseases collected via mobile devices [8]. In addition, many studies have been conducted to compare the performance based on model architectures or to improve the performance by modifying the model structures [9–11]. M. Arsenovic et al. [12] and J. Ahmad et al. [13] corrected low-quality crop images in real environments and used data augmentation methods that simulate real-world conditions to train deep learning models for crop disease diagnosis that perform well in actual environments.

The increase in adoption of smart farms, combined with recent advancements in IoT technology, has facilitated the collection of a broad spectrum of time-series data including temperature, humidity, and CO₂ concentration. Not all environmental data refer to time-series data, but in this study, we assume that environmental data refer to time-series data. Nalini et al. utilized the KNN (K-Nearest Neighbor) classifier and the ensemble method to predict crop diseases based on measured temperature data [14]. Mishra et al. designed a system equipped with sensors for the real-time measurement of temperature and humidity, employing a decision tree algorithm to predict the suitability of the environment for crop cultivation [15].

Previous studies have focused mostly on utilizing either image data or environmental data alone. Currently, few studies predict crop diseases using multimodal data, which comprise diverse types of information with distinct characteristics. Furthermore, there is a lack of research on data augmentation techniques to compensate for insufficient data in methods utilizing time-series data.

To address these issues, we propose a deep learning model that utilizes both crop images and environmental data, including temperature, humidity, and dew point, to predict crop diseases. The assumption that image data and environmental data for crops can play a complementary role in disease prediction is inherent in this approach. For crop leaves affected by powdery mildew, if the image quality is poor or the size of the powdery mildew spots is too small, a deep learning model might incorrectly classify the crop as healthy. However, if environmental variables such as temperature and humidity are also provided, it could help make a more accurate diagnosis. We will experimentally confirm that this assumption is correct in Section 5.

We conduct a comparative analysis of disease prediction performance for the one utilizing either image data or environmental data alone, and for utilizing multimodal data. We also introduce a multimodal mixup augmentation method, which enables the simultaneous application of crop images and environmental data and experimentally demonstrates its effectiveness in improving prediction performance.

2. Related Works

2.1. Deep Learning Approach to Crop Disease Diagnosis Using Multimodal Data

Multimodal data refer to the input of data from various sensors, including visual, auditory, and temperature sensors, encompassing diverse types of information with distinct characteristics. In deep learning, the utilization of multimodal data for learning purposes is commonly known as multimodal deep learning [16].

The analysis of multimodal data using deep learning has been studied in various fields, including healthcare [17,18], automotive [19], and facial recognition [20] domains. Rastgoo et al. introduced a deep learning-based method for the driver's stress level classification using ECG (electrocardiogram) signals, vehicle dynamics data, and environmental data [19]. It was found that when using multimodal data, the accuracy increased by up to 14.9% compared to when using unimodal data (ECG signals). J.C. Vázquez-Correa et al. analyzed data from speech, handwriting, and gait using a multimodal deep learning model to diagnose Parkinson's disease [18]. When multimodal data were used, the accuracy improved by 5.3% compared to using unimodal data (speech).

However, most of the deep learning studies for crop disease diagnosis have relied heavily on unimodality (e.g., mostly images). They have leveraged structural improvements of a CNN (convolutional neural network), utilization of multiple resolutions, and ensembles the maximization performance from a given single datum [21–23].

In this study, we would like to verify whether performance improves when environmental data as well as images are used to diagnose diseases of crops. We would like to emphasize that in this study, we do not propose a novel model structure for processing multimodal data. To date, most studies on diagnosing crop diseases using deep learning have focused on data using a single modality. However, in this study, we aim to demonstrate that a deep learning model using multimodal data outperforms a model using unimodal

data. To the best of our knowledge, this is the first attempt to perform crop disease diagnosis using multimodal data and confirm performance improvement compared to unimodal data.

2.2. Mixup Augmentation

Conventional deep learning models are trained using the ERM (Empirical Risk Minimization) method, which aims to minimize the occurrence of mispredictions. However, the ERM method often exhibits overconfidence in predicting a specific class, even for data points that are located in the vicinity of the distribution boundaries. To address the limitation of ERM, mixup—a data augmentation technique—is introduced [24]. It involves creating augmented samples by applying linear interpolation to pairs of images and labels. Specifically, two samples with different labels are randomly selected from the training dataset, and their images and labels are linearly interpolated to generate new samples. This process results in the creation of new data points that lie on the line connecting the original data points in the feature space. By doing so, mixup allows the model to leverage the distribution information present in the neighborhood of the training data. As a result, the model becomes more robust and capable of better generalizing better to unseen data, as it learns from a wider range of data points and distributions. Mixup effectively regularizes the model, preventing it from memorizing the training samples and improving its generalization ability. We note that mixup augmentation applications have primarily focused on unimodal data augmentation, such as video or audio data [25].

In this paper, we propose a multimodal mixup augmentation method that extends the unimodal mixup augmentation approach. In multimodal mixup, we can combine images with environmental data, leading to more comprehensive and robust models that can handle diverse data types and improve performance across different domains.

3. Materials and Methods

3.1. Dataset Description

We used the dataset from Korea AI Challenge, the ‘Crop Disease Diagnosis based on Agricultural Environmental Changes’. The data collection for the challenge was conducted by LG AI Research, and the challenge was hosted by Dacon Inc. (Seoul, Republic of Korea). The data can be downloaded from the Dacon website [26]. The dataset containing labels consists of 4675 samples. The provided samples consist of image data and environmental data from the same crop. The infected leaves, vegetables, or fruits with diseases are placed in the center of the image. The environmental data include the information of temperature, humidity, and dew point recorded at 10 min intervals for at least 48 h before capturing the crop disease images. The environmental data are provided in the form of comma-separated value files. One of the examples of the given crop image and environmental data is shown in Figure 1.

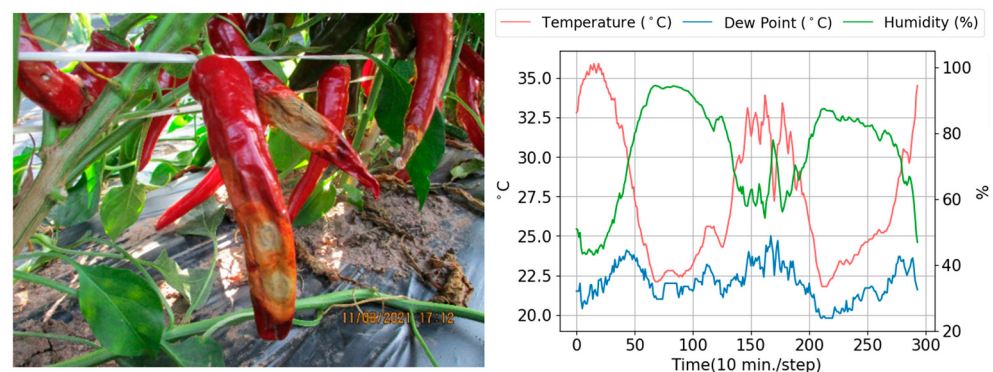


Figure 1. An example of dataset: image (left) and environmental data (right) for “Chili–Anthracnose–Intermediate Stage”.

The objective of the AI challenge is to predict the crop type, disease type, and severity of the disease using the given crop image and environmental data. In this study, we intend to define the term “Crop Disease Diagnosis” in the same way as in the AI Challenge, meaning the simultaneous classification of crop type, disease type, and disease severity. This approach allows us to use the class labels from the dataset utilized in the AI Challenge. The dataset provides class labels for 6 types of crops (including tomatoes, strawberries, and bell peppers), 9 types of diseases (such as normal, powdery mildew, and various nutrient deficiencies), and different levels of infection severity (early, intermediate, and terminal). Table 1 shows 25 classes of the dataset provided.

Table 1. Class types and the number of samples utilized for crop disease classification. N, P, K, and Ca denote deficiencies in the macronutrient elements nitrogen, phosphorus, potassium, and calcium, respectively.

Crop	Disease	Severity	The Number of Samples
Strawberry	Normal		810
Tomato	Normal		143
	Powdery Mildew	Intermediate	189
Paprika	Normal		1177
	Powdery Mildew	Early	154
		Intermediate	111
		Terminal	42
	Ca	Early	166
	N		142
	P		156
	K		153
Cucumber	Normal		917
Chili	Normal		69
	Anthracnose	Intermediate	99
	N	Early	148
	P		159
	K		157
Grape	Normal		828
	Anthracnose	Early	40
		Intermediate	12
	Powdery Mildew	Early	13
		Intermediate	29
	Sunscald	Early	18
		Intermediate	14
	Corky Core	Early	21
Total			5767

3.2. Multimodal Deep Learning

The deep learning model architecture used in this study is illustrated in Figure 2. This architecture includes an LSTM (long short-term memory) module for processing environmental data and a CNN (convolutional neural network) for image data. For the CNN model, we used a pretrained Resnet 50 [27] from the ImageNet dataset [28], which is one of the most commonly used networks for image-based deep learning research. After passing through the CNN, crop images are embedded into low-dimensional features with a dimension of 25. Environmental data include temperature, humidity, and dew point. These data are concatenated and passed through an LSTM module, which are then mapped to a

feature with a dimension of 200. Next, they are reduced to a low-dimensional feature with a dimension of 25, similar to the image data, by passing through two fc (fully connected) layers. The low-dimensional features from each modality are then concatenated to create a single dimensional feature, which is passed through a final fc layer to generate scores for 25 classes. To prevent overfitting, dropout layers were added at the output of the CNN and at the final layer of the LSTM.

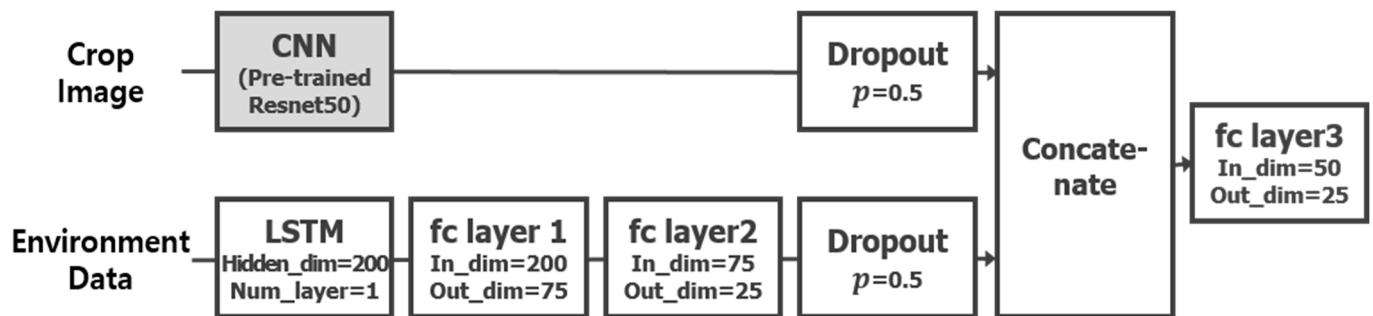


Figure 2. The deep learning model architecture for diagnosing crop diseases by analyzing crop image and environmental data.

3.3. Multimodal Mixup Augmentation

Equation (1) describes the newly generated data by unimodal mixup augmentation. x_1 and x_2 represent randomly sampled image data, and y_1 and y_2 represent the corresponding labels. λ refers to the weight used when mixing two data, and it is sampled from a beta distribution with parameters α and β . That is, $\lambda \sim \text{Beta}(\alpha, \beta)$ for a positive constant α . For convenience, we used both α and β values of 1.0. This implies that λ is sampled from a uniform distribution over the interval $[0, 1]$. Then, the augmented data $\bar{x}$ and $\bar{y}$ can be generated by linearly interpolating each raw data and label.

$$\bar{x} = \lambda x_1 + (1 - \lambda)x_2 \text{ and } \bar{y} = \lambda y_1 + (1 - \lambda)y_2 \quad (1)$$

We can also infer that the change in the value of lambda affects the intensity and impact of data augmentation. As λ approaches 0.5, the augmentation becomes stronger, while approaching 0 or 1 makes the augmentation weaker. Also, as the α value decreases, the mixup augmentation effect decreases.

To apply multimodal mixup augmentation properly, we added the augmented environmental data $\bar{z}$ as Equation (2). z_1 and z_2 represent environmental data, with the corresponding labels y_1 and y_2 .

$$\bar{z} = \lambda z_1 + (1 - \lambda)z_2 \quad (2)$$

If the data are multimodal, it is desirable to generate augmented data by linearly combining all modality data. For example, if mixup is performed only for the image modality as in the conventional method, the synthesized labels may play the role of noisy labels for the sequence model that takes environmental data as input, potentially leading to unfavorable learning outcomes.

To linearly interpolate each sampled data, it is necessary to first align their data sizes. The size of the image data sampled was equal; however, the length of the environmental data varied. In general, this can be achieved through resampling, interpolation, or other methods to match the predetermined size. In this study, a crop or padding strategy was used for the convenience of implementation. When the length of the environmental data was excessively long, we cropped them—fixing the size at 320—whereas when they were too short, we applied zero-padding to ensure it reached a length of 320.

Figure 3 shows an example of a proposed mixup augmentation for multimodal data. The top images represent crop images, while the bottom graphs represent environmental

data normalized with min-max normalization. Figure 3a represents Paprika-Deficiency (P)-Early, (b) shows Paprika-Normal, and (c) illustrates the results of mixup augmentation. The multimodal data from (a) and (b) were linearly interpolated with a weight of 0.5 each, resulting in synthesized data (c). The synthesized data have labels of $0.5 \times$ 'Paprika-Deficiency (P)-Early' and $0.5 \times$ 'Paprika-Normal'.

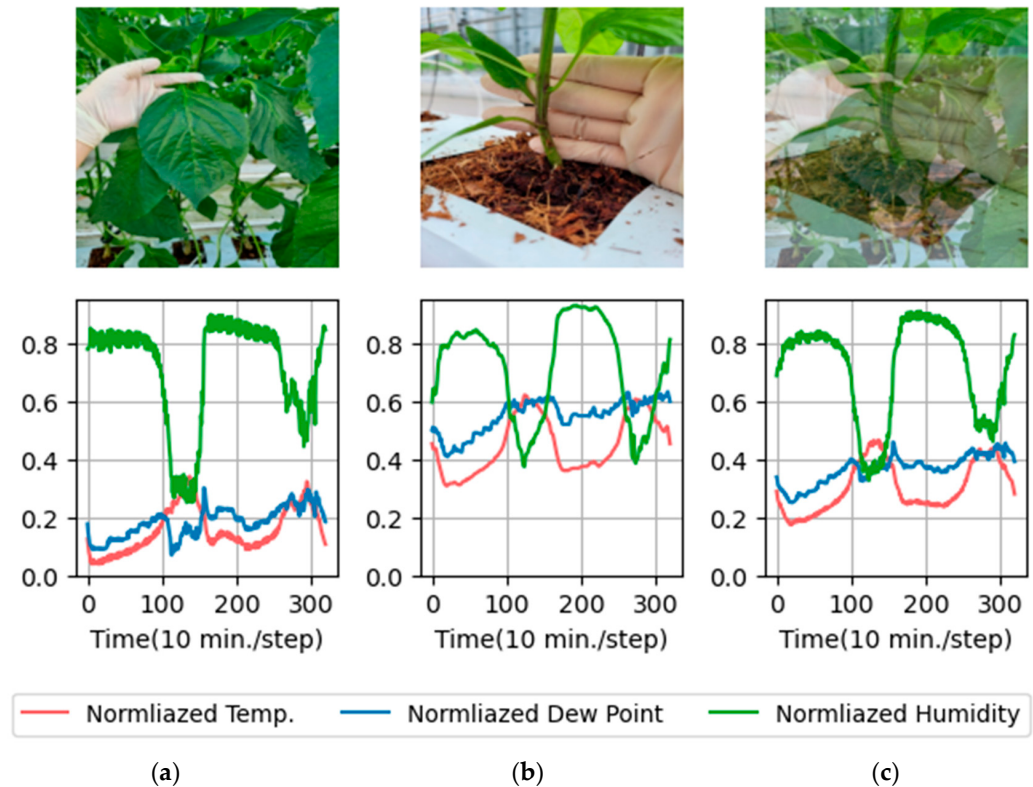


Figure 3. An example of multimodal mixup (a) 'Paprika-Deficiency (P)-Early'; (b) 'Paprika-Normal'; (c) Multimodal mixup result. Multimodal data (a,b) are linearly interpolated to generate augmented data (c). The label of augmented data (c) is also generated by linearly interpolating the labels of (a,b).

4. Experiments

We experimentally analyzed the crop disease classification performance. We compared the result from both image and environmental data with that from crop image data alone. We also compared the classification performance of the proposed multimodal mixup data augmentation method with that of unimodal mixup data augmentation on image data. We utilized the dataset from the Korean AI Challenge, which is introduced in Section 3.1. We divided the dataset into a ratio of 5:3:2 for training, validation, and testing sets, respectively. Due to the imbalanced distribution of data samples, we oversampled for the classes with insufficient sample size during the training phase. The macro $F1$ score was employed as the performance metric for crop disease diagnosis prediction. This is commonly used for multi-class classification and was also employed as a performance metric in the AI challenge. The $F1$ score can be expressed as shown in Equation (3), where TP_i , FP_i , and FN_i are, respectively, the number of true positive samples (correctly classified positive samples), false positive samples (incorrectly classified positive samples), and false negative samples (incorrectly classified negative samples) for a given i -th class, and n is the number of classes.

$$F1 = \frac{1}{n} \sum_i^n \frac{TP_i}{TP_i + (FP_i + FN_i)/2} \quad (3)$$

All experiments were conducted in an environment with an Intel(R) Core(TM) i9-10940x CPU@ 3.30 GHz, DDR 128 GB, and NVIDIA RTX3090. The implementation was done using PyTorch version 1.13.0. In all experiments, we used a minibatch size of 8. Also,

we used an Adam optimizer [29] to train the deep learning model over 300 epochs, where the initial learning rate was set to 0.001, and the learning rate scheduler was the cosine-annealed warm restart learning scheduler. Adam optimizer is one of the most commonly used in deep learning. It adaptively adjusts learning rates based on the first and second moments of the gradients, which allows for faster convergence and robustness to sparse gradients.

All image data were resized to 256×256 and normalized using the mean and standard deviation values used in ImageNet training [28]. Environmental data were normalized using min-max normalization so that the data values could fall between 0 and 1. Then, zero-padding or cropping was performed so that the length of all data could be 320. Augmentation methods including scaling, rotation, translation, flipping, and coarse dropout were only applied to image data.

5. Results

5.1. Comparative Analysis of Crop Disease Diagnosis Performance

Table 2 presents the classification performance of crop disease diagnosis for the testing dataset with respect to data modality. We conducted the same experiment three times, only varying the initial training conditions (e.g., random seed). Furthermore, we present the average *F1* score, along with its estimation under a 95% confidence interval, in Table 2.

Table 2. Crop disease diagnosis *F1* score (%) for the testing dataset by data modality.

CNN Model	Environmental Data	Image	Multimodal
Resnet 50		87.79 ± 0.70	90.37 ± 0.59
DenseNet 121	80.84 ± 0.91 (LSTM)	90.60 ± 0.46	92.02 ± 0.74
Xception		88.07 ± 0.89	90.68 ± 0.33
MobileNet V3		89.27 ± 0.73	90.56 ± 0.46
EfficientNet V2-Small		90.34 ± 0.53	92.15 ± 1.18

We expected that multimodal data would demonstrate a wider range of features compared to unimodal data, such as environmental or image data. As expected, the best classification performance was achieved when using multimodal data (e.g., average *F1* score of Resnet 50 is 90.37%). However, we also noticed that using only environmental data had a noticeable impact on the occurrence of crop diseases, as a performance of *F1* score 80.84% was observed.

Additionally, we conducted iterative experiments by changing CNN models to show that the proposed method was not limited to a particular model. And, we confirmed once again that the performance was the best when environmental data and image data are used together. The models used were DenseNet 121 [30], Xception [31], MobileNetV3 [32], and EfficientNetV2-Small [33]. To improve image classification performance, these models reference the output of all previous layers (DenseNet 121), change the layer structure to increase accuracy while reducing computational cost (Xception, MobileNetV3), or systematically optimize the model's structure (EfficientNetV2-Small). Since these models have different architectures, they can extract different features and make different judgments, even from the same images. Our experiments with these different architectures demonstrate that the proposed method is robust across various CNN models.

In Section 1, we assumed that image data and environmental data play a complementary role in predicting crop disease diagnosis. Here, we will confirm this assumption through two case studies of crop disease prediction results based on different data modalities. Figure 4 shows an image of a paprika leaf infected with intermediate-stage powdery mildew along with the associated environmental data. When only environmental data were used, the model accurately predicted the crop type and disease. However, when only images were used, the model correctly identified the crop type but incorrectly classified it as healthy. In fact, since the disease was in its early stage, the powdery mildew spores were barely visible in the images. When both image and environmental data were used, the model correctly identified both the crop type and the disease.

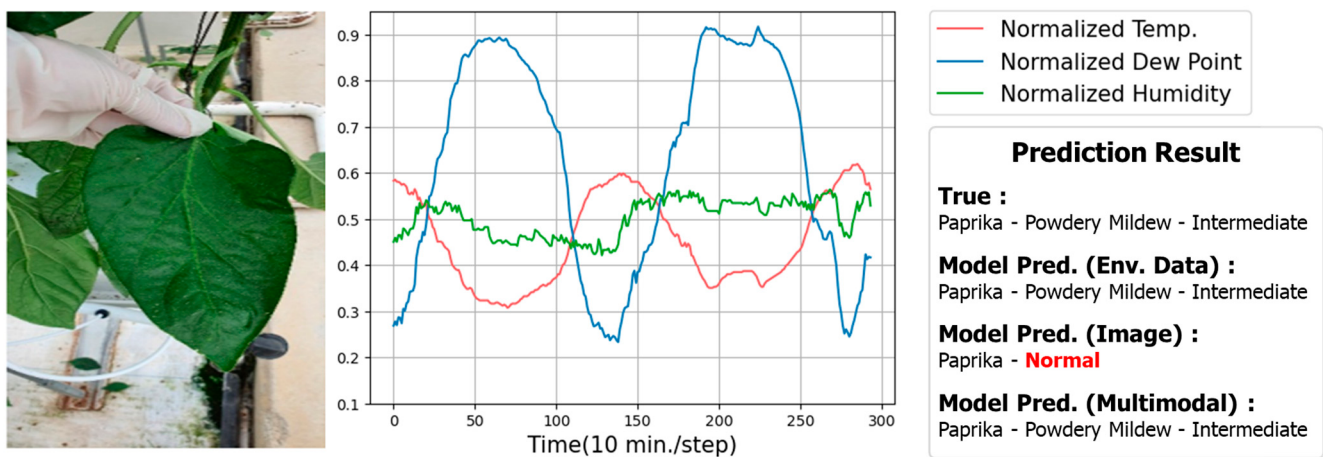


Figure 4. Crop disease prediction result by data modality—Case I: When using image data, crop type was predicted correctly, but disease type was not. However, both crop and disease type were correctly predicted when using environmental data and multi-modal data respectively.

Figure 5 shows an image of chili in the early stage of macronutrient element N (nitrogen) deficiency, along with its environmental data. When only environmental data were used, the deep learning model misclassified the crop type and disease. However, when using only image data, the model accurately identified both the crop type and disease. When using both image and environmental data, the model also accurately identified the crop type and disease.

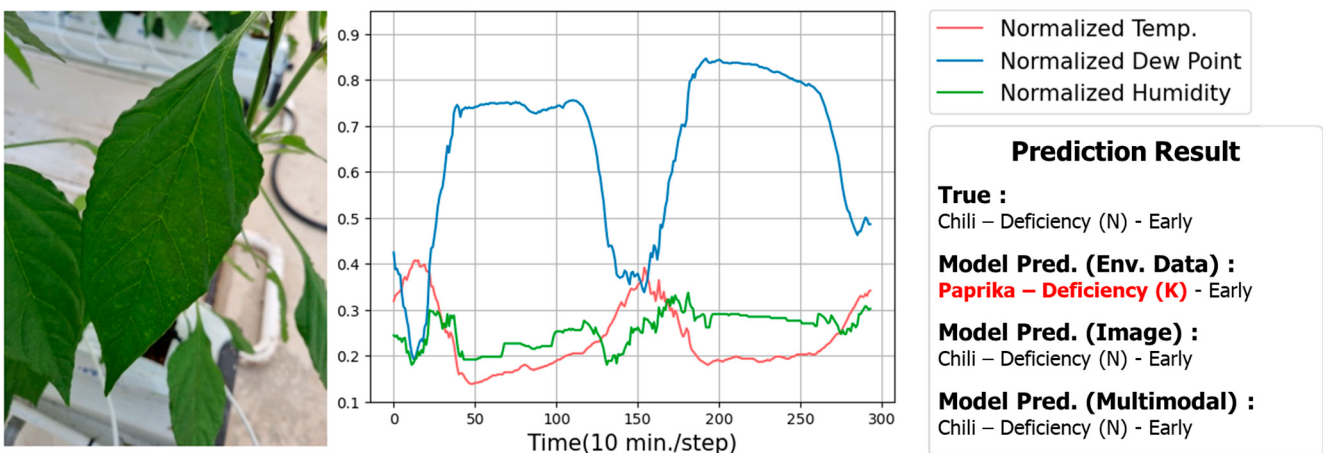


Figure 5. Crop disease prediction result by data modality—Case II: When using environmental data, both crop and disease type were misclassified. However, both crop and disease type were correctly predicted when using image data and multi-modal data, respectively.

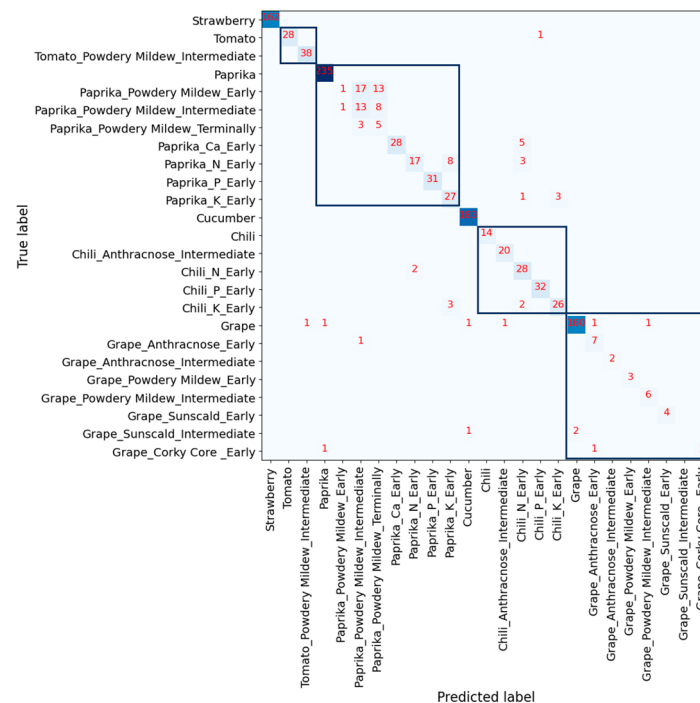
These two cases show that while a deep learning model using a single modality might misclassify crop diseases, it predicted accurately when using multimodal data. This implies that multimodal data can complement each other and enhance the accuracy of crop disease diagnosis.

Table 3 shows the result from crop disease diagnosis with respect to mixup augmentation. Using multimodal mixup augmentation resulted in the best performance, regardless of CNN model types. This implies that the proposed multimodal mixup augmentation method shows generalized performance not limited to specific model architectures.

Table 3. Crop disease diagnosis F1 score (%) for the testing dataset by data augmentation method.

CNN Model	Without Mixup	Image-Only Mixup	Multimodal Mixup
Resnet 50	90.37 $\pm$ 0.59	90.93 $\pm$ 0.34	92.26 $\pm$ 0.79
DenseNet 121	92.02 $\pm$ 0.74	92.08 $\pm$ 0.42	92.66 $\pm$ 0.39
Xception	90.68 $\pm$ 0.33	90.97 $\pm$ 0.90	91.95 $\pm$ 0.56
MobileNet V3	90.56 $\pm$ 0.46	91.17 $\pm$ 0.80	91.94 $\pm$ 0.64
EfficientNet V2-Small	92.15 $\pm$ 1.18	90.85 $\pm$ 0.46	93.22 $\pm$ 0.61

In Figures 6 and 7, the confusion matrix illustrates the predictions of a deep learning model trained with different data modality. In the confusion matrix, a square box drawn with bold blue lines means that the labels in the box belong to the same crops but have normal status or different disease severity. For example, in the box located at the top left corner, there are two labels for normal and powdery mildew with intermediate severity for tomato. The intensity of the box color represents the number of data. The darker blue, the more data it represents, and the lighter one, the fewer data it represents.

**Figure 6.** Confusion matrix by data modality: environmental data (F1 score 80.9%).

The F1 score for crop disease classification performance is 89.5% when only image data are used, which is 8.6%p higher than when only environmental data are used. On the other hand, when the type of crop is correctly predicted, there are quite a few cases where the severity of disease is also correctly predicted. From the facts above, we conclude that environmental data alone do not have a significant impact on crop type classification, but they provide sufficient information to classify crop diseases. When only image data are used, most of the types of crops are well predicted.

However, there are quite a few cases where, even though the type of crop is correctly predicted, the type and severity of the disease were incorrectly predicted. Consequently, environmental data and image data seem to be able to play complementary roles in deep learning model learning. In fact, using multimodal data, as shown in Table 2, resulted in more than 2%p performance improvement compared to using unimodal data.

Figures 8 and 9 show the confusion matrix for two mixup augmentation methods, that is, image-only mixup and multimodal mixup. Incorporating both environmental and image data notably enhances performance in both cases compared to using data from a single modality. In particular, when the mixup augmentation method is applied to both image and

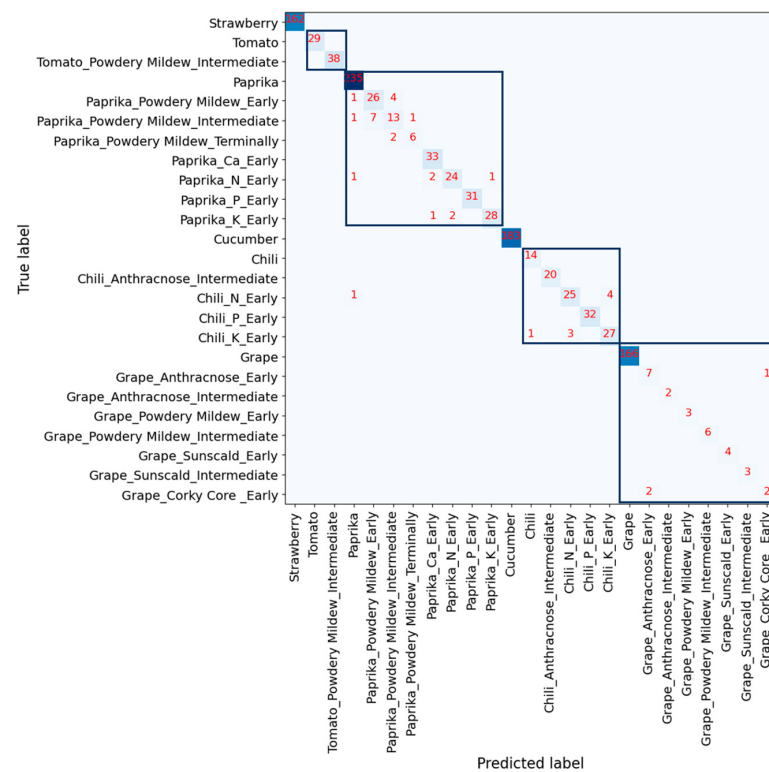


Figure 9. Confusion matrix by augmentation method: multimodal mixup (F1 score 92.4%).

5.2. Computational Cost Comparison

We also investigated the impact of including environmental data on the model's complexity and inference speed. For the environmental data, we utilized the simplest forms of the LSTM layer and two fully connected layers, resulting in a model with low complexity, as shown in Table 4. However, this model exhibited the lowest F1-score among the three modalities. Interestingly, when combining the model for environmental data with Resnet 50 for image data, we observed a 2.5%p increase in F1-score compared to Resnet 50 alone, despite the slight increase in model complexity and computational volume being negligible in comparison.

Table 4. Comparison of model complexity and inference time.

Modality of Data	The Number of Params. (M)	FLOPs (G)	F1 Score (%)
Environmental Data	0.19	0.11	80.84 ± 0.91
Image	23.56	10.76	87.79 ± 0.70
Multimodal	23.75	10.88	90.37 ± 0.59

6. Discussion

In this study, we proposed a multimodal-based crop disease diagnosis method that uses both image and environmental data. We also extended the mixup augmentation, commonly used in the image domain, to be applicable to multimodal data.

Previous studies on crop disease diagnosis have focused primarily on a single modality, such as images or environmental data. To improve accuracy, they have worked on refining model architectures [34–36], collecting new data to simulate real-world conditions [12,13], or compressing models for real-time detection [37–39]. Recently, with the advancement of IoT technologies like low-power sensors and remote-sensing, it has become easier to acquire various multimodal data. Given these changes, we explored a crop disease diagnosis approach using multimodal data, rather than focusing on enhancing performance with a single modality data. The underlying assumption of this approach is that multimodal data,

combining image and environmental data, can be complementary. Experimental results showed that the crop disease diagnosis performance was as follows: environmental data (80.84%) < image data (87.79%) < multimodal data (90.37%). Additionally, we expanded the mixup augmentation method, widely used in deep learning for images, for multimodal deep learning, and observed a 1.33%p performance improvement compared to the original mixup augmentation [24].

This study has several limitations. The first is having performed validation with different datasets, but it is challenging to obtain paired image and environmental data from crops. Our team is searching for new datasets to validate the proposed study and to explore new follow-up research. Another limitation is the difficulty of applying the method in real-world environments. The data used in this study were carefully curated, but real-world images can have different shooting angles and varying lighting conditions. Finding the best combination of data augmentation methods might be important. Environmental data collected from sensors can be noisy. Sensor errors might require signal processing, and missing data may need imputation. The model for environmental data used a simple LSTM structure with two fully connected layers, but more complex models might be required in real-world scenarios.

7. Conclusions

A multimodal deep learning model architecture has been presented to enhance the prediction performance of crop disease diagnosis models. Applying multimodal data, using both the image and environmental data simultaneously in experiments, shows better performance (by 2.58%p in the test dataset) against applying unimodal data on an image. This enhancement comes with a negligible increase in model complexity. The comparison between the result from the multimodal mixup augmentation approach and that from the unimodal one indicates that the multimodal mixup augmentation approach has a 1.33%p performance improvement in the test dataset *F1* score for the Resnet 50 CNN model. Our results demonstrate that the best performance on the test dataset was achieved when the multimodal mixup augmentation was applied to both images and environmental data. This finding underscores the fact that utilizing multimodal data, which includes crop images and environmental data, along with appropriately extended mixup augmentation not only enhances disease classification performance but also improves the model's generalization abilities.

Author Contributions: Conceptualization, H.L. (Hyunseok Lee) and S.Y.; software, H.L. (Hyunseok Lee) and D.Y.; data curation, S.Y. and Y.-S.P.; visualization, Y.-S.P. and H.L. (Hoyul Lee); writing—original draft preparation, H.L. (Hyunseok Lee); writing—review and editing, T.-J.P. and D.Y.; funding acquisition T.-J.P. and D.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant, funded by the Korean government (Ministry of Science and ICT, MIST) (No. RS-2022-00144000).

Data Availability Statement: Restrictions apply to the availability of these data. The data were obtained from ‘Crop Disease Diagnosis based on Agricultural Environmental Changes’ hosted by LG AI Research and are available at <https://dacon.io/competitions/official/235870/overview/description> (accessed on 16 May 2024) with permission. Our code is available at https://github.com/hyunseoki/crop_disease (accessed on 7 March 2024).

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

1. Gascoyne, P.R.; Vykoukal, J.V.; Schwartz, J.A.; Anderson, T.J.; Vykoukal, D.M.; Current, K.W.; McConaghy, C.; Becker, F.F.; Andrews, C. Dielectrophoresis-based programmable fluidic processors. *Lab Chip* **2004**, *4*, 299–309. [\[CrossRef\]](#)
2. Lin, C.H.; Tsai, C.H.; Pan, C.W.; Fu, L.M. Rapid circular microfluidic mixer utilizing unbalanced driving force. *Biomed. Microdevices* **2007**, *9*, 43–50. [\[CrossRef\]](#)
3. Taylor, M.T.; Belgrader, P.; Furman, B.J.; Pourahmadi, F.; Kovacs, G.T.; Northrup, M.A. Lysing bacterial spores by sonication through a flexible interface in a microfluidic system. *Anal. Chem.* **2001**, *73*, 492–496. [\[CrossRef\]](#)

4. López, M.M.; Llop, P.; Olmos, A.; Marco-Noales, E.; Cambra, M.; Bertolini, E. Are molecular tools solving the challenges posed by detection of plant pathogenic bacteria and viruses? *Curr. Issues Mol. Biol.* **2009**, *11*, 13–46.
5. Mumford, R.; Boonham, N.; Tomlinson, J.; Barker, I. Advances in molecular phytodiagnostics—New solutions for old problems. *Eur. J. Plant Pathol.* **2006**, *116*, 1–19. [\[CrossRef\]](#)
6. Nandhini, M.; Kala, K.U.; Thangadarshini, M.; Verma, S.M. Deep Learning model of sequential image classifier for crop disease detection in plantain tree cultivation. *Comput. Electron. Agric.* **2022**, *197*, 106915. [\[CrossRef\]](#)
7. Ferentinos, K.P. Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* **2018**, *145*, 311–318. [\[CrossRef\]](#)
8. Picon, A.; Alvarez-Gila, A.; Seitz, M.; Ortiz-Barredo, A.; Echazarra, J.; Johannes, A. Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. *Comput. Electron. Agric.* **2019**, *161*, 280–290. [\[CrossRef\]](#)
9. Atila, Ü.; Uçar, M.; Akyol, K.; Uçar, E. Plant leaf disease classification using efficientnet deep learning model. *Ecol. Inform.* **2021**, *61*, 101182. [\[CrossRef\]](#)
10. Yoon, H.S.; Jeong, S.B. Performance comparison of base CNN models in transfer learning for crop diseases classification. *J. Soc. Korea Ind. Syst. Eng.* **2021**, *44*, 33–38. [\[CrossRef\]](#)
11. Pandian, J.A.; Kumar, V.D.; Geman, O.; Hnatiuc, M.; Arif, M.; Kanchanadevi, K. Plant disease detection using deep convolutional neural network. *Appl. Sci.* **2022**, *12*, 6982. [\[CrossRef\]](#)
12. Arsenovic, M.; Karanovic, M.; Sladojevic, S.; Anderla, A.; Stefanovic, D. Solving current limitations of deep learning based approaches for plant disease detection. *Symmetry* **2019**, *11*, 939. [\[CrossRef\]](#)
13. Ahmad, J.; Jan, B.; Farman, H.; Ahmad, W.; Ullah, A. Disease detection in plum using convolutional neural network under true field conditions. *Sensors* **2020**, *20*, 5569. [\[CrossRef\]](#)
14. Nalini, T.; Rama, A. Impact of temperature condition in crop disease analyzing using machine learning algorithm. *Meas. Sens.* **2022**, *24*, 100408. [\[CrossRef\]](#)
15. Mishra, D.; Deepa, D. Automation and integration of growth monitoring in plants (with disease prediction) and crop prediction. *Mater. Today Proc.* **2021**, *43*, 3922–3927.
16. Ngiam, J.; Khosla, A.; Kim, M.; Nam, J.; Lee, H.; Ng, A.Y. Multimodal deep learning. In Proceedings of the 28th International Conference on Machine Learning (ICML-11), Bellevue, WA, USA, 28 June–2 July 2011; pp. 689–696.
17. Xu, T.; Zhang, H.; Huang, X.; Zhang, S.; Metaxas, D.N. Multimodal deep learning for cervical dysplasia diagnosis. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, 17–21 October 2016; Proceedings, Part II 19; Springer: Berlin/Heidelberg, Germany, 2016; pp. 115–123.
18. Vásquez-Correa, J.C.; Arias-Vergara, T.; Orozco-Arroyave, J.R.; Eskofier, B.; Klucken, J.; Nöth, E. Multimodal assessment of Parkinson’s disease: A deep learning approach. *IEEE J. Biomed. Health Inform.* **2018**, *23*, 1618–1630. [\[CrossRef\]](#)
19. Rastgoo, M.N.; Nakisa, B.; Maire, F.; Rakotonirainy, A.; Chandran, V. Automatic driver stress level classification using multimodal deep learning. *Expert Syst. Appl.* **2019**, *138*, 112793. [\[CrossRef\]](#)
20. Tzirakis, P.; Trigeorgis, G.; Nicolaou, M.A.; Schuller, B.W.; Zafeiriou, S. End-to-end multimodal emotion recognition using deep neural networks. *IEEE J. Sel. Top. Signal Process.* **2017**, *11*, 1301–1309. [\[CrossRef\]](#)
21. Chen, Z.; Wu, R.; Lin, Y.; Li, C.; Chen, S.; Yuan, Z.; Chen, S.; Zou, X. Plant disease recognition model based on improved YOLOv5. *Agronomy* **2022**, *12*, 365. [\[CrossRef\]](#)
22. Li, J.; Qiao, Y.; Liu, S.; Zhang, J.; Yang, Z.; Wang, M. An improved YOLOv5-based vegetable disease detection method. *Comput. Electron. Agric.* **2022**, *202*, 107345. [\[CrossRef\]](#)
23. Yu, H.; Liu, J.; Chen, C.; Heidari, A.A.; Zhang, Q.; Chen, H.; Mafarja, M.; Turabieh, H. Corn leaf diseases diagnosis based on K-means clustering and deep learning. *IEEE Access* **2021**, *9*, 143824–143835. [\[CrossRef\]](#)
24. Zhang, H.; Cisse, M.; Dauphin, Y.N.; Lopez-Paz, D. mixup: Beyond empirical risk minimization. *arXiv* **2017**, arXiv:1710.09412.
25. Xu, K.; Feng, D.; Mi, H.; Zhu, B.; Wang, D.; Zhang, L.; Cai, H.; Liu, S. Mixup-based acoustic scene classification using multichannel convolutional neural network. In Proceedings of the Advances in Multimedia Information Processing–PCM 2018: 19th Pacific-Rim Conference on Multimedia, Hefei, China, 21–22 September 2018; Proceedings Part III 19; Springer: Berlin/Heidelberg, Germany, 2018; pp. 14–23.
26. DACON AI Challenge Website. Available online: <https://dacon.io/competitions/official/235870/data> (accessed on 7 March 2024).
27. 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.
28. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
29. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the International Conference on Learning Representations (ICLR), Banff, AB, Canada, 14–16 April 2014.
30. 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.
31. 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.

32. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.
33. Tan, M.; Le, Q. Efficientnetv2: Smaller models and faster training. In Proceedings of the 2021 International Conference on Machine Learning (PMLR), Virtual, 18–24 July 2021; pp. 10096–10106.
34. Divyanth, L.G.; Ahmad, A.; Saraswat, D. A two-stage deep-learning based segmentation model for crop disease quantification based on corn field imagery. *Smart Agric. Technol.* **2023**, *3*, 100108. [[CrossRef](#)]
35. Raj, N.; Perumal, S.; Singla, S.; Sharma, G.K.; Qamar, S.; Chakkaravarthy, A.P. Computer aided agriculture development for crop disease detection by segmentation and classification using deep learning architectures. *Comput. Electr. Eng.* **2022**, *103*, 108357. [[CrossRef](#)]
36. Alqahtani, Y.; Nawaz, M.; Nazir, T.; Javed, A.; Jeribi, F.; Tahir, A. An improved deep learning approach for localization and recognition of plant leaf diseases. *Expert Syst. Appl.* **2023**, *230*, 120717. [[CrossRef](#)]
37. Mishra, S.; Sachan, R.; Rajpal, D. Deep convolutional neural network based detection system for real-time corn plant disease recognition. *Procedia Comput. Sci.* **2020**, *167*, 2003–2010. [[CrossRef](#)]
38. Garg, G.; Gupta, S.; Mishra, P.; Vidyarthi, A.; Singh, A.; Ali, A. CROPCARE: An intelligent real-time sustainable IoT system for crop disease detection using mobile vision. *IEEE Internet Things J.* **2021**, *10*, 2840–2851. [[CrossRef](#)]
39. Schaad, N.W.; Frederick, R.D. Real-time PCR and its application for rapid plant disease diagnostics. *Can. J. Plant Pathol.* **2002**, *24*, 250–258. [[CrossRef](#)]

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