Crop Disease Detection Enhancement via EfficientNet and Convolutional Group-Wise Transformer Integration

Jyotika
PG Student, E&CE(DCN)Dept.
Guru Nanak Dev Engineering College Bidar
jyotikahalburge@gmail.com

Prof. Ramesh Patil
Associate.Professor, E&CE Dept
Guru Nanak Dev Engineering College Bidar
rameshpatil.gndec@gmail.com

Abstract: Crop diseases are a major problem for farming around the world since they can kill plants, reduce yields, and even wipe out some crops completely. This hurts farmers and the world's food supply. traditional methods for finding diseases, which typically depend on farmers and agricultural professionals looking at plants, are time-consuming, subjective, and likely to be wrong. New developments in "artificial intelligence (AI)" offer a viable alternative through real-time monitoring, automatic recognition, and smart decision-making. this is possible because to the combination of "internet of things (IoT)" and cloud computing technologies. This study looks at how different algorithms can be used for classification tasks. these algorithms include "Convolutional Neural Networks (CNN), vision Transformers (ViT), InceptionV3, EfficientNetB0 and B7, global Wavelet transform (GWT), Xception, NasNetMobile, CVT (CNN + ViT), ICVT (Inception + ViT), and EGWT (EfficientNet + GWT). We use advanced models like YOLOV5x6, YOLOV5s6, YOLOV8n, and YOLOV9n to find problems with crop leaves". in addition, datasets like PlantVillage data, Cassava data, and Tomato data are used to make crop disease identification more accurate and quicker. The goal of this research is to give farmers the latest technologies to help lessen the effects of crop diseases on agricultural productivity.

***"Index Terms -** Crop Disease Detection, Convolutional Neural Networks (CNN), Vision Transformers (Vit), Hybrid Models, Transfer Learning, Object Detection, YOLO, Deep Learning, Global Wavelet Transform (GWT)".*

1. INTRODUCTION

For the health and production of crops, it is important to find and diagnose crop diseases early and accurately. good identification and management could lead to more crops, better quality crops, and better protection of the agroecosystem. Deep learning algorithms are a great technique to analyze crop images.

"Since 2012, Convolutional Neural Networks (CNNs)" have changed and improved, and they are now vital for computer vision[10]. "Vision Transformers (ViTs)" have become more popular recently since they can handle long-range dependencies and "do well on visual tasks[facsimiles[1][16]. CNNs, ViTs, and hybrid designs" that combine their strengths have done quite well at finding diseases in crops [3, 19].

these models work well because they were trained on large "public datasets like ImageNet[10], common objects in Context (COCO)[11], and MIT places[12]", which provide a strong base for transfer learning. it is hard to employ these DL models because they have so many parameters, which makes the calculations more complicated. ViTs, which sometimes improve model parameters when performance improves, are especially affected [19]. To solve this problem, researchers are improving deep learning approaches for finding agricultural diseases.

2. RELATED WORK

A lot of studies have used AI and deep learning to find and categorize illnesses in crops. "Convolutional Neural Networks (CNNs)" have proven very important to this study since they are good at tasks that include images. Kurmi et al. built a deep CNN model that does a decent job of classifying crop diseases from pictures of leaves[4]. Agarwal et al. also employed CNNs to find diseases in tomato crops, which shows how strong and scalable they are [5].

"Vision Transformers (ViTs)" are based on the success of CNNs and show long-range picture interdependence. Fu et al. improved ViT-based crop pest photo detection for patterns that are hard to see [1]. Using convolutional networks and transformers, Wang et al. also built a hybrid architecture for classifying regional agricultural diseases that worked better than single models[3].

these models are better now that they employ transfer learning. Hassan et al. employed CNN-based transfer learning to find plant diseases on different types of datasets more quickly and accurately[6]. Chen et al. used deep transfer learning to find illnesses in rice plants, showing that pretrained models may be adapted [20][21].

not just the design of the model, but also the quality and variety of the dataset affect how well it can find things. "Many people have utilized PlantVillage to train and test DL models[6][20]". we have also used cassava and tomato datasets to test how well models work with different types of crops [5, 21].

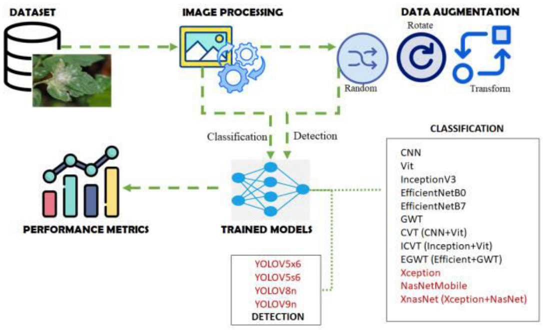
recent improvements in models for identifying objects have also aided the field. "YOLOv5x6, YOLOv5s6, and YOLOv8n are examples of YOLO-based" designs that can swiftly and accurately find and identify crop leaves that are sick[2][5].

Even with recent improvements, the costs of computing and the complexity of models, especially those based on transformers, are still high. "The goal of making lightweight hybrid models like CVT (CNN + ViT) and ICVT (Inception + ViT) is to find a balance between performance and efficiency[3][18]. EGWT (EfficientNet + GWT) can make the best use of resources with greathigh accuracy[19]".

Advanced DL, hybrid architectures, transfer learning, and datasets are all helping to come up with new ways to find agricultural diseases. this could be good for sustainable farming.

3. MATERIALS AND METHODS

The proposed system uses strong AI algorithms in a strong framework to change the way agricultural diseases are found and classified. The system uses a number of classification methods to look at the health of plants. "these include CNN[6], ViT[1][16][19], InceptionV3[10], EfficientNetB0 and EfficientNetB7[4], GWT[19], Xception[4], and NasNetMobile[21]. Hybrid models like CVT (CNN + ViT)[3][18], ICVT (Inception + ViT)[1][7], and EGWT (efficient + GWT)[19]" are used with these algorithms to make classification more accurate. The method uses "the YOLO family of models, such as YOLOV5x6, YOLOV5s6, YOLOV8n, and YOLOV9n, to quickly" and accurately find problems with crop leaves[2][5]. The system works with datasets as "Plant Village data[6][20], Cassava data[21], and Tomato data[5][20] to include crop diseases". Farmers can use this set of tools to get real-time information about the health of their crops and the amount of food they produce.



“Fig.1 Proposed Architecture”

This figure (Fig. 1) depicts how a deep learning-based image processing system sorts and finds things. After a dataset is produced, image processing steps like random transformations and rotations are used to add more information. “Trained CNNs, Vits, and hybrid models like EfficientNetB0, EfficientNetB7, and Xception are used to sort images. Detection is done with YOLO models like YOLOv5”. using the right metrics for evaluation makes sure that the analysis is quick and the classification and detection results are accurate.

i) “Dataset Collection:”

The suggested method uses a lot of datasets to train and test models that may find crop diseases. these datasets include a lot of information on how well models work with different types of crops and diseases.

Plant Village Dataset: PlantVillage is one of the most used databases for classifying plant diseases. It has pictures of more than 50 crop types and 14 disease types[6][20]. This collection of labeled pictures is useful for training DL models to identify and classify plant diseases. This dataset has a wide range of crops and diseases, which lets models work with many types of crops, making them more useful in real-world farming situations.

“Fig.2 Dataset Collection 38 classes”

Cassava Dataset: Cassava is a big crop in a lot of places. The Cassava dataset finds sick cassava plants. The photos of CBSD and CMD are labeled so that people may tell what kind of ailment they have. people typically use this dataset to teach ML algorithms how to find diseases in cassava crops [21]. this means that the suggested strategy can be utilized in cassava farming, where keeping an eye on diseases is important for keeping the production high.

“Fig.3 Dataset Collection - 5 classes”

Tomato Dataset: The Tomato dataset has a lot of photographs of tomato diseases, which makes it valuable for finding and classifying them [5][20]. This collection has photographs of early blight, late blight, and the tomato yellow leaf curl virus. Tomato growers need this dataset because it has a wide range of photographs that help the algorithm find and sort tomato plant diseases.

“Fig.4 Dataset Collection - 11 classes”

in order for the suggested system to perform well with all types of crops and diseases, its algorithms must be developed, verified, and tested on various datasets. using these different datasets, the system gives farmers real-time, useful information to help them keep their crops healthy.

ii) Image Pre-Processing:

Preprocessing images makes DL models better at finding agricultural diseases and making predictions. ImageDataGenerator is used by the proposed system to add to and process incoming photographs. This method gives the model a variety of well-processed data, which makes it stronger and better able to handle changes in pictures. these are the steps that are taken before processing:

(a) Re-scaling the Image: To preserve input "data between 0 and 1, pixel values are normalized by dividing them by 255". keeping the input values the same during normalization speeds up model convergence and makes it more accurate [4].

(b) Shear Transformation: Shear transformations change the picture's angles and points of view by moving it in a random direction. by making the model less sensitive to small geometric changes, this strategy makes it less sensitive to different viewing angles and crop orientations[6]. Shearing adds to the training data variety by changing the data.

(c) Zooming the Image: Random zooming makes the distances between objects in pictures the same. when you show the model different levels of magnification during preprocessing, it gets better at finding diseases at different sizes. Zooming also makes the model less sensitive to the size of the plants and the cropping[1][5].

(d) Horizontal Flip: This preprocessing step adds differences to the dataset, especially for models that look at plant photos from different angles. Horizontal flipping helps the model learn more general traits, which makes it more resistant to plant orientations when it is deployed. [4][19].

(e) Reshaping the Image: The design of the neural network needs a picture size that stays the same. This stage always processes all of the images in a way that makes it easy for the model to use them. To train a model well, the pictures need to be the same size all the time so the network can handle them correctly. [3][6].

The ImageDataGenerator does these steps to prepare the training data, which gives the model a lot of different picture types to work with. This makes the model better at generalizing and finding crop diseases in real-world farming situations. [6][5][19].

iii) "Training & Testing:"

We make training and validation datasets using processed pictures to teach the crop disease identification algorithm. The training dataset is sent through the network to teach the model how to recognize and understand patterns and traits of crop diseases. "CNN, ViT, and hybrid models like CVT and ICVT let the model learn from data. Stochastic gradient descent (SGD)" or Adam minimize the loss function, which is commonly categorical cross-entropy, during training to make the model more accurate [6][4].

We evaluate the model's ability to generalize on a dataset that is different from the training data. "After training, the model is tested on data it hasn't seen before to check how accurate, precise, and recall it is, as well as its F1 score". tests show that the model is strong and can find crop diseases in new, real-world pictures[1][5].

iv) Algorithms:

(a) "Classification Models:"

CNNs CNNs are DL algorithms that look at structured grid data, mostly photographs. CNNs are great at finding and classifying features in crop disease detection because they learn how to group data in space through convolutional layers. these models are very good at spotting patterns in agricultural diseases, which makes it easier to find and suggest ways to treat them. [4][6].

ViT (vision Transformer) treats image patches as sequences, which makes it easy to quickly extract features and classify crop diseases. ViT improves sickness diagnosis by leveraging self-attention processes to make the model better at recognizing complex image patterns and connections. [3] [7].

InceptionV3 does a good job at classifying agricultural diseases by using varying filter sizes in a layer to capture a number of distinct features. This model finds intricate patterns with high accuracy and speed, making it perfect for finding agricultural diseases in real time. [5] [6].

EfficientNetB0 is a lightweight convolutional neural network that uses compound scaling to find the right balance between model size and performance. it can find crop illnesses very quickly and accurately, even when there aren't a lot of computers available. "This makes it perfect for farming situations where resources are restricted. [5] [6]. advanced EfficientNetB7 increases depth, breadth, and resolution" to improve categorization. because of its optimized layout, the model can as it should be diagnose crop diseases in real time for precision farming. [5][6].

Group-wise attention and convolutional networks enable GWT find and show features. This mixed method helps provide a better picture of how different crop images interact with each other and makes categorization more accurate by focusing on the parts of the picture that are most important to the disease. [7]. "Xception uses depthwise separable convolutions" to make feature extraction better. it can tell when a plant is sick and grasp sophisticated patterns, "which makes it a powerful tool for classifying crop diseases [5]. [6]. NasNetMobile is very accurate and cheap" to run on mobile and edge devices. devices with limited resources can run its design well, which makes it possible to monitor crop diseases in real time in agricultural settings. [6] [7].

CVT uses CNNs and vision Transformers together to get the most out of each. This hybrid method increases feature extraction and classification by recognizing complex patterns in crop images. This makes it easier to find crop diseases. [4] [7].

ICVT (Inception + ViT): This combines Inception and ViT to make the most of what they do best. It captures complicated features and interactions, which makes it perfect for finding and classifying crop diseases in different situations. [4] [7].

Combining ***EfficientNet*** with GWT to focus on important picture features makes categorization better while keeping costs down. This technology makes it easier to diagnose crop diseases and lets farmers respond quickly to problems. [5] [7].

By merging ***Xception and NasNet, XNasNet*** strikes a compromise between speed and accuracy. This lightweight architecture lets you monitor diseases in real time on devices with limited resources, which makes it great for finding and telling apart agricultural diseases[5][6].

(b) "Detection Models:"

YOLOV5x6 is a fast and accurate improvement on the "YOLO (You only look once)" object identification method. YOLOV5x6 finds problems in crop pictures to find crop illnesses in real time and push for quick fixes in farming. [4] [6].

YOLOV5s6 is a small, optimized way to find objects with YOLO. It swiftly finds signs of crop disease since it is fast and accurate. This model is great for real-time applications that need to be analyzed and acted on right away. [4][7].

YOLOV8n is a new and improved version of YOLO for finding objects. It accurately finds disease-related features in pictures of crops, which makes it possible to keep an eye on and control crop health. [6] [7].

YOLOV9n is the newest YOLO model, and it is great at finding objects in real time. Its advanced design finds and sorts crop diseases, which lets farmers respond quickly and effectively. [6][7].

4. RESULTS & DISCUSSION

Accuracy: The test's accuracy is how well it can tell the difference between sick and healthy people. We need find out how many of the evaluated cases were true positive and true negative in order to figure out how accurate a test is. this can be said mathematically as:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} (1)$$

Precision: Precision looks at how many of the samples or cases that were labeled as positives were actually right. So, the formula for figuring out the precision is as follows:

$$Precision = \frac{\text{True Positive}}{\text{True Positive } + \text{False Positive}} (2)$$

Recall: In ML, recall is a measure of how well a model can find all the relevant examples of a certain class. it is the number of correctly predicted positive observations divided by the total number of real positives. This gives you an idea of how well a model captures instances of a certain class.

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} (3)$$

F1-Score: The F1 score is a way to check how accurate a ML model is. "It adds together the precision and recall scores of a model". The accuracy statistic counts how many times a model produced a valid prediction on the whole dataset.

$$F1\ Score = 2 * \frac{Recall \text{ X } Precision}{Recall + Precision} * 100 (4)$$

mAP: "mean average Precision (MAP)" is a way to measure how good a ranking is. It looks at how many relevant suggestions there are and where they are on the list. To find MAP at k, you take the average of the "average Precision (AP)" at k for all users or queries.

$$mAP = \frac{1}{n} \sum_{k=1}^{k=n} AP_k \ (5)$$

Table 1 shows how well each method does on "the performance metrics: accuracy, precision, recall, and F1-score. The Extension and YoloV5s6 for classification and Detection get the best marks". We also show the stats of other algorithms for comparison.

"Table.1 Performance Evaluation Metrics of Classification"

Model	Accuracy	Precision	Recall	F1 Score
CNN	0.688	0.710	0.673	0.685
ViT	0.615	0.615	0.615	0.615
InceptionV3	0.515	0.523	0.385	0.431
EfficientNetB0	0.615	0.615	0.615	0.615
EfficientNetB7	0.615	0.615	0.615	0.615
GWT	0.615	0.615	0.615	0.615
Xception	0.978	0.978	0.978	0.978
NASNetMobile	0.979	0.980	0.979	0.979
CVT	0.649	0.734	0.600	0.645
ICVT	0.513	0.512	0.423	0.453
EGWT	0.607	0.655	0.572	0.600
XNASNet	0.999	1.000	0.997	0.999

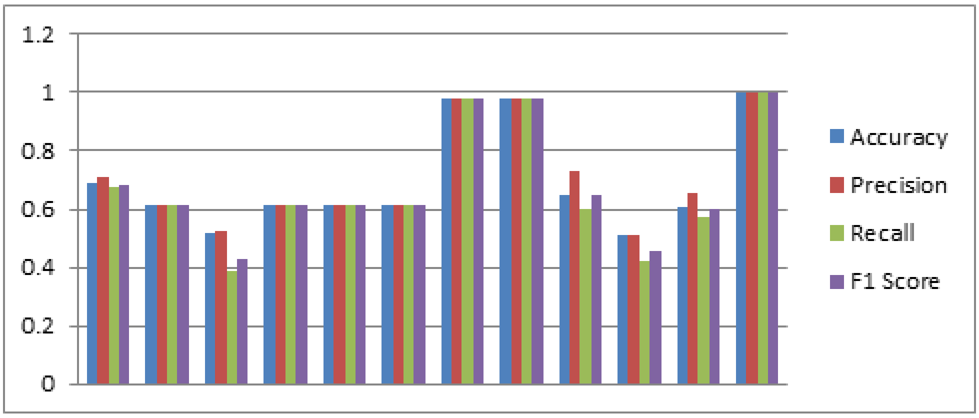
“Table.2 Performance Evaluation Metrics of Classification – Plant Village”

Model	Accuracy	Precision	Recall	F1 Score
CNN	0.895	0.914	0.885	0.898
ViT	0.992	0.993	0.992	0.992
InceptionV3	0.987	0.987	0.987	0.987
EfficientNetB0	0.198	0.198	0.198	0.198
EfficientNetB7	0.197	0.197	0.197	0.197
GWT	0.928	0.936	0.924	0.929
Xception	1.000	1.000	1.000	1.000
NASNetMobile	1.000	1.000	1.000	1.000
CVT	0.989	0.989	0.989	0.989
ICVT	0.995	0.995	0.995	0.995
EGWT	0.852	0.947	0.765	0.839
XNASNet	1.000	1.000	1.000	1.000

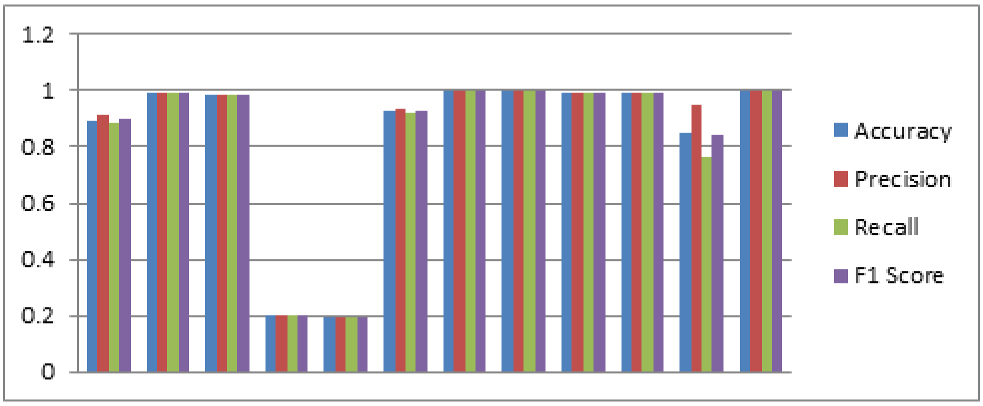
“Table.3 Performance Evaluation Metrics of Classification – Tomato”

Model	Accuracy	Precision	Recall	F1 Score
CNN	0.898	0.923	0.880	0.894
ViT	0.119	0.000	0.000	0.000
InceptionV3	0.827	0.833	0.826	0.828
EfficientNetB0	0.134	0.134	0.134	0.134
EfficientNetB7	0.119	0.119	0.119	0.119
GWT	0.118	0.118	0.118	0.118
Xception	0.991	0.995	0.984	0.988
NASNetMobile	0.867	0.868	0.866	0.988
CVT	0.906	0.942	0.861	0.888
ICVT	0.861	0.887	0.829	0.848
EGWT	0.145	0.000	0.000	0.000
XNASNet	0.918	0.940	0.893	0.908

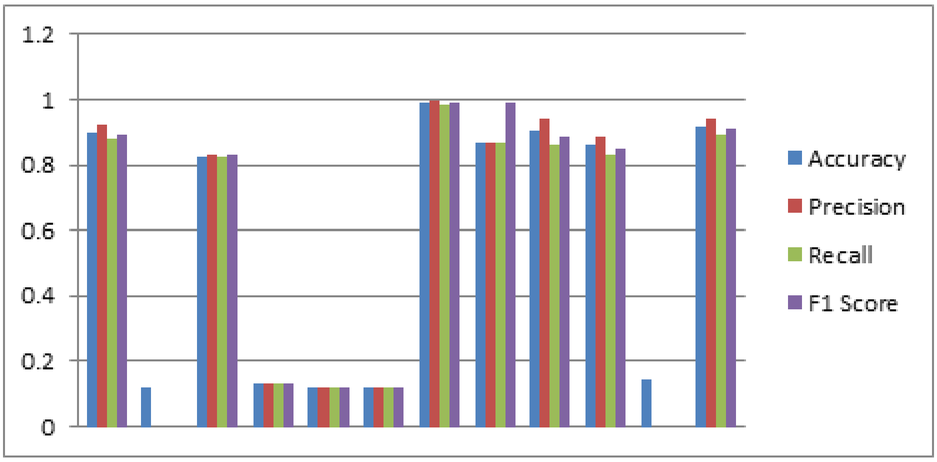
“Graph.1 Comparison Graphs of Classification – Cassava”



“Graph.2 Comparison Graphs of Classification – Plant Village”



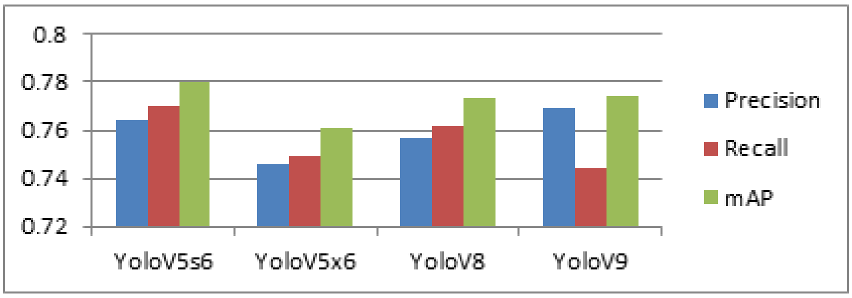
“Graph.3 Comparison Graphs of Classification – Tomato”



“Table.4 Performance Evaluation Metrics of Detection- All 3 Datasets”

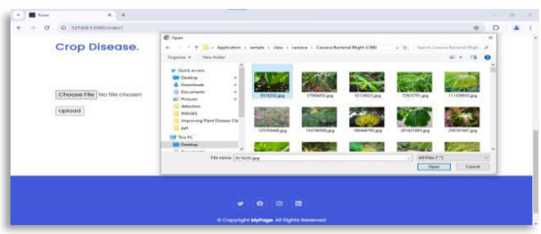
Model	Precision	Recall	mAP
YoloV5s6	0.764	0.770	0.780
YoloV5x6	0.746	0.749	0.761
YoloV8	0.757	0.762	0.773
YoloV9	0.769	0.744	0.774

“Graph.4 Comparison Graphs of Detection – All 3 Datasets”



In graphs that show “classification, accuracy is light blue, precision is maroon, recall is green, and F1-score is violet. In detection graphs (1–4), precision is light blue, recall is maroon, and mAP is green. Xception and YOLO have the greatest values”, which means they do better than all the other models. these results are shown in the graphs above.

“Fig.5 Test case 1 for Classification Dataset-1”

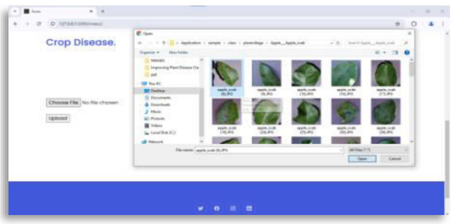


This picture seems like it shows a file upload interface for your system for finding crop diseases.

“Fig.6 Results of Test case 1 for Classification Dataset 1”

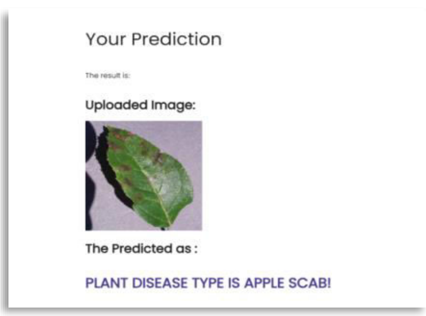
This picture displays what your crop disease detection system thinks will happen. It correctly recognized the uploaded picture as "Cassava Bacterial Blight (CBB)."

“Fig.7 Test case 1 for Classification Dataset-2”



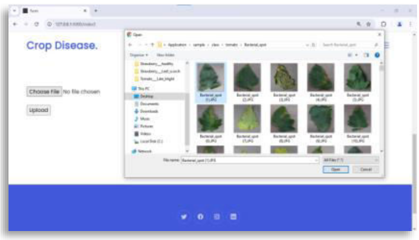
This picture displays a popup for choosing a file to upload an apple leaf with "Apple Scab" illness to a system that detects agricultural diseases. users can choose and upload an image for prediction using the interface.

“Fig.8 Results of Test case 1 for Classification Dataset 2”



This picture shows the results of a plant disease categorization system's forecast. The algorithm accurately identifies the ailment as Apple Scab in the uploaded picture of an apple leaf.

“Fig.9 Test case 1 for Classification Dataset-3”

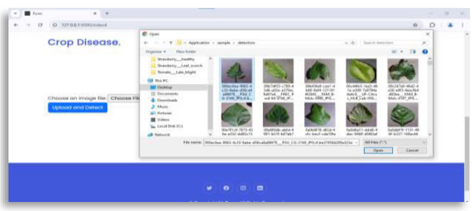


the part of your Crop disease Detection net app where you choose a tomato leaf picture with Bacterial Spot to upload.

“Fig.10 Results of Test case 1 for Classification Dataset-3”

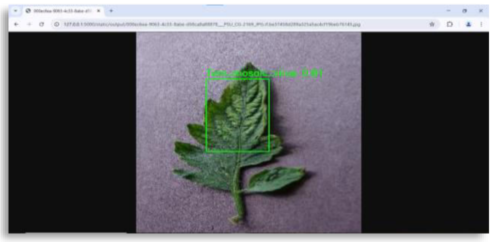
The model properly named the plant disease Tomato Bacterial Spot!

“Fig.11 Test case 1 for Detection Dataset-1”



a web software that helps you find plant diseases by letting you submit files with different leaf photos for examination and selection.

“Fig.12 Results of Test case 1 for Detection Dataset-1”



A plant disease detection algorithm found "Tomato Mosaic Virus with 81% confidence” and used a bounding box to show the area of the leaf that was afflicted.

5. CONCLUSION

To sum up, “the Xception, NasNetMobile, and XNasNet” models all operate well with crop disease datasets. these models did a better job of classifying since they used DL architectures and better feature extraction. “NasNetMobile's” lightweight yet powerful design worked effectively on mobile and resource-limited devices for real-time agricultural surveillance. “Xception's depthwise” separable convolutions, on the other hand, accurately recognized complex disease patterns. XNasNet had the best balance of accuracy and efficiency because it combined the strengths of Xception and NasNet. This made dataset classification more accurate.

The “YOLOV5x6, YOLOV5s6, YOLOV8n, and YOLOV9n algorithms” found problems in leaf images that might be used to diagnose diseases. YOLO models, which can find things in real time, swiftly and accurately found indicators of disease. YOLOV9n was quite good at finding things and digesting them quickly. these strong algorithms for discovering and classifying crop diseases give quick and precise results that help farmers manage their crops better.

Future Scope could add pest detection and yield prediction to this model and try it out on a lot of different types of crops and weather conditions. To make sure the model can grow and work well, its hyperparameters will be tuned to find the right balance between performance and complexity. more varied datasets and improved generalization will be added to the model to make it more useful in farming.

REFERENCES

[1] X. Fu, Q. Ma, F. Yang, C. Zhang, X. Zhao, F. Chang, and L. Han, ‘‘Crop pest image recognition based on the improved ViT method,’’ Inf. Process. Agricult., Feb. 2023.

[2] R. Reedha, E. Dericquebourg, R. Canals, and A. Hafiane, ‘‘Transformer neural network for weed and crop classification of high resolution UAV images,’’ Remote Sens., vol. 14, no. 3, p. 592, Jan. 2022.

[3] Y. Wang, Y. Chen, and D. Wang, ‘‘Convolution network enlightened transformer for regional crop disease classification,’’ Electronics, vol. 11, no. 19, p. 3174, Oct. 2022.

[4] Y. Kurmi, P. Saxena, B. S. Kirar, S. Gangwar, V. Chaurasia, and A. Goel, ‘‘Deep CNN model for crops’ diseases detection using leaf images,’’ Multidimensional Syst. Signal Process., vol. 33, no. 3, pp. 981–1000, Sep. 2022.

[5] M. Agarwal, S. K. Gupta, and K. K. Biswas, ‘‘Development of efficient CNN model for tomato crop disease identification,’’ Sustain. Comput.,Informat. Syst., vol. 28, Dec. 2020, Art. no. 100407.

[6] S. M. Hassan, A. K. Maji, M. Jasiński, Z. Leonowicz, and E. Jasińska, ‘‘Identification of plant-leaf diseases using CNN and transfer-learning approach,’’ Electronics, vol. 10, no. 12, p. 1388, Jun. 2021.

[7] H.-P. Wei, Y.-Y. Deng, F. Tang, X.-J. Pan, and W.-M. Dong, ‘‘A comparative study of CNN- and

transformer-based visual style transfer,'' J. Comput. Sci. Technol., vol. 37, no. 3, pp. 601–614, Jun. 2022.

[8] R. Ahuja and S. C. Sharma, ''Transformer-based word embedding with CNN model to detect sarcasm and irony,'' Arabian J. Sci. Eng., vol. 47, no. 8, pp. 9379–9392, Aug. 2022.

[9] J. Fang, H. Lin, X. Chen, and K. Zeng, ''A hybrid network of CNN and transformer for lightweight image super-resolution,'' in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2022, pp. 1103–1112.

[10] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, ''ImageNet large scale visual recognition challenge,'' Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, Dec. 2015.

[11] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick, ''Microsoft COCO: Common objects in context,'' in Computer Vision—ECCV, Zurich, Switzerland. Springer, 2014, pp. 740–755.

[12] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, ''Places: A 10 million image database for scene recognition,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 6, pp. 1452–1464, Jun. 2018.

[13] S. U. Khan, N. Khan, T. Hussain, and S. W. Baik, ''An intelligent correlation learning system for person re-identification,'' Eng. Appl. Artif. Intell., vol. 128, Feb. 2024, Art. no. 107213.

[14] A. Parashar, A. Parashar, A. F. Abate, R. S.Shekhawat, and I. Rida, ''Real-time gait biometrics for surveillance applications: A review,'' Image Vis. Comput., vol. 138, Oct. 2023, Art. no. 104784.

[15] H. Wang, Y. Zhu, B. Green, H. Adam, A. Yuille, and L.-C. Chen, ''Axial-DeepLab: Stand-alone axial-attention for panoptic segmentation,'' in Proc. Eur. Conf. Comput. Vis. Springer, 2020, pp. 108–126.

[16] Y. Xu, Z. Zhang, M. Zhang, K. Sheng, K. Li, W. Dong, L. Zhang, C. Xu, and X. Sun, ''Evo-ViT: Slow-fast token evolution for dynamic vision transformer,'' in Proc. AAAI Conf. Artif. Intell., 2022, vol. 36, no. 3, pp. 2964–2972.

[17] M.-H. Guo, C.-Z. Lu, Z.-N. Liu, M.-M. Cheng, and S.-M. Hu, ''Visual attention network,'' Comput. Vis. Media, vol. 9, no. 4, pp. 733–752, Dec. 2023.

[18] X. Pan, T. Ye, Z. Xia, S. Song, and G. Huang, ''Slide-transformer: Hierarchical vision transformer with local self-attention,'' in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2023, pp. 2082–2091.

[19] Y. Fang, X. Wang, R. Wu, and W. Liu, ''What makes for hierarchical vision transformer?'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 10, pp. 12714–12720, Oct. 2023.

[20] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, ''Using deep transfer learning for image-based plant disease identification,'' Comput. Electron. Agricult., vol. 173, Jun. 2020, Art. no. 105393.

[21] J. Chen, D. Zhang, Y. A. Nanehkaran, and D. Li, ''Detection of Rice plant diseases based on deep transfer learning,'' J. Sci. Food Agricult., vol. 100, no. 7, pp. 3246–3256, May 2020.