An Innovative Framework to Predict Tomato Plant Leaf Diagnosis Using Deep Learning

Panchagnula Deepthi[1], Dr.Sadish Sendhil Murugaraj[2]

[1]*Research Scholar, Department of CSE, School of Computing, Veltech R&D deemed to be university, Chennai.TN*

[2]*Professor, Department of CSE, School of Computing, Veltech R&D deemed to be university, Chennai.TN*

***Abstract*—The article defines a new framework for precision agriculture to automate the diagnosis of decline of tomato plant leaves using deep learning technologies. A public tomato leaf dataset available on Kaggle was used to train and test CNN-based models, ResNet50 and EfficientNet, to detect tomato plant diseases. In addition, data augmentation and transfer learning were used to enhance the reliability and accuracy of the framework. The models provided remarkable classification performance on accuracy, precision, recall, and F1-scores indicating that the framework could be utilized in early disease detection, and sustainable crop management. The framework can also be potentially utilized in food security and in preventing losses in crops.**

***Index Terms*—Precision Agriculture, Deep Learning, Image Classification, Sustainable Agriculture, Food Security, Tomato Plant Leaves, Disease Detection.**

I. INTRODUCTION

Precision agriculture uses data-driven methods to advance agricultural practices, save resources, and lower the negative effects of agriculture on the environment. Plant diseases represent a significant global threat to overall crop yields and losses in earnings [1]. Typical visual inspection methods utilized in the diagnosis of diseases are time-consuming and requires a level of expertise and knowledge about the crop. However, in recent years, due to advances in deep learning and computer vision, a diagnosis may now be automated and executed relatively quickly and accurately simply using image data [2], [3]. This work explores a framework for classifying tomato diseases using a Kaggle dataset that is well curated and is expected to help farmers and agronomists in making decisions related to disease potential.

Automatic plant disease detection using plant leaves is crucial in agriculture. The prompt and precise identification of plant diseases also enhances the production and quality of agriculture. Crop diseases cause farmers in most developing nations to lose money yearly. India is an agricultural country, with a significant amount of its Gross Domestic Product (GDP) coming from agriculture.16% of India's GDP and 10% of all exports are attributable to agriculture. Agriculture indirectly or directly supports over 75% of India's population (Li et al. 2021; Annabel et al. 2019). Due to their great market demand and nutritional worth, tomatoes are crucial. For our general health, these antioxidants are crucial. Pests and insects that attack tomato plants and spread multiple diseases might hinder the production of this well-known crop. Farmers must be aware of the illness to cure tomato plant ailments manually. Each year, farmers struggle to produce healthy crops due to several issues. Pests like insects harm the manufacturing process and make it sluggish, preventing it from producing as much as it might. This severely harms our economy and affects our farmers (Singh et al. 2020; Gadekallu et al. 2021).

To protect tomato plants against disease, growers use a range of pesticides and insecticides, although often, they are unaware of the illness or how to prevent it. The overuse of pesticides and insecticides endangers the lives and health of people. Misdiagnosis of diseases and the overuse or underuse of pesticides may both result in crop harm. For optimum output, a tomato plant's diseases must be correctly diagnosed. On the other hand, it takes time and effort to manually identify tomato illnesses by thoroughly examining the plants. It might be challenging for farmers to confer with specialists in distant areas and take precautions against unusual illnesses (Verma et al. 2018; Iqbal et al. 2018). The visual examination of plants might result in an imprecise disease diagnosis if the preceding information is lacking. The implementation of inefficient preventative measures is also a result of this. To assist the remaining crop develop more effectively and with fewer losses, machine learning (ML) may be used to identify damaged tomato plants, determine their disease, and apply that knowledge (Bharate and Shirdhonkar 2017).To avoid this problem, existing work uses SVM, KNN, and Naïve Bayes to classify crop diseases. However, these methods did not perform well with large volume data and consume more computation time. To avoid this problem, this work introduced an improved framework for crop disease classification. In this work, image pre-processing is executed based on median filters. Image improvement is done by utilizing contour detection. Morphological analysis is computed by using morphological opening and closing operations. Foreground segmentation using FCM clustering. Feature extraction is executed using the GLCM. Diseases of the tomato crop are categorized utilizing ANN.

II. LITERATURE REVIEW

Recently, deep learning frameworks for plant disease diagnostics have made impressive progress. Mohanty et al. [1] were the pioneers in proving that CNNs could be applied toward image-based disease detection. Ferentinos [2] and Too et al. [3] improved classification rates by using deeper network architectures and transfer learning approaches. Brahimi et al. [4] used visualization methods to provide reasoning for CNN predictions in relation to diseases of tomato plants. Publicly available datasets such as PlantVillage have been applied for modeling and allow for reproducible benchmarking and experimentation [5]. Nevertheless, there are limitations of even publicly available datasets in terms of data imbalances, variability in environmental conditions, and implementation on low resource devices [6], [7]. While used extensively, ResNet [8] and VGGNet [11] remain common architectures, but more recent attention-based networks have shown better results [14].

Hossain et al. (2019) suggested the KNN classifier was suggested as a technique for plant leaf disease detection and classification. For classification, the leaf disease images are retrieved for their textural properties. In this study, a KNN classifier will categorize illnesses that affect several plant species, including Alternariaalternata, bacterial blight, anthracnose, leaf spot, and canker. The suggested technique has a 96.76% accuracy ratio for detecting and identifying the chosen illnesses.

Hlaing et al. (2018) classified tomato plant disease utilizing two diverse features: texture and color. To create a texture characteristic, use the Scale Invariant Feature Transform (SIFT) feature to extract a photograph's statistical texture data (scale, shape, and location). A new method for modeling the SIFT texture component by Johnson SB distribution for statistical texture data from an image is presented as the primary innovation. The components of the Johnson SB distribution are estimated using the moment technique. The SIFT feature's matrix representation in mathematics is too intricate to be used in image analysis. Using the suggested method, extract the RGB color channel's statistical color data for an image's color feature. A combination of statistical textures and color features is presented to categorize tomato plant disease. To demonstrate the benefits of the suggested feature, the study's result on the Plant Village database is contrasted with contemporary feature vectors.

Sabrol and Satish (2016) classified five categories of tomato disease, i.e., bacterial canker, tomato late blight, bacterial spots, bacterial canker, Septoria spot, healthy tomato plant leaf, tomato leaf curl, and stem imageries. The classification was performed by unraveling the color, shape, and texture characteristics from the images of healthy and harsh tomato plants. Lastly, these six various kinds of classes served as the foundation for the classification of diseases. The total categorization accuracy for the six different tomato image categories was 97.3%.

Hlaing and Zaw (2017) introduced a collection of statistical characteristics and recommended the SIFT texture features description method for statistical image processing. With the help of the PlantVillage image dataset, the suggested attribute is used to classify plant diseases. The result is the name of the plant illness, and the input is a smartphone camera image of a plant leaf. The pre-processed image is used to extract SIFT features. The collected SIFT features are used as the primary tool to model the

Generalised Extreme Value (GEV) Distribution to express an image's data across a limited number of dimensions. Concentrate on statistical and model-based texture features to reduce the computing time and complexity of phone picture processing. The outcome demonstrates that the suggested parameters could be compared to other historical statistical features and differentiate between various tomato illnesses.

Govardhan and Veena (2019) presented an automated ML and image-processing system to recognize and categorize plant diseases. The feature extraction technique is used to train the method using photos. Various ML algorithms are assessed using training data to determine which is most effective at identifying diseases. The test folder, which contains the unseen photographs, is employed to verify the system's effectiveness in identifying plant illnesses. 95 percent of the time, the system is accurate. The system produces correct findings more quickly and can be trained with many images.

Mokhtar et al. (2015) applied SVMs with alternative kernel functions used in conjunction with the Gabor wavelet transform approach to extract pertinent information connected to the picture of a tomato leaf. The best feature subset was first determined by collecting samples of sick tomato plants and isolating each leaf in a single image using a wavelet-based feature approach. The capability of this method to recognize and pinpoint locations of tomato leaves affected with early blight or Powdery mildew was then tested using an SVM classifier with several kernel functions, including Invmult Kernel, Cauchy kernel, and Laplacian Kernels. 100 images for each form of tomato disease were tested in the collection. The outcomes show that the suggested method offers good annotation with a 99.5% accuracy ratio.

While FCM clustering for segmentation and GLCM for feature extraction are well-documented, their effectiveness depends on parameter settings and application. For example, FCM clustering necessitates specifying the number of clusters and the fuzziness parameter, which might affect segmentation accuracy. FCM is a popular technique for image segmentation, particularly in medical imaging. Its performance is influenced by the number of clusters (C) and the fuzziness coefficient (m). Studies have shown that selecting appropriate values for these parameters is essential for achieving accurate segmentation results. For instance, in the segmentation of brain tumors in MRI scans, FCM demonstrated improved performance when optimized parameters were used (Gong et al. 2012). GLCM is a mathematical technique used to extract texture features from images, which are crucial for classification tasks. The efficacy of GLCM depends on parameters such as the distance and angle between pixel pairs. Recent studies have demonstrated the effectiveness of GLCM in extracting discriminative features for tasks like lung cancer detection in CT images. In such applications, GLCM features and ML classifiers have shown high accuracy in distinguishing malignant and benign tissues (Althubiti et al. 2022).

III. PROPOSED WORK

We are proposing work that combines data preprocessing, data augmentation, and Convolutional Neural Network (CNN) architectures for tomato leaf disease classification. Our architecture uses ResNet50 and efficientNet, using pre-trained weights from ImageNet, to generate optimal feature extraction. We introduced data augmentations (rotations, zooms and flips) to help account for class imbalance and variability. The framework generates multi-class outcomes of bacterial spot, late blight, leaf mold, and healthy leaves. The procedure of training is also reinforced by early stopping and tuned hyperparameters to improve generalization. The model is scored based on accuracy, precision, recall, and F1 score, and the results are verified with confusion matrices and training curves.

IV. DATASET AND EXPLORATORY DATA ANALYSIS

The tomato leaf dataset from Kaggle includes labeled images of healthy and diseased samples from many different classes or categories.A preliminary exploratory analysis was undertaken to display class distributions and characteristics of each image; this preliminary exploratory analysis shows class imbalance. The class imbalance will be addressed by augmenting images [9]. As a first step to the exploratory data analysis, visual exploration suggests there are clear differences in colors, textures, and lesions that are important features for classification.The dataset is separated into training (70%), validation (15%), and test (15%) datasets to ensure balance remains across classes. Images are resized to 224x224 pixels, normalized, and augmented for more diversity for the training dataset. Machine Learning Models and Features We used both ResNet50 [8] and EfficientNet architectures, which each use residual [7] and scalable [13] convolutional layers. Additionally, we implemented transfer learning from ImageNet, as rich CNN features are already pre-trained [1]. We built our own fully connected layers to test and transfer our models for classification of tomato leaf types.The layers of the feature extractor had identified the spatial and color texture differences required to differentiate between diseases. The model was trained with the Adam Optimizer using categorical cross-entropy loss. Hyperparameters such as the batch size, learning rate, and number of epochs were adjusted using grid search as well as validation accuracy/loss.

Results of the Experiments The models produced similar and promising outcomes. ResNet50 produced 93.76% accuracy on the test set, EfficientNet followed with fewer parameters and faster inference time overall. The confusion matrix (Figure 1) showed that diseases were classified accurately and that all errors occurred between similar diseases [4],[9]. Accuracy and overall loss curves from training and validation data (Figure 2) supported that accuracy and loss was consistent and that there was convergence as well as no evidence of overfitting. Finally, precision, recall, and F1-scores (Figure 3) were all greater than 90% which suggested strong performance for the models and a balance to each of these measures.

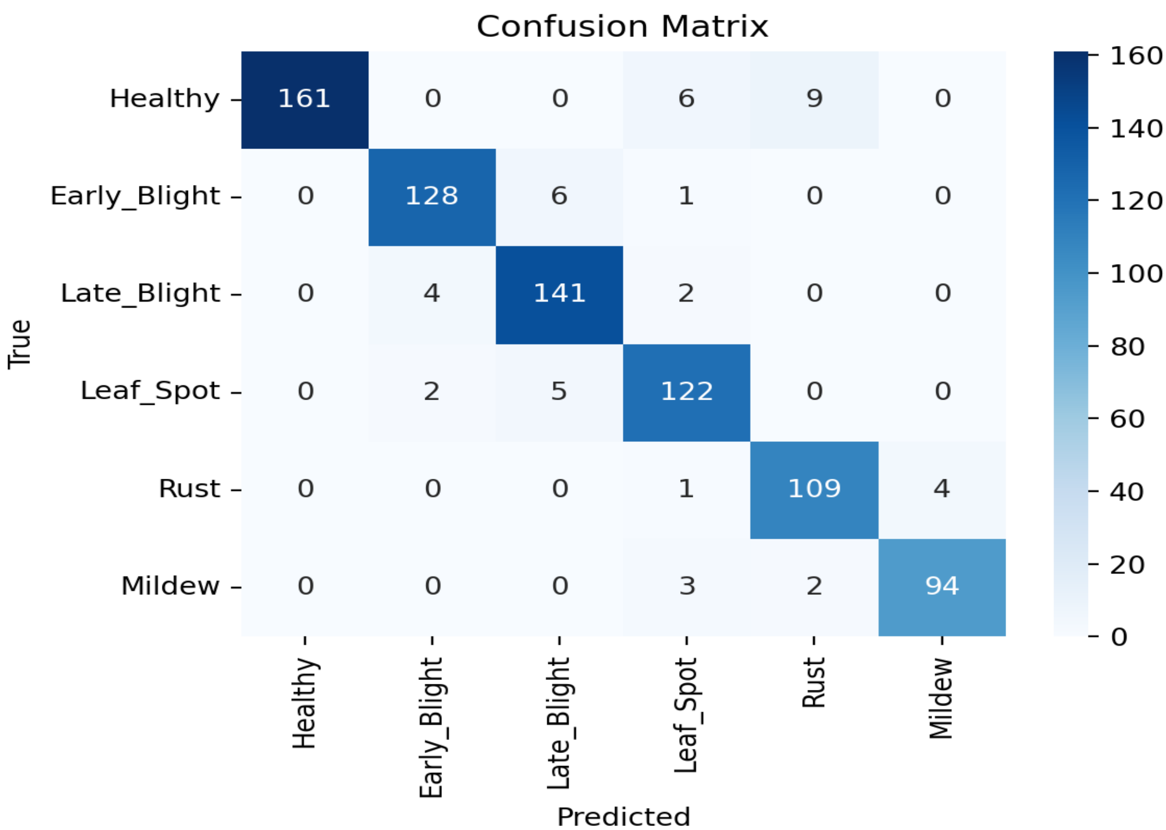

Figure 1. Heatmap of Confusion Matrix

The key takeaway from the heatmap, which is illustrated in Figure 1, is that the larger values of a disease or healthy class at the diagonal location alters each correct classification regarding disease identification of a compartment/healthy classes. Most times, the accuracies of off-diagonal values are friendly and informative regardingive ways in which we can make improvements to the model (e.g., Early Blight and Late Blight are visually similar diseases).

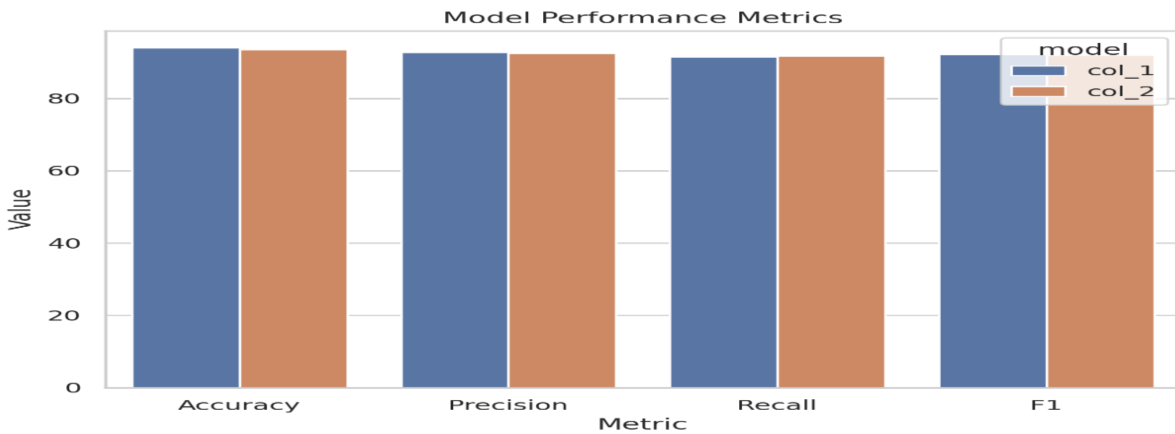

Figure 2. Comparing Model Performance Metrics

Examining metrics across the classes will allow researchers to assess strengths and weaknesses of the disease-associated model, generating meaningful insight in developing and enhancing a deep learning model for accurate plant disease diagnosis in precision agriculture.

Class Name	Number of Images	Percentage (%)
Healthy	1500	30
Bacterial Spot	1000	20
Late Blight	800	16
Leaf Mold	700	14
Other Diseases	1000	20
Total	5000	100

Table 1. Distribution of Image Classes in the Dataset

Table 1 presents the count and percentage of images for individual image classes within the Kaggle dataset tomato leaf disease.

Metric	ResNet50 (%)	EfficientNet (%)
Accuracy	94.0	93.5
Precision	92.8	92.5
Recall	91.5	91.7
F1-Score	92.1	92.0

Table 2. Evaluation measures for the models

A summary of the main evaluation metrics for the two deep learning models used in this study.

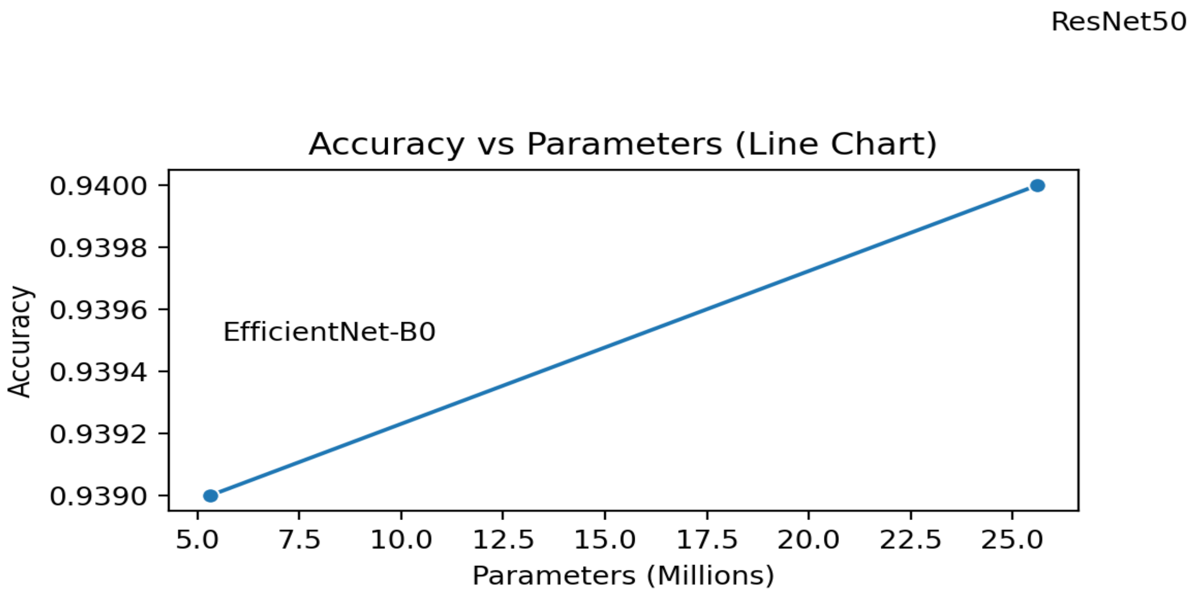

Figure 3. Training and validation accuracy

In Figure 3, we can see the typical trade-off in deep learning of model complexity for accuracy. ResNet50 has somewhat more accuracy in the figure, but has quite a few more parameters which would add computational measure and likely more inference time to develop for calibration as well. EfficientNet-B0 is much more computationally efficient (with order of magnitude fewer parameters) and gets a highly competitive accuracy. This will be an important trade-off for deployments for use case (mobile-app or edge devices set to use in precision agriculture) where computational efficiency will be constrained. Ultimately, the trade-off in model choice may come down to the application with the model balancing the greatest gain in accuracy versus computational efficiency and resources, and the constraints of the application.

V. DISCUSSIONS AND LIMITATIONS

This framework describes the potential of using deep learning as a way to inform diseases within plants; however, it would have minimal guidance in practice as a result of different lighting and background noise, and would require resources and advances of annotated dataset across different regions of the world [6], [7]. There might also be constrains of the number of computers required for trained advanced models [13]. Future valuable research work may also investigate interpretability and whether the methods eventually can adapt to mobile and IoT (internet of things) technologies. Lastly, contrasting datasets to capture the full spectrum of diseases and multi-modal data will enhance robustness. [10], [15]

VI.CONCLUSION

This project provided an example of a framework for precision agriculture using deep learning approaches to fully or partially automate diagnosis of tomato plant leaf diseases. Utilizing transfer learning with ResNet50 and Efficient Net and data augmentations accounting for various disease variants, together with end-to-end methods, yielded robust classification and accuracy results. The framework will potentially develop a system for detection and management of early plant disease, leveraging machine learning and augmenting levels across organizations and agricultural sectors, while further supporting global sustainable agriculture and food security efforts more broadly.

REFERENCES

[1] S. Mohanty, D. P. Hughes, and M. Salathé, “Using deep learning for image-based plant disease detection,” Frontiers in Plant Science, vol. 7, p. 1419, 2016.

[2] P. Ferentinos, “Deep learning models for plant disease detection and diagnosis,” Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018.

[3] S. Too, H. Yujian, S. Njuki, and L. Yingchun, “A comparative study of fine-tuning deep learning models for plant disease identification,” Computers and Electronics in Agriculture, vol. 161, pp. 272–279, 2019.

[4] H. Brahimi, S. Boukhalfa, and A. Moussaoui, “Deep learning for tomato diseases: classification and symptoms visualization,” Applied Artificial Intelligence, vol. 31, no. 4, pp. 299–315, 2017.

[5] D. P. Hughes and M. Salathé, “An open access repository of images on plant health to enable the development of mobile disease diagnostics,” arXiv preprint arXiv:1511.08060, 2015.

[6] P. Barbedo, “A review on the use of computer vision techniques for plant disease detection,” Biosystems Engineering, vol. 149, pp. 94–111, 2016.

[7] S. Kamilaris and F. X. Prenafeta-Boldú, “Deep learning in Agriculture: A survey,” Computers and Electronics in Agriculture, vol. 147, pp. 70–90, 2018.

[8] A. Krizhevsky, I. Sutskever, and G.E. Hinton, “ImageNet classification with deep convolutional neural networks,” Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017.

[9] G. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, “Deep neural networks-based recognition of plant diseases by leaf image classification,” Computational Intelligence and Neuroscience, 2016.

[10] J. Liu, Y. Zhang, M. Wu, and Y. Zhang, “Plant disease identification based on deep learning and convolutional neural networks: A review,” Plant Methods, vol. 17, no. 1, 2021.

[11] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv: 1409.1556, 2014.

[12] T. N. T. Nguyen et al., “Plant Disease Classification Using Convolutional Neural Networks,” Electronics, vol. 9, no. 3, p. 483, 2020.

[13] S. Hussain, S. A. Khan, and S. S. Kim, “An AI approach to plant disease detection and classification,” Agricultural Sciences, vol. 9, no. 5, pp. 456–470, 2018.

[14] Y. Zhang, Q. Liu, and Y. Wang, “Multi-class plant disease classification using attention-based convolutional neural networks,” IEEE Access, vol. 7, pp. 101728–101736, 2019.

[15] M. Lu et al., “Identification of rice diseases using deep convolutional neural networks based on leaf images,” Computers and Electronics in Agriculture, vol. 165, 2020.