Indian Journal of Computer Science and Technology
https://www.doi.org/10.59256/indjcst.20260501034
Volume 5, Issue1 (January-April 2026), PP: 243-250.
www.indjcst.com
ISSN No: 2583-5300

Hybrid EfficientNet-B0 and Vision Transformer Framework for Context-Aware Crop Disease Detection from Agricultural Images

T C Swetha Priya[1], Poosa Rushika[2], Naragudam Sanjana[3], Megi Sindhuja[4]

[1]Assistant Professor, Department of Information Technology, Stanley College of Engineering & Technology for women, Hyderabad, Telangana, India.

[2,3,4]UG Scholar, Department of Information Technology, Stanley College of Engineering & Technology for women, Hyderabad, Telangana, India.

To Cite this Article: *T C Swetha Priya[1], Poosa Rushika[2], Naragudam Sanjana[3], Megi Sindhuja[4], "Hybrid EfficientNet-B0 and Vision Transformer Framework for Context-Aware Crop Disease Detection from Agricultural Images, Volume 05, Issue 01 (January-April 2026), PP: 243-250.*

Abstract: *Crop diseases are one of the major problems in agriculture. It has a significant impact on the productivity of the crops as well as the lives of the farmers. Detection of diseases in plants at an early stage is important in order to avoid damage to the crops and enhance the productivity of agriculture. However, the traditional method of plant disease detection is based on observation, which is a tedious process and might result in incorrect identification of diseases due to a lack of knowledge. This paper suggests an efficient method of plant disease detection using images and the EfficientNet-B0 and ViT.The EfficientNet-B0 model is a Convolutional Neural Network model that is used in the extraction of important features in images related to plant leaves, such as color changes, textures, and diseases, whereas the Vision Transformer model is used in the identification of the relationships between different areas in the images using a self-attention mechanism in order to ensure better accuracy in the classification of the images. The model is trained and validated using different datasets, including PlantVillage, PlantDoc, Vegetable images, etc. Depending upon the disease, fertilizer suggestions will also be provided with images, usage guidelines, dosage, and precautions for proper usage of the fertilizer by the farmers. Moreover, a multilingual voice assistant is also integrated into the system for providing information in English, Hindi, and Telugu, thus making it more user-friendly for farmers models.*

Key Words: *Crop Disease Detection, Deep Learning, CNN, Vision Transformer (ViT), EfficientNet-B0, Image Classification, Fertilizer Recommendation System, Multilingual Voice Assistant.*

I. INTRODUCTION

Agriculture plays a vital role in ensuring food security at the global level and supporting the lives of millions of farmers worldwide. Despite its importance, crop diseases pose a major challenge that has been affecting crop yield, crop quality, and the economy. Accurate and timely identification of crop diseases is essential to manage the crops effectively. Traditional approaches to crop disease diagnosis depend on visual inspection by experts or farmers. This approach may take more time and may not be accurate due to human error.Computer vision and deep learning techniques have been successful in developing an automated system for crop disease detection using digital images. Convolutional Neural Networks have been successful in extracting local features such as texture, shape, and color from leaf images. Even though the CNN approach has been successful in crop disease detection, there may be difficulties in capturing long- range dependencies and global context information in an image. Recently, Vision Transformers have been proposed as an efficient approach to crop disease detection using the self-attention mechanism, which captures global context information effectively. In this paper, an image-based crop disease detection system using the combined power of CNN and Vision Transformer architectures will be proposed. Such a hybrid approach helps to ensure accurate classification results through precise local feature extraction as well as overall global context understanding. The system is trained and tested using well-known datasets such as PlantVillage, PlantDoc, and vegetable images. In addition to disease detection, this proposed system proves to be beneficial for farmers as it also suggests the usage of fertilizers with images and detailed information regarding their usage, including dosage, frequency, etc. To ensure that this system is easily accessible to farmers who may not speak English, a voice assistant is also integrated, which speaks out recommendations in English, Hindi, and Telugu using text-to-speech technology. Such a system, which includes state-of-the-art deep learning techniques as well as information regarding agricultural practices, proves to be highly beneficial for reducing losses, increasing productivity, etc.

II. LITERATURE REVIEW

In (Shandilya et al.,2025), a hybrid deep learning method for maize leaf disease detection by using Convolutional Neural Networks (CNN) and Vision Transformers (ViT) has been proposed. Maize crops are usually infected with leaf diseases. Therefore, the early and accurate detection of maize leaf diseases plays a vital role in improving crop yields. However, traditional approaches are not effective enough to classify disease patterns on leaf images. To avoid this problem, a proposed model uses a combination of CNN for local feature extraction and ViT for global contextual relationships by using a self-attention mechanism. The proposed model achieved high accuracy by training on maize leaf images obtained from the Kaggle and Mendeley datasets with the help of various preprocessing and data augmentation techniques. The proposed model achieved an accuracy of 99.15%, precision, recall, and F1-score of 99.13%, indicating that the proposed model performs well when compared with CNN model results[1].

A hybrid deep learning model was suggested in (Sinamenye et al .,2025) to enhance potato plant disease detection. Crop diseases play a crucial role in crop productivity. Manual detection of crop diseases is not accurate in real-world conditions. Therefore, a hybrid model was suggested to improve crop disease detection. The suggested model is a hybrid model of EfficientNetV2B3 (CNN) and Vision Transformer (ViT). The EfficientNetV2B3 model can accurately detect visual features from images of potato leaves. The suggested model also considers the global relationship within images using ViT. The model was trained using the Potato Leaf Disease Dataset, in which images are taken in real-world conditions. The experimental results of the suggested model were accurate to an extent of 85.06%, which is 11.43% higher than previous models[2].

In the year (Aboelenin et al.,2025), a hybrid deep learning framework was proposed for the detection and classification of plant leaf diseases in intelligent agriculture. The detection of plant diseases is difficult due to the various crops and the variations in the diseases. The proposed method is based on the combination of Convolutional Neural Network (CNN) and Vision Transformers (ViT) for the improvement of classification accuracy. In the proposed method, CNN models such as VGG16, Inception-V3, and DenseNet20 are used for the extraction of crucial features from the images of the plant leaves, and Vision Transformers are used for the detection of the patterns for the improvement of plant disease detection. The proposed method was tested using the Apple and Corn plant datasets, each containing four classes. The accuracy of the proposed method was found to be 99.24% for the Apple plant and 98% for the Corn plant[3].

(Roy and Kukreja, 2025) A research paper proposed the use of Vision Transformers (ViT) for the detection of rice leaf diseases and the estimation of their severity levels. Rice leaf diseases play a significant role in the quality and quantity of the rice crop. Traditional approaches to detecting rice leaf diseases are time- consuming and costly. A research paper proposed the application of Vision Transformers to identify the type of disease and the level of severity. A dataset of 3,345 annotated rice leaf images with 10 types of diseases and 3 levels of severity were used to train the proposed model. Rotation and flipping were used as data augmentation techniques. In the proposed model, the Vision Transformer architecture uses the multi-head self-attention mechanism. In the experimentation process, the proposed model showed a weighted F1- score of 54.17% and 77.94% with an AUC score of 0.86 to identify the type of disease and the level of severity. It showed good capability to distinguish the difference between the various conditions of the diseases[4].

Furthermore, in the year (Maeda-Gutiérrez et al.,2025), a research paper has proposed a method to diagnose the diseases of the leaves of the plant "Taro," which is scientifically known as "Colocasia Esculenta," with the use of AI techniques. In this paper, the authors compared the performance of the Inception V3, ResNet50, and Vision Transformer models. In this study, SHAP (Shapley Additive explanations) has also been used to improve the interpretability of the models. The results were compared, showing that Inception V3 performed well to diagnose diseases with 99.85% accuracy, F1 score, and recall, along with 99.92% specificity. SHAP has also been used to emphasize the features in the leaf images that are important to the decision-making process. It can be concluded that the proposed method can be used to improve plant disease detection using deep learning techniques along with explainable AI[5]

.In (Mehnaz and Islam.,2025), research work was conducted on the comparison of different computer vision approaches in the detection of rice leaf diseases using the Dhan-Shomadhan dataset, which originated in Bangladesh. The early detection of diseases in crops helps in the prevention of losses, thus improving productivity in agriculture. In the research work, different models were employed, including CNN- based models and Vision Transformers (ViT), as well as traditional machine learning approaches like SVM. The use of transfer learning helps in improving the performance of the models even with fewer images. In the experimental results, it has been demonstrated that ResNet50 has the highest performance among all the models, including CNN-based models and Vision Transformers, to detect rice leaf diseases[6]. A (Salman et al.,2025) study introduced a hybrid deep learning model. This model combined a Vision Transformer (ViT) and a Mixture of Experts (MoE) to improve the classification of plant diseases in real- world situations. Normally, deep learning models trained on a dataset cannot perform well under real-world conditions due to changes in illumination conditions, backgrounds, and severity of diseases. The suggested model used a Vision Transformer model to obtain global features from images of plant leaves. The model was then tested using the PlantVillage dataset to obtain results on the PlantDoc dataset to check its performance in real-world conditions. The experimental results obtained from this model were up to 20% more accurate than a normal Vision Transformer model, with a high accuracy of about 68%[7].

In (Hemalatha and Jayachandran 2024), a new model named PDLC-ViT, based on a Vision Transformer model, was presented for automatic plant disease detection. The model is based on a multitask learning approach to simultaneously carry out plant disease classification and localization. The model utilizes advanced mechanisms such as co-scale, co-attention, and cross-attention to identify crucial features from images of plant leaves. The model was trained using the PlantVillage dataset with optimized parameters to avoid overfitting using an early stopping mechanism. The experimental results indicate that the model achieved high performance with an accuracy of 99.97%, 99.18% mean Average Precision (mAP), and 99.11% mean Average Recall (mAR). The above study proves that the PDLC-ViT model is a highly efficient and dependable solution for automatic plant disease detection in modern agriculture[8] .

For the year (Borhani et al.,2022), a technique using deep learning has been proposed to enhance the precision of the detection of diseases, especially for tomato plants. Conventional ML techniques, such as SVM, k-NN, and Decision Tree, cannot effectively detect the disease region. Although the CNN-based methods have achieved improved results over the conventional ML methods by virtue of their improved feature extraction ability, these methods might not be effective enough to capture the global features of the objects. To overcome this drawback, the authors have proposed the use of Vision Transformers, along with the CNN-based approaches. In the proposed E- TomatoDet approach, the transformer-based global feature learning is enhanced with the improved local feature extraction techniques. In the experiments, the proposed approach has shown promising results on the COCO, PlantDoc, and CCMT datasets with a high mAP50 of 97.2%, compared to the performance of existing approaches, including YOLOv8, YOLOv9, and YOLOv10 [9]. In (Sun et al.,2025), the advanced deep learning network, termed E-TomatoDet, was proposed to address the issue of detecting diseases from the leaves of the tomato crops growing in complicated conditions. In the leaves of the tomato crops, small spots of disease occur. It is difficult to use deep learning-based techniques to address this issue. Therefore, the advanced network architecture based on the concept of global and local feature learning is proposed. In this network, the CSWin Transformer network is used to conduct the global contextual feature learning. In addition, the Comprehensive Multi-Kernel Module (CMKM) and the Local Feature Enhance Pyramid (LFEP) modules are proposed to conduct the local multi-scale feature learning from the leaves of the tomato crops. In the proposed network, the performance is validated on the tomato leaf disease detection dataset. In the validation, the performance is found to be 97.2% mAP50, which is better than the performance of the baseline network by 4.7%. In addition, the performance is better compared to the YOLOv10s network.[10]

III. PROBLEM STATEMENT

Manual verification of diseases is a slow process that is often inaccurate. This has created the need for an intelligent system that uses CNN EfficientNet-B0 and ViT to ensure high accuracy in the classification of crop diseases, but the existing systems lack the complete remedy for farmers, such as fertilizer advice, dosage advice, and voice guidance in multiple languages.

IV. LIMITATIONS OF EXISTING SYSTEM

In manual crop disease identification, it is completely dependent upon the farmer's experience and observation. It is a slow and inaccurate process. In rural areas, it is very hard to get expert advice regarding crop diseases, which leads to late identification of diseases. It may result in the use of improper fertilizers or improper amounts of fertilizers, which reduces crop yield. Although some digital agriculture technologies are available, they are limited to providing general textual information and do not support real-time image-based crop disease identification. Moreover, most of these available digital agriculture technologies do not support basic features like recommending fertilizers or treatments after identifying diseases. Most of the applications also do not support regional language and voice assistance, which makes the application difficult for farmers who are not educated. Most traditional machine learning algorithms, such as Support Vector Machines and K-Nearest Neighbours, are based on manual feature extraction techniques. These techniques are less accurate in handling the complexities of plant disease. At the initial stages of disease development, the symptoms are usually hard to identify because of poor image quality, lighting conditions, and noise in the background. These symptoms are usually missed or misclassified. The traditional techniques are also not able to effectively combine local and global feature learning because no such system exists in the prior art that is able to combine Convolutional Neural Networks and Vision Transformers

V. DATASET DESCRIPTION

The dataset used in the current research is downloaded from the Kaggle platform. It is a publicly available dataset for machine learning and research purposes. The images used in the current research are downloaded from the popular plant disease datasets named PlantVillage and PlantDoc. This dataset contains images of plant leaves with healthy leaves and leaves affected by various plant diseases.

The images contain various symptoms of diseases such as spots on leaves, color changes, texture changes, and various patterns visible on the leaves. This helps the machine learning model classify plant diseases. This dataset contains various crops and diseases, which helps the machine learning model classify plant diseases effectively.

For the proper training and testing of the machine learning model, the dataset is divided into three sets. 70% of the images are used for training the machine learning model. This helps the machine learning system learn the patterns of diseases. 15% of the images are used for validation. This helps the machine learning system check the performance of the model during training. This also helps the machine learning system avoid overfitting. The remaining 15% of the images are used for testing the machine learning model.

The dataset consists of images of leaves from nine different crops: apple, bell pepper, cherry, corn or maize, grape, peach, potato, strawberry, and tomato. The dataset consists of both diseased and healthy leaves and covers 18 classes in total. This variety in the dataset will enable the model to learn different kinds of disease and help it detect crop disease more accurately.

VI.PROPOSED METHODOLOGY

The proposed system intends to identify the diseases that affect crops based on the images of their leaves by using a hybrid deep learning method that includes both EfficientNet-B0 and Vision Transformer algorithms. The process that the system will undergo includes data preprocessing, feature extraction, and disease classification. The EfficientNet-B0 algorithm will be used for feature extraction of key features in the images of leaves, and the Vision Transformer algorithm will also be used for better results in classification by analyzing relationships in the images. The process will enable accurate identification of diseases that affect crops and will also offer recommendations for fertilizers that can be used for the crops.

a) Data Preprocessing

Data preprocessing is an essential step before training the deep learning model. The images of the plant leaves, which are collected from the datasets of PlantVillage and PlantDoc, may differ in their sizes, lighting conditions, and orientations. Therefore, various techniques are applied for preprocessing the data to effectively train the deep learning model, which can improve the performance of the proposed system.

b) Image Resizing

The images of the plant leaves are resized to a specific size so that they can be efficiently processed by the deep learning model. Resizing the images of the leaves of plants is an important technique, and all the images are resized to a specific size, which is essential for training the neural networks. In the proposed system, all the images are resized to a specific size, i.e., 224 x 224, which is essential for training the model.

c) Normalization

Normalization is carried out in order to normalize the pixel values of the images. This is important in order to stabilize the training process and make the model train efficiently. This is because the values are in a similar range. Normalization also helps the model to converge and train faster.

d) Data Augmentation

There are various data augmentation techniques used in order to increase the diversity of the data and reduce the possibility of overfitting. This is achieved through the use of rotation, flipping, zooming, and shifting the images. This increases the diversity of the images while maintaining the original characteristics of the disease. This makes the model robust and efficient in predicting the disease using images.

i. Feature Extraction Using EfficientNet-B0

The first stage of the proposed system utilizes a Convolutional Neural Network (CNN), which is known as the EfficientNet-B0 model, for the feature extraction of the given images of the plant leaves. The reason for choosing the EfficientNet-B0 model for the proposed system is the high accuracy of the model in classifying the given images, and the model is computationally efficient. The compound scaling technique is used in the model. This compound scaling approach allows the model to learn from the images using fewer parameters compared to the traditional CNN model.

In this stage of the proposed system, the preprocessed images of the leaves are given as an input to the EfficientNet-B0 model. The convolutional layers of the model are used to extract the important features from the given images. These features include patterns such as the difference in colors, textures, spots, discoloring, etc., which are generally found on the leaves of the infected plants. By analyzing these features of the leaves, the model is able to differentiate between the healthy leaves and the leaves affected by different kinds of diseases.

In the process of training, the model is able to identify certain disease patterns contained in the leaf images. Through continued analysis of many images in the training process, the model is able to improve its capacity to identify the disease patterns with great accuracy. This process helps the system differentiate between different crop diseases. This stage's output includes the main visual features extracted from the leaf images. These visual features are then subjected to nalysis in the next stage of the system using the Vision Transformer model to improve the accuracy of disease classification.

ii. Disease Classification Using Vision Transformer

The second stage of the proposed system utilizes the Vision Transformer (ViT) model to improve the accuracy level of the crop disease classification. The Vision Transformer is different from the traditional Convolutional Neural Network as it focuses on the relationships between the different parts of the image.

a) Patch Embedding

To start with the image embedding process, the image is first divided into different patches, and then each patch is embedded into a numerical form that is known as the embedding vector. The embedding vectors represent the different parts of the image, and the positional information is also added to the embedding vectors so that the image can be understood by the system.

b) Self-Attention Mechanism

The embedded patches are then passed to a transformer encoder to which self-attention mechanism is applied. The self-attention mechanism aids in the analysis of how different components of the image are interconnected. Through this, it is able to concentrate on different areas of the leaf that are of utmost importance in showing signs of diseases.

c) Global Feature Learning

Unlike CNN models, which are centered on local feature detection, Vision Transformers are able to detect global features in images. Through this, it is able to comprehend different complexities of diseases that may be visible in different areas of the leaf. Therefore, by integrating EfficientNet-B0 for feature extraction and Vision Transformers for global feature detection, it is able to improve its accuracy in detecting diseases in crops.

Lastly, the model's output layer is responsible for determining the type of disease occurring in the plant leaf image. Through this, fertilizer recommendations are made to the farmer, depending on the type of disease occurring in the crops.

Fig 1. Demonstrates the System Architecture of crop disease detection system

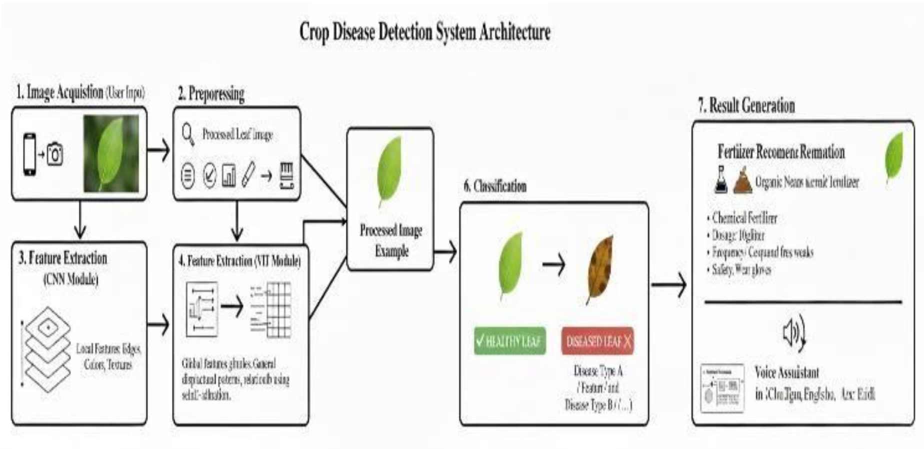

Fig 1. System Architecture

iii. Fertilizer Recommendation Module

The fertilizer recommendation module assists farmers in performing the right action after a crop disease is identified. Once the system detects the disease in the leaf image, it recommends fertilizers that can help in controlling or treating the disease. This module is created to provide farmers with information that is useful and easily understandable.

a) Fertilizer Names

The system offers farmers information about the names of the fertilizers that can be used to treat the identified crop disease. These fertilizers are usually chosen based on agricultural practices that are used to treat certain plant diseases.

b) Fertilizer Images

In addition to the name of the fertilizers, the system also shows images of the recommended fertilizers. The images will help farmers recognize the fertilizers easily when they are purchasing them from stores.

c) Treatment Guidance

The system will also provide farmers with treatment guidance on how to use the fertilizers. The guidance will provide farmers with the necessary information regarding the amount of fertilizers to be applied and how they should be applied in order to treat the affected crop.

iv. Multilingual Voice Assistant

The multilingual voice assistant has been included in the system, which will allow the crop disease detection application to be more accessible to the farmer. This is because the farmer might not be able to read the information on the screen, which will be technical in nature. This will enable the farmer to understand the information more clearly.

a) Voice Output

The system will include voice output, which will be used to display the crop disease detected and the fertilizer information. The voice assistant will also be used to explain the dosage, usage, and precautions, which will enable the farmer to understand how to use the information to cure the crop disease.

b) Multiple Language Support

To accommodate farmers of different geographical locations, the voice assistant has been designed to support multiple languages. In this system, the information has been provided in English, Hindi, and Telugu. This helps the farmers to access the information in the language they are most comfortable with.

c) Farmer-Friendly Interface

The voice assistant has been designed to have a user- friendly interface for the farmers. Farmers can use this system by simply uploading the image of the leaf, and the system will detect the disease for them. This helps the farmers, even if they are not tech-savvy.

VII.RESULTS AND DISCUSSION

The proposed crop disease detection system was trained using the images of the leaves obtained from the PlantVillage and PlantDoc datasets from the Kaggle platform, which were split into 70% for training, 15% for validating, and 15% for testing the data.

Fig. 2 shows the interface of the proposed crop disease detection system, in which the image of the crop's leaves is uploaded, the disease in the crop is detected, and the fertilizer required with the amount and precautions is suggested.

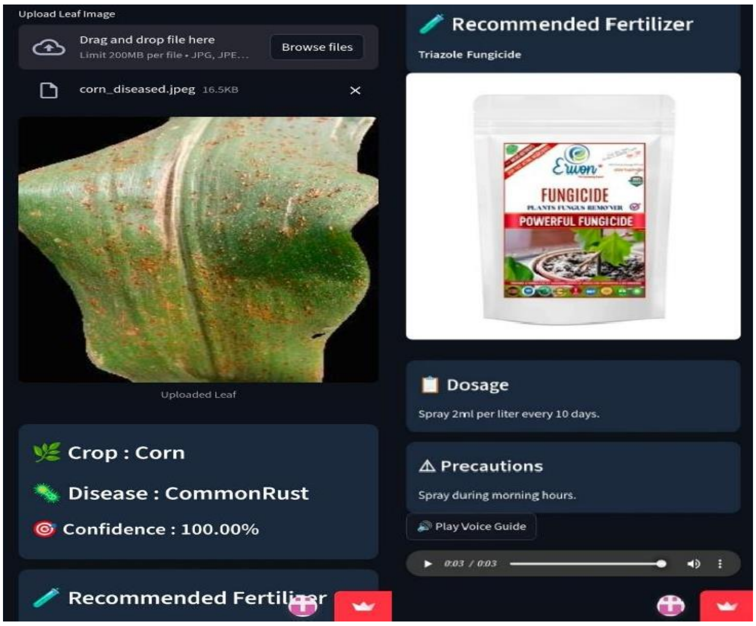

Fig 2. Crop disease identification & fertilizer recommendation

Fig. 3 shows the accuracy graph, in which the accuracy increases as the model is trained, reaching a maximum accuracy of nearly 100%.

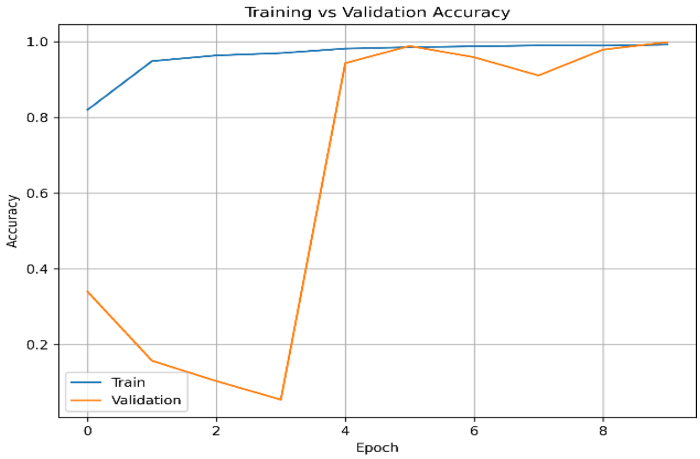

Fig 3. Accuracy graph

Fig. 4 shows the loss graph, in which the loss decreases as the model is trained, indicating that the proposed model learns effectively.

Fig 4. Loss graph

Fig. 5 and Fig. 6 show the precision and recall values, in which the values are nearly equal to 1, indicating that the proposed model works very effectively.

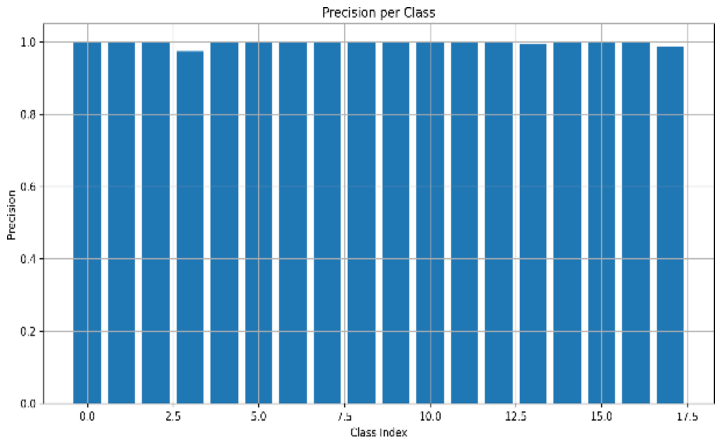

Fig 5. Precision graph

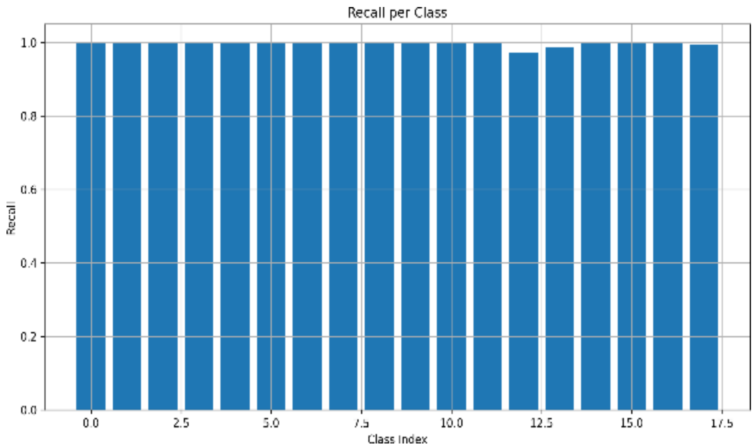

Fig 6. Recall graph

Fig. 7 shows the confusion matrix, in which the values are closer to the main diagonal, indicating that the proposed model classifies the images of the crop disease very effectively.

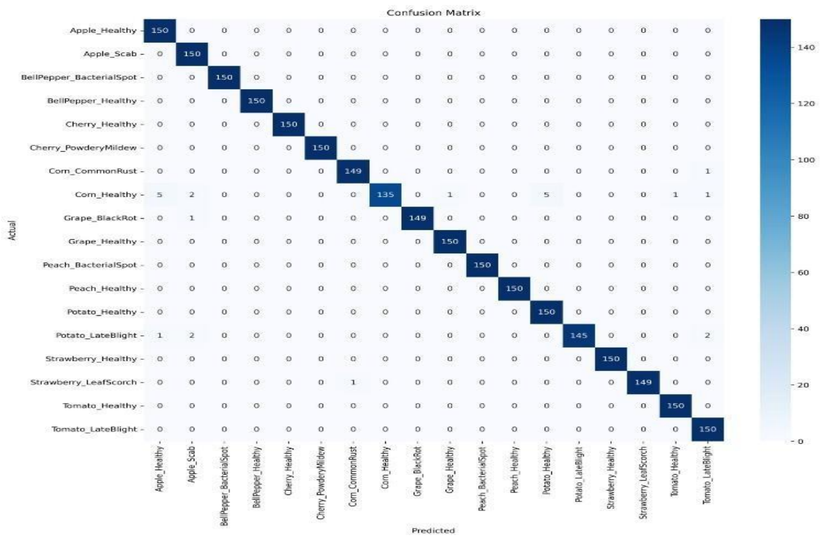

Fig 7. Confusion matrix

VIII.CONCLUSION

In this research, an intelligent system of crop disease detection has been proposed with the use of EfficientNet-B0 and Vision Transformer models. With the use of this system, accurate detection of crop diseases is carried out from images of plant leaves. For accurate detection of plant diseases, EfficientNet-B0 is utilized to extract essential visual features from plant leaf images. At the same time, Vision Transformer is utilized to extract accurate plant leaf images.

In this research, the dataset is collected from Kaggle with images from PlantVillage and PlantDoc datasets. For proper evaluation of the model and accurate results, 70% of the dataset is utilized for training, 15% is utilized for validation, and 15% is utilized for testing.

In addition to accurate detection of plant diseases with images of plant leaves, fertilizer suggestions with images of fertilizers, names of fertilizers, and usage of

fertilizers are provided to farmers. Furthermore, a multilingual voice assistant is included with this system to provide accurate plant disease detection and fertilizer suggestions to farmers in English, Hindi, and Telugu.

With the use of this system, farmers would be able to detect plant diseases at an early stage and use accurate fertilizers to improve agricultural productivity.

Acknowledgement

The authors are very much thankful to the management of Stanley College of Engineering and Technology for Women for providing the necessary facilities to conduct the experiments in the institution's laboratory. The authors also express their gratitude to the faculty members of the Department of Information Technology for their continuous support and encouragement throughout this work.

REFERENCES

1) Shandilya, G., S.Gupta, H. G.Mohamed, S.Bharany, A. U.Rehman, and S.Hussen. 2025. "Enhanced Maize Leaf Disease Detection and Classification Using an Integrated CNN-ViT Model." Food Science & Nutrition13, no.7:e70513. https://doi.org/10.1002/fsn3.70513.
2) Sinamenye, J.H., Chatterjee, A. & Shrestha, R. Potato plant disease detection: leveraging hybrid deep learning models. BMC Plant Biol 25, 647 (2025).https://doi.org/10.1186/s12870-025-06679-4
3) Aboelenin, S., Elbasheer, F.A., Eltoukhy, M.M. et al. A hybrid Framework for plant leaf disease detection and classification using convolutional neural networks and vision transformer. Complex Intell. Syst. 11, 142 (2025). https://doi.org/10.1007/s40747- 024-01764-x
4) Roy, P.S., Kukreja, V. Vision transformers for rice leaf disease detection and severity estimation: a precision agriculture approach. J. Saudi Soc. Agric.Sci. 24,3(2025).https://doi.org/10.1007/s44447- 025-00007-w
5) Maeda-Gutiérrez, V., Oropeza-Valdez, J. J., Reveles-Gómez, L. C., Padron-Manrique, C., Resendis-Antonio, O., Solís-Sánchez, L. O., Guerrero-Osuna, H. A., & Olvera Olvera, C. A. (2025). AI-Driven Plant Health Assessment: A Comparative Analysis of Inception V3, ResNet-50 and ViT with SHAP for Accurate Disease Identification in Taro.Agronomy,15(1),77.https://doi.org/10.3390/agro nomy15010077
6) Mehnaz, S., & Islam, M. T. (2025). Rice leaf disease detection: A comparative study between CNN, transformer and non-neural network architectures (arXiv:2501.06740).
7) Salman Z, Muhammad A and Han D (2025) Plant disease classification in the wild using vision transformers and mixture of experts. Front. Plant Sci. 16:1522985. doi: 10.3389/fpls.2025.1522985
8) Hemalatha, S., Jayachandran, J.J.B. A Multitask Learning-Based Vision Transformer for Plant Disease Localization and Classification. Int J Comput Intell Syst 17, 188 (2024). https://doi.org/10.1007/s44196-024-00597-3
9) Borhani, Y., Khoramdel, J. & Najafi, E. A deep learning based approach for automated plant disease classification using vision transformer. Sci Rep 12,11554(2022).https://doi.org/10.1038/s41598- 022-15163-0
10) Sun H, Fu R, Wang X, Wu Y, Al-Absi MA, Cheng Z, Chen Q, Sun Y. Efficient deep learning-based tomato leaf disease detection through global and local feature fusion. BMC Plant Biol. 2025 Mar 11;25(1):311. doi: 10.1186/s12870-025-06247-w. PMID: 40069604; PMCID: PMC11895386.