E-Palli International Conferences (EIC)
23 & 24 December 2022
Kuala Lumpur, Malaysia

International Conference of Agricultural Science, Engineering (ICASE)

www.conferences.e-palli.com

DECISION SUPPORT SYSTEM FOR AGRICULTURE: CROP DISEASE RECOGNITION AND CLASSIFICATION THROUGH AN OPTIMIZE CONVOLUTION NEURAL NETWORK (CNN)

Md. Taslim[1], Md Shafiuzzaman[1], Mostafijur Rahman Akhond[1*], and Md. Alam Hossain[1]

ABSTRACT

Crop leaf diseases cause great damage to agriculture, causing significant crop losses in Bangladesh every year. Crop economic loss can be significantly reduced by accurately recognizing and classifying crop leaf diseases. This study developed an optimized Convolution Neural Network (CNN) model to recognize and classify crop leaf diseases. The proposed dataset of this study was collected from the field with the help of Bangladesh Agriculture University (BAU) and Bangladesh Agricultural Research Institute (BARI) experts. This dataset includes 5 types of crops (bean, cauliflower, paddy, potato, and tomato), 21 types of diseases, and 14624 sample images. The Adam optimizer is used as an optimizer in this study. Our developed CNN model can recognize and classify crop species and crop leaf diseases with the best accuracy of 99.67% and 96.55%, respectively. Furthermore, the proposed model is more accurate than the previous study.

Keywords: Crop Leap Disease, CNN, Crop Species, Disease Classification, Species Classification.

INTRODUCTION

Despite being the eighth most populous country in the world, Bangladesh has been plagued by a scarcity of arable land resources. According to the Ministry of Agriculture survey, less than 5% of Bangladesh's total land area is cultivated. Despite this, Bangladesh has the fastest growth rate in fruit production among the world's fruit-producing countries. It is the tenth largest producer of tropical crops, according to the United Nations Food and Agriculture Organization (FAO). According to FAO estimates (Osborne BG, 2006), the country's population has increased by 11.5 percent on average over the last 18 years. Despite this success in crop cultivation, huge amounts of crops are wasted due to a lack of timely disease identification in developing countries like Bangladesh, which is harmful to the country's economy. Natural disasters that affect a country's crop production have a negative impact on agricultural production and development. So, how to develop agriculture sustainably, particularly in a complex environment, is critical for Bangladesh. On the other hand, agricultural production is improving as science and technology advance. However, crop yield has not improved significantly due to a variety of natural and non-natural factors. Crop leaf diseases and insect problems account for the majority of the various causes. According to statistics, in Bangladesh, 20% of crops are lost due to pests and disease before they reach stable (April in Bangladesh, 15.10.22.). This problem has grown in recent years and is seriously jeopardizing the development of the plantation industry. Crop disease diagnosis and prevention have become increasingly important. Today, (Yong et al., 2020) agricultural workers often use books and networks, as well as local experts, to protect against and manage crop diseases. However, for a variety of reasons, misjudgments and other issues often arise, seriously affecting agricultural production. Crop disease research is currently divided into two areas. The first one is the traditional physical method, which detects various diseases primarily through spectral detection. Different diseases and pests cause different types of leaf damage, resulting in different spectral absorption and reflectance of diseased and healthy crops. Another is image recognition using machine learning techniques. That is, disease images are extracted using computer technology, and diseased and healthy trees are identified based on various characteristics. Some of the researchers conducted several machine learning techniques, such as (Alam et al., 2022 and Militante et al., 2019) proposed an optimized model for detecting various crop diseases. This study used 35,000 disease images, and their model achieved 96.5% accuracy, as well

[1] *Department of Computer Science and Engineering, Jashore University of Science and Technology, Jashore, Bangladesh*
[*] *Corresponding author:* mr.akhond@just.edu.bd

as 100% accuracy in detecting crop species. This paper (Hu WJ et al.,2020) addresses the issue of crop fine-grained disease. They built an IoT system using IoT technology and deep learning to identify this disease. They proposed a multidimensional feature compensation residual neural network (MDFC-ResNet) as a model, and their model identifies coarse-grained disease, species, and fine-grained disease. Their proposed model performed well in recognizing and detecting crop fine-grained disease. The authors of this study (Zhou et al.,2021) proposed a restructured residual dense network to detect tomato leaf disease. This hybrid deep learning model combines dense and deep residual networks to improve training performance and solve the vanishing gradient problem. Initially, residual deep networks are used to increase image resolution and then use the combined model to classify the disease. Finally, in the testing phase, their model achieves a top-1 average accuracy of 95%. In this study, we developed an optimized deep-learning solution for detecting crop species and diseases. Our model efficiently detects crop species and diseases within less training time. We build an efficient model to reduce model complexity and overfitting issues. In addition, we found the misclassification reason with misclassified images from our model. In this research, we conducted five species and twenty-one diseases with 14,624 sample images.

The rest of the paper is organized as follows: Section 2 Related works. Section 3 Structure of our model. Section 4 Methodology, Section 5 Result and discussion, and Section 6 Conclusions.

LITERATURE REVIEW

Crop leaf disease is a concerning issue all over the world. Researchers all over the world are working to solve this problem in order to reduce crop losses caused by leaf diseases. In this section, we review some related work on leaf disease recognition and classification.

This research (Roy et al., 2021) proposed a novel model based on state-of-the-art computer vision techniques to classify apple plant diseases. Their model addresses the existing model classification problem of apple disease. In addition, their model increases both the detection speed and the accuracy of classifying the disease. As a result, the model finds the F1-score and mean average precision (mAP) at 91.2% and 95.9%, respectively, and at a detection rate of 56.9 FPS. Moreover, their model increases the precision and the F1-score by 9.05% and 7.6%, respectively.

In addition, the authors (Cap et al., 2018) of this paper proposed a leaf localization method using on-site wide-angle images and a deep learning approach to detect plant diseases in their early stages. Their proposed model received the highest F1 score of 78% at 2 fps.

Additionally, this (Türkoğlu et al., 2019) study conducts numerous studies to assess the effectiveness of the model utilizing diverse methodologies. They identified the crop disease using nine widely used deep learning architectures. They also used transfer learning and deep feature extraction techniques, followed by classification using extreme learning machine (ELM), K-nearest neighbor (KNN), and support vector machine (SVM). For model evaluation accuracy, F1-score, specificity, and sensitivity are considered for F1-score. Finally, SVM/ELM performs better than transfer learning.

In this paper (Militante et al., 2019), they propose an optimized model for detecting various crop diseases. This study used 35,000 disease images, and their model achieved 96.5% accuracy, as well as 100% accuracy in detecting crop species.

Furthermore, this (Hu WJ et al., 2020) paper addresses the issue of crop fine-grained disease. They built an IoT system using IoT technology and deep learning to identify this disease. They proposed a multidimensional feature compensation residual neural network (MDFC-ResNet) as a model, and their model identifies coarse-grained disease, species, and fine-grained disease. Their proposed model performed well in recognizing and detecting crop fine-grained disease.

The authors of this study (Zhou et al., 2021) proposed a restructured residual dense network to detect tomato leaf disease. This hybrid deep learning model combines dense and deep residual networks to improve training performance and solve the vanishing gradient problem. Initially, residual deep networks are used to increase image resolution and then use the combined model to classify the disease. Finally, in the testing phase, their model achieves a top-1 average accuracy of 95%.

Also, the author of this (Ahmed et al., 2021) paper proposed a dual-phase CNN strategy in which small crops, including diversity, could be used to effectively analyze disease data. First, a faster RCNN method is used to remove a significant portion of the image (rice husk) and generate a secondary dataset of rice husks with no distinguishing background. CNN Architecture divides diseases into two categories: developing infectious diseases and general models. The proposed method, when combined with CNN's dual phase method, achieves an accuracy of 88.92% when applied directly to small grain datasets, which is five times better with mutual efficiency.

METHODOLOGY

In this research, our main goal is to develop an efficient CNN model to recognize and classify crop diseases and crop species using crop leaf images. For that crop, leaf images are used as the input of our model. Then the convolution operation extracts the feature and it is classified by the fully connected layer. Our entire experiment is depicted in Fig.1.

Data Description

The proposed dataset is collected from the field with the help of Bangladesh Agriculture University (BAU) and Bangladesh Agricultural Research Institute (BARI), Gazipur, Bangladesh. There are 14624 sample images of five types of crops with 21 disease classes. Five crops are Bean, Cauliflower, Paddy, Potato, and Tomato. In addition, twenty-one disease classes are Bean Angular Leaf Spot, Bean Rust, Cauliflower Alternaria Leaf Spot, Cauliflower Cabbage Aphid Colony, Cauliflower Ring Spot, Paddy Bacterial Leaf Blight, Paddy Brown Spot, Paddy Leaf Smut, Potato Early Blight, Potato Late Blight, Potato Healthy, Tomato Bacterial Spot, Tomato Early Blight, Tomato Late Blight, Tomato Leaf Mold, Tomato Septoria Leaf Spot, Tomato Spider Mites, Two-Spotted Spider Mite, Tomato Target Spot, Tomato Yellow Leaf Curl Virus, Tomato Mosaic Virus, Tomato Healthy. Fig. 2 and Fig. 3 show some sample images for both crop species and crop diseases.

Data Preprocessing

Data preprocessing is the crucial stage of any deep learning task. Preprocessing was done on the images before putting them into the model to make it easier to extract features. The pixel value of the magnified image is a single integer in the range of 0-255 that represents the brightness of the pixel. A pixel with a value of 0 is considered black, and a pixel with a value of 255 is considered white. All input images were resized into a 256x256@3 dimension. Scaling every image to the same range [0,1]. Augmentation of images is done by the random flip with horizontal and vertical flip and random rotation with 0.2 factor. Then the preprocessed image is used as input to fit the proposed CNN model.

Classification By CNN

The color of crop leaf disease contributes significantly to need-based nitrogen management (Torikul et al., 2020). The convolutional layers in the proposed model are responsible for extracting optimal features. We demonstrated that CNN can train for classifications based

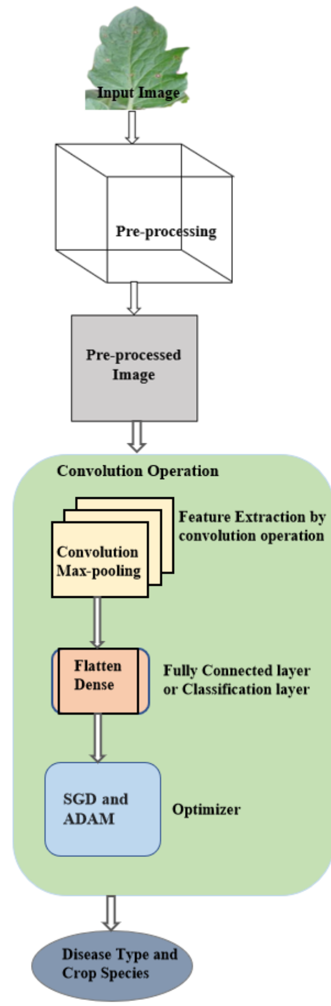

Figure 1: Proposed CNN Model

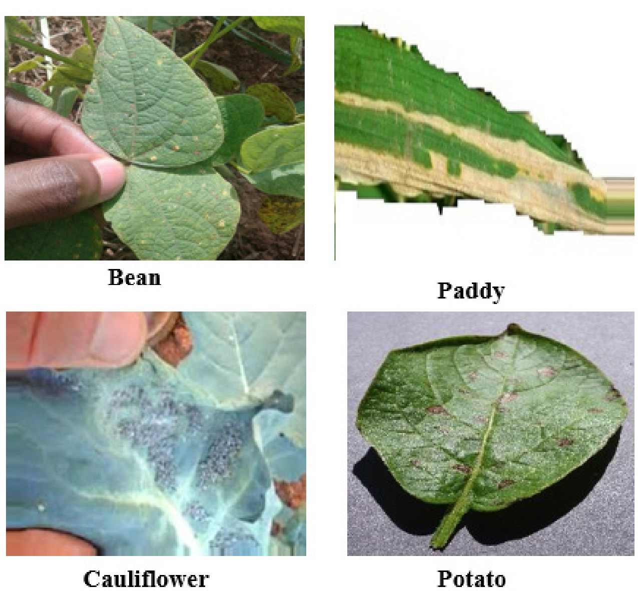

Figure 2: Crop Species

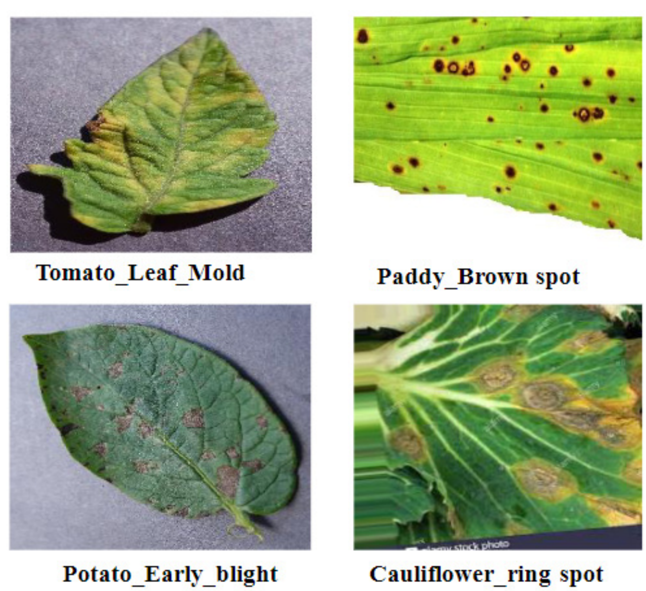

Figure 3: Crop Disease

on shape information as well as classifications based on color. We developed an optimized CNN model for crop leaf disease and species classification.

Six convolutional layers, six pooling layers, two fully connected layers, and one SoftMax layer make up our CNN architecture. Convolution is the first layer in our CNN architecture. after the convolution process pooling processes are performed. In Fig. 4, our CNN model's architecture is shown.

In the CNN, the convolution operation extracts the feature from the image. This process is done based on the given eq. (1). If P_i is the input image, k is the kernel of convolution then the output of the convolution operation P_0 can be written as,

$$P_0[x,y] = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} P_n\,[n,m].\,k[x,y] \qquad (1)$$

Where P [x, y] is the pixel value of the P coordinate. In neural networks (NN), the activation function is used to provide the non-linearity of the hidden layers. The Activation function activates the network between hidden layer nodes. Different activation functions range differently. So, the choice of the specific activation function has a vital task for the NN. In this research, Rectified Linear Unit (ReLU) activation functions are used. The ReLU activation function range is 0 to x, where x means input value. ReLU activation functions are used in all hidden layers, and SoftMax activation functions are used in the output layer. ReLU activation functions are used to solve the gradient problem. Then the normalization process is done by the eq. (2).

$$P_{a,b}^{i} = \frac{h_{a,b}^{i}}{(1+\frac{\propto}{n}\sum_{j=i-\frac{n}{2}}^{i+\frac{n}{2}}\left(h_{a,b}^{i}\right)^{2})^{\beta}} \qquad (2)$$

Here, the output of the normalization process is P_(a,b)^iand h_(a,b)^i is the output of the activation function at (a, b) coordinate. After the first convolution layer, pooling operations are performed. There are two types of pooling, namely max pooling and mean pooling. Max pooling performs based on the sharp edge's techniques. In this research, max pooling is used with size 2x2 and stride 1. After the first pooling process, the second convolution layer is performed, and the second pooling, and so on. After six convolutions and pooling, layer features are extracted, and then these features are provided as the fully connected layer for classification.

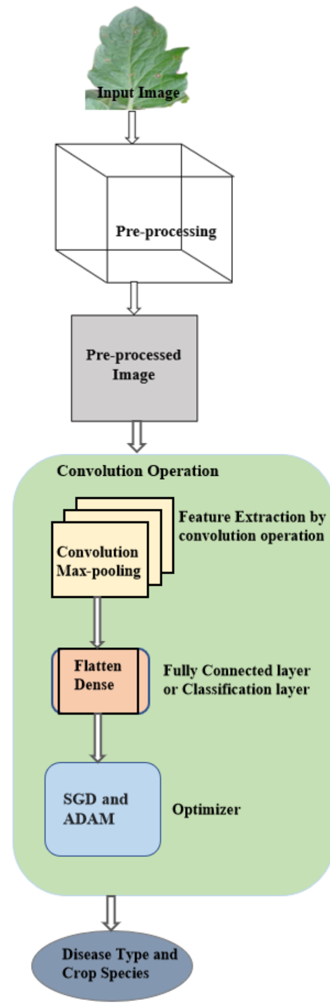

Figure 4: Proposed CNN Model Architecture.

In addition, our developed CNN architecture consists of six convolution and pooling layers. The first layer consists of 32 kernels with a 3x3 kernel size, and the second to sixth layer consists of 64 kernels with a 3x3 kernel size. Additionally, the pooling size of 2x2 is used in those layers. Finally, the flatten and output layer consists of 16448–21 neurons for disease classification and 16448–5 neurons for crop species classification.

Our CNN model is trained by two optimizers, namely stochastic gradient descent (SGD) and adaptive moment estimation (Adam). The batch size is 64, and three channels are used with 256x256@3 image resolution. The SGD is used to find the maximum or minimum value of some function. It will work with the gradient dissent for all functions. In addition, SGD is used for minimizing loss or error. The weight loss function updates are done by the eq. (3). While Adam optimizers are used for solving non-convex issues.

$$W_i = W_{i-1} - \alpha * \Delta l(p, W_{i-1}) \qquad (3)$$

Here, W_i is the current weight, W_(i-1)is the previous weight, α is the learning rate and Δl(p,W_(i-1)) is the gradient of loss function of l(p,W_(i-1)).

In this experiment, we used 14624 sample images with 21 class for disease and 14624 sample images with 5 class for crop species. For training and testing split the dataset into 80% training, 10% for testing and 10% for validation and before train the model to reduce the bias set the shuffle is true.

Experimental Analysis

In this section, we describe our experimental process and model evaluation criteria. For model performance evaluation, consider accuracy and loss of both train and validation accuracy. In this experiment, 100 epochs are used for both crop species and disease classification. After completing the training process get a 96.66% accuracy for disease classification and a 99.01% accuracy for crop species classification in the validation phase. In addition, in the testing phase, we get a 96.42% accuracy for disease classification and 99.67% for crop species classification. Furthermore, training and validation loss are low, which means the model is less overfit. All experimental outcomes are listed in table 1.

Table 1: CNN model performance comparison for Adam optimizer.

Parameter	Crop Disease	Crop Species
Training Accuracy	96.97%	99.31%
Validation Accuracy	96.66%	99.01%
Training Loss	9%	1.88%
Validation Loss	9.95%	2.17%
Testing accuracy	96.55%	99.67%

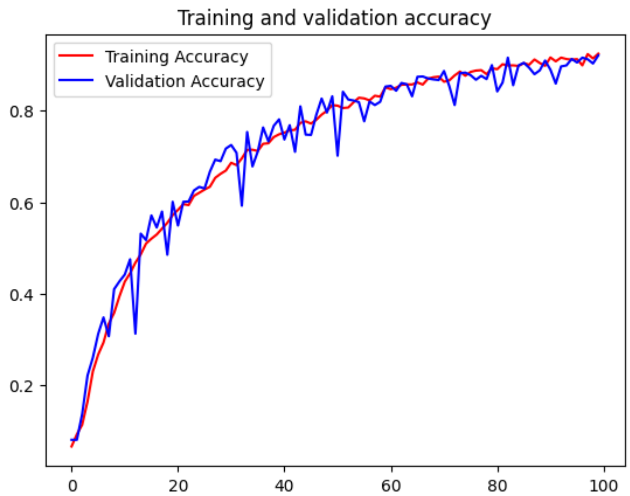

Figure 5: Crop Disease Accuracy

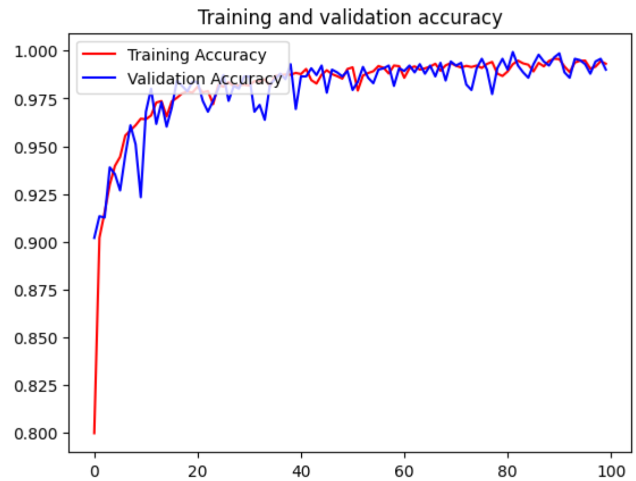

Figure 6: Crop Species Accuracy

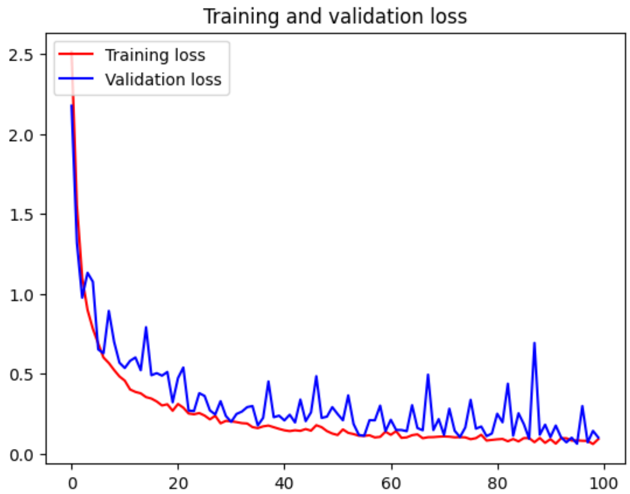

Figure 7: Crop Disease Loss

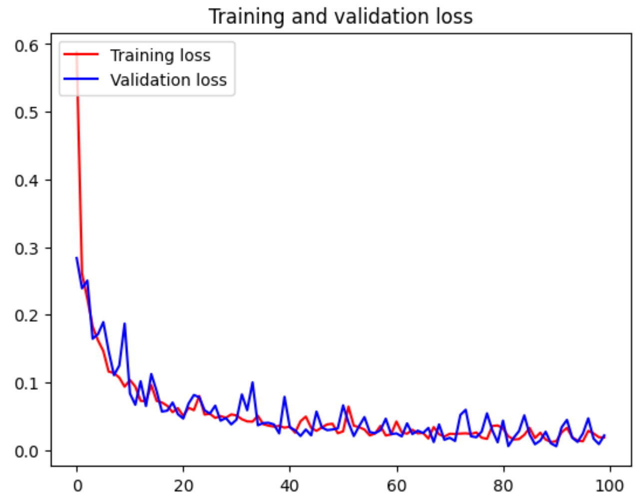

Figure 8: Crop Species Loss

In figs. 5 and 6, the training and validation accuracy is shown. Here, it is clear that in both figures, increasing the number of epochs means the proposed model learns better and almost the same as the validation set, which means the proposed model is less overfit.

On the other hand, fig.7 and fig.8 represent the model loss on the training and validation sets. Also, here we see that the proposed model produces less error and loss when going to converge.

RESULT ANALYSIS

In this section, we discuss mechanism of our develop CNN model. We see the experimental section that CNN model achieved the 96.42% and 99.67% accuracy disease and species classification respectively. Due to provide good result of our proposed CNN model that our proposed model kernels and filter work coherent way. As a result, misclassification rate is reduced. The proposed CNN model learns in better way than raw CNN model and recognized the crop disease and crop species in better way.

Misclassification

In this section, we present misclassification issues with misclassified images. The proposed CNN model misclassifies those images which are affected by the sunlight. In sunlight-affected images, pixel values change, so the proposed CNN model cannot recognize the correct diseases and species in those images. In addition, in the proposed datasets, some images have low pixel values, and those images increase the chance of misclassification. Fig.9 represents some misclassified images with their actual and predicted levels.

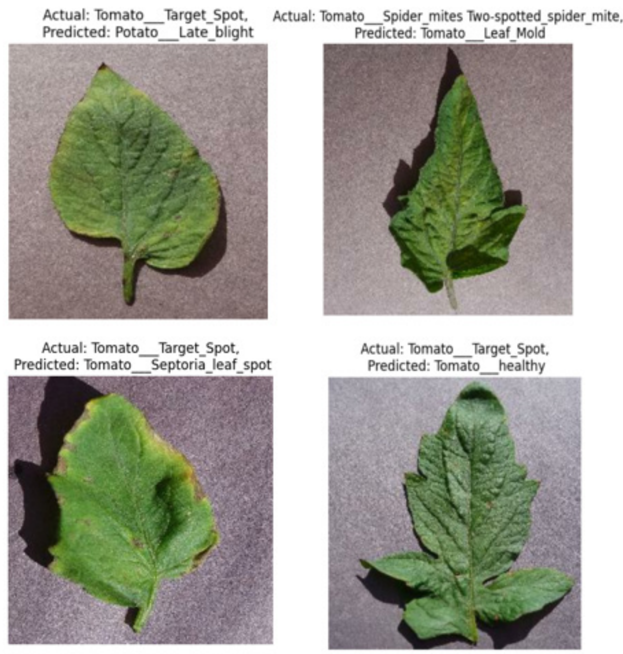

Figure 9: Misclassified Image

Here, actual level of first image is Tomato_Target_Spot but predicted level is Potato_Late_blight, actual level of second image is Tomato_Spider_mites_Two_spotted_spider_mite but predicted level is Tomato_leaf_mold, third image actual level is Tomato_Target_Spot but predicted level is Tomato_Septorial_leaf_spot and also fourth image actual level is Tomato_Target_Spot but predicted level is Tomato_healthy.

Table 2: Represents the comparison with the previous study.

Ref.	Previous Study	Proposed Study (testing accuracy)
(Adedamola et al., 2019)	93.82%	96.55%
(Adesh et al., 2021)	81.4%	
(Sambasivam et al., 2021)	93%	
(JANARTHAN et al., 2020)	95.04%	
(QINGMAO et al., 2020)	92.60%	
(Siddharth et al., 2018)	86.21%	

Comparison Of Performance Proposed CNN Model with Previous Study

The proposed model efficiently recognizes the crop species and crop disease with good accuracy. The proposed model provides 96.42% accuracy for disease classification and 99.67% accuracy for species classification on the test dataset. whereas other previous studies found less accuracy than our proposed model. Table 2 represents a comparison with the previous study.

CONCLUSION

In this research, we mainly developed a CNN model to accurately recognize the crop species, crop disease, and classification. The whole work divides into three-part (1) pre-processing (2) feature extracting (3) model training and disease recognition and classification. This process is done by the Convolution Neural Network. To reduce the sunlight effect of image data acquisition techniques are used. This work performs on five crops Bean, Cauliflower, Paddy, Potato, and Tomato. In addition, we have implemented an optimized CNN model to classify the crop species and crop disease with the best accuracy of 99.67% and 96.55% respectively.

This study helps the farmers to detect the automatically of disease class and crop species using leaf images. This implementation of the smart device can monitor the automatically of crop fields that helps the village farmer. In the future, researchers improved the data collection process to remove the sunlight effect on the image. Our study will bring unprecedented success to the farmers as well as the economy of Bangladesh.

REFERENCE

Osborne BG. Near-infrared spectroscopy in food analysis. Encyclopedia of analytical chemistry: applications, theory and instrumentation. 2006 Sep 15.

April in Bangladesh. (n.d.). Retrieved 23 October 2022, from https://people.umass.edu/wick/fulbright/2008.html

Militante SV, Gerardo BD, Dionisio NV. Plant leaf detection and disease recognition using deep learning. In 2019 *IEEE Eurasia conference on IOT, communication and engineering (ECICE)* 2019 Oct 3 (pp. 579-582). IEEE.

Hu WJ, Fan J, Du YX, Li BS, Xiong N, Bekkering E. MDFC–ResNet: an agricultural IoT system to accurately recognize crop diseases. *IEEE Access*. 2020 Jun 9; 8:115287-98.

Zhou C, Zhou S, Xing J, Song J. Tomato leaf disease identification by restructured deep residual dense network. *IEEE Access*. 2021 Feb 12; 9:28822-31.

Roy AM, Bhaduri J. A deep learning enabled multi-class plant disease detection model based on computer vision. *AI.* 2021 Aug 26;2(3):413-28.

Cap HQ, Suwa K, Fujita E, Kagiwada S, Uga H, Iyatomi H. A deep learning approach for on-site plant leaf detection. In 2018 *IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA)* 2018 Mar 9 (pp. 118-122). IEEE.

Türkoğlu M, Hanbay D. Plant disease and pest detection using deep learning-based features. *Turkish Journal of Electrical Engineering and Computer Sciences*. 2019;27(3):1636-51.

Ahmed T, Rahman CR, Abid MF. Rice grain disease identification using dual phase convolutional neural network-based system aimed at small dataset. *AgriRxiv*. 2021 Jul

8(2021):20210263534.

Islam T, Rizan RU, Tusher YA, Shafiuzzaman M, Hossain MA, Galib S. *Nitrogen fertilizer recommendation for paddies through automating the Leaf Color Chart (LCC)*. International

Journal of Advanced Computer Science and Applications. 2020;11(8).

Adedoja A, Owolawi PA, Mapayi T. Deep learning based on nasnet for plant disease recognition using leave images. In 2019 *International Conference on Advances in Big Data, Computing and Data Communication Systems (ICABCD)* 2019 Aug 5 (pp. 1-5). IEEE.

Panchal AV, Patel SC, Bagyalakshmi K, Kumar P, Khan IR, Soni M. I*mage-based plant diseases detection using deep learning. Materials Today: Proceedings*. 2021 Aug 3.

Sambasivam G, Opiyo GD. A predictive machine learning application in agriculture: Cassava disease detection and classification with imbalanced dataset using convolutional neural networks. *Egyptian Informatics Journal*. 2021 Mar 1;22(1):27-34.

Janarthan S, Thuseethan S, Rajasegarar S, Lyu Q, Zheng Y, Yearwood J. Deep metric learning based citrus disease classification with sparse data. *IEEE Access*. 2020 Sep 3; 8:162588-600.

Zeng Q, Ma X, Cheng B, Zhou E, Pang W. Gans-based data augmentation for citrus disease severity detection using deep learning. *IEEE Access*. 2020 Sep 18; 8:172882-91.

Chouhan SS, Kaul A, Singh UP, Jain S. Bacterial foraging optimization based radial basis function neural network (BRBFNN) for identification and classification of plant leaf diseases: An automatic approach towards plant pathology. *IEEE Access*. 2018 Feb 12; 6:8852-63.

Alam BR, Khatun MS, Taslim M, Hossain MA. Handling Class Imbalance in Credit Card Fraud Using Various Sampling Techniques. *American Journal of Multidisciplinary Research and Innovation*. 2022 Oct 3;1(4):160-8.

Ai Y, Sun C, Tie J, Cai X. Research on recognition model of crop diseases and insect pests based on deep learning in harsh environments. *IEEE Access*. 2020 Sep 21; 8:171686-93.