A Comparative Study of Fine-Tuning Deep Learning Models for Leaf Disease Identification and Classification

Bh. Prashanthi
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, Vaddeswaram, India | Department of Computer Science and Engineering, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India
bhupathi.prashanthi@gmail.com (corresponding author)

Anne Venkata Praveen Krishna
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, Vaddeswaram, India
praveenkrishna@kluniversity.in

Ch. Mallikarjuna Rao
Department of Computer Science and Engineering, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India
professorcmrao@gmail.com

Received: 16 September 2024 | Revised: 12 October 2024, 31 October 2024, and 08 November 2024 | Accepted: 26 November 2024

 DOI: https://doi.org/10.48084/etasr.9017

ABSTRACT

Innovative agricultural solutions are needed to detect and classify leaf diseases early across crop species and environments. This study compares deep learning approaches, focusing on Convolutional Neural Networks (CNN) and Vision Transformers (VTs), to identify leaf diseases early and accurately for scalable crop management and productivity. Optimizing CNNs, Explainable Transfer Learning (ETPLDNet) using ResNet50 architecture, and LEViT leaf disease diagnosis are compared. The CNN model, optimized with dynamic hyperparameters, achieved an impressive 99.58% accuracy for leaf disease classification, demonstrating its effectiveness in feature extraction and classification precision. On the other hand, the VT-based LEViT model, which leverages self-attention mechanisms and Explainable AI (XAI), achieved 95.22% accuracy but offers enhanced interpretability and generalization capabilities due to its transformer-based architecture. This distinction illustrates that while CNNs excel in accuracy, VTs provide a more transparent decision-making process and better handle the complex variances in plant leaf diseases, making them ideal for precision agriculture. The combined use of CNNs and VTs showcases the strengths of each model, with CNN focusing on high classification precision and VTs offering improved interpretability and adaptability for various leaf disease conditions. The use of XAI enables the models to highlight important areas in plant leaf images that influence the model's decisions, offering a transparent and interpretable decision-making process that allows researchers and farmers to understand why a particular diagnosis or classification was made. This ability to visualize and explain the reasoning behind the model predictions is crucial to increasing trust in AI-driven solutions in agriculture. By combining the high precision of CNN and the interpretability of VT with XAI, this study offers a robust approach to improving crop disease management and precision agriculture.

Keywords-leaf diseases; convolutional neural networks; transfer learning; vision transformers; explainable artificial intelligence; classification

I. INTRODUCTION

Agriculture drives GDP, making crop damage a major issue for national productivity and food security [1]. Leaves are the most sensitive part of plants and often show early signs of disease, highlighting the need for crop management throughout its lifecycle. Large land expanses, infectious diseases, multiple diseases on one leaf, and a lack of agricultural competence in remote areas complicate the diagnosis and management of agricultural diseases. These obstacles hinder disease prevention and treatment, reducing production and quality. Global food security depends on the diagnosis of infectious plant diseases [2]. New deep learning methods, such as Vision Transformer (VT) models, improve leaf disease detection and classification. This study improves the accuracy of leaf disease classification using VTs and transfer learning for agricultural production and food security. Machine learning, particularly deep learning, can improve crop management and meet food demand [3], with applications in insect and crop management, fruit picking, and leaf recognition [4]. Transfer learning helps to perform tasks with limited data, especially in rare data domains, with large image datasets. Plant diseases reduce global grain production, emphasizing the importance of disease detection in agriculture.

Leaf diseases, caused by fungal, bacterial, viral, and environmental factors, can damage crop productivity and quality [5]. Symptoms include discoloration, spotting, wilting, deformation, and defoliation. Detection and classification using visual inspection, laboratory analysis, and remote sensing, along with machine learning and computer vision, can help reduce disease costs. Many studies used deep learning models to detect and classify leaf diseases. Deep learning methods can detect plant anomalies, such as parasites or diseases, despite weak contrast and minor changes, reducing agricultural losses and improving precision agriculture [6]. CNNs are recommended for early plant disease detection and yield prediction. Deep CNN architectures extract images and classify leaf diseases using complex convolutional layers, instead of metadata analysis and hand-crafted features [7]. Computer vision applications have made deep transfer learning popular for leaf disease diagnosis, improving automated diagnostic systems, increasing crop yield and reducing species losses. CNNs reduce agricultural losses and improve management by detecting plant parasites and diseases [8]. Pre-trained models, such as VGG16 and ResNet50, can recognize plant and leaf diseases. These transfer learning-refined models can accurately diagnose and classify leaf diseases, improving agriculture and crop health [9]. Deep learning architectures trained on curated plant disease datasets improve disease diagnosis and classification along with farming outcomes.

Many studies on plant disease detection used high-quality datasets, such as aerial and non-aerial images. Effective CNN models, including DenseNet, achieved high accuracy on the PlantVillage dataset [10]. CNNs improve plant disease detection but increase processing requirements. Vision Transformers (ViTs) can address CNN limitations with self-attention in identifying leaf diseases [11]. ViTs and CNNs can identify plant diseases [12] but require larger models and more processing resources. CvT [13] and PlantViT [14] hybrid models improve plant disease diagnosis. In [15], the focus was on apple leaf disease detection using computer vision techniques.

The primary motivation for using deep learning models for leaf disease detection is based on their ability to provide a highly accurate and efficient diagnosis, which is essential for early intervention. These models, such as CNNs and ViTs, can handle large datasets and complex variations in disease symptoms, making them scalable and adaptable across different crops and environments. By automating the detection process, deep learning models reduce the need for manual inspections, reducing operational costs while ensuring consistent monitoring. Additionally, the integration of Explainable AI (XAI) techniques, such as Grad-CAM, enhances transparency, allowing users to understand the reasoning behind model predictions, which fosters trust and improves decision-making. Ultimately, deep learning models help boost agricultural productivity by preventing the spread of diseases, ensuring better crop management, and contributing to food security.

This study uses deep learning models to classify leaf diseases, reducing detection costs and increasing crop yield on large farms. The selection of specific deep learning models, such as CNN and ViTs, is driven by their ability to address key gaps and meet real-world agricultural needs. CNNs are chosen for their proven effectiveness in image-based recognition, particularly their ability to automatically extract features from leaf images, handle complex patterns, and provide high accuracy for leaf disease detection. CNNs' hierarchical structure, with convolutional and pooling layers, enables them to capture spatial dependencies and subtle disease features in large agricultural datasets, making them ideal for early and accurate disease diagnosis. However, CNNs can sometimes lack transparency in decision-making, leading to the integration of ViTs that offer improved interpretability through self-attention mechanisms and XAI techniques such as Grad-CAM. ViTs also excel in processing complex variations in leaf diseases and are more robust in handling diverse environmental conditions, making them particularly useful for scalable and explainable solutions in agriculture. These models bridge the gap between the high-accuracy detection of CNNs and the need for transparency and better generalization in varying agricultural environments through ViTs, ensuring more reliable and interpretable disease management practices.

The main challenges in applying deep learning models to diverse plant leaf images include variations in disease symptoms, differences in leaf shapes and textures, environmental factors, such as lighting and background noise, and the limited availability of labeled datasets. These challenges are addressed through data augmentation techniques (e.g., random flipping, rotation, zooming, resizing) to enhance model generalization and preprocessing steps to isolate diseased leaves for improved feature extraction. Additionally, dynamic hyperparameter tuning in CNN models allows for better adaptation to complex data, while ViTs provide enhanced interpretability and robustness in handling environmental variability. Transfer learning with models such as ResNet50 can further mitigate the issue of limited labeled data by leveraging pre-trained knowledge, enabling accurate and scalable disease detection across different plant species. These strategies ensure more effective and reliable performance in leaf disease detection.

This study uses an optimized CNN model for leaf disease classification. Dynamic hyperparameters, such as learning rate and regularization, improve model accuracy and data adaptability. ETPLDNet, an explainable transfer learning model based on ResNet50 and modified with dropout techniques, dense layers for accuracy enhancement, and Grad-

CAM for interpretability, outperformed existing models in the domain with 92.53% training, 97.33% validation, and 97.58% testing accuracy across 38 disease classes and 14 plant types. The LEViT model [16] clarified decisions with the ViT_B_32 transformer with 95.22% training, 96.19% validation, and 92.33% test accuracy. The proposed method was compared with modern leaf disease diagnosis and categorization models, showing its superiority. These models demonstrate that dynamic hyperparameter tuning, explainability in deep learning, and transformer-based architectures improve leaf disease identification and classification, promising agricultural advances.

II. METHODS

Deep learning, transfer learning, and ViTs are extensively studied for plant leaf recognition but can provide misdiagnoses due to leaf, type, and environmental variations that are particularly challenging in rural areas. CNNs are effective for image-based recognition but face challenges due to limited-sized and diverse datasets. Figure 1 illustrates the proposed method for evaluating deep learning models in leaf disease diagnosis and classification.

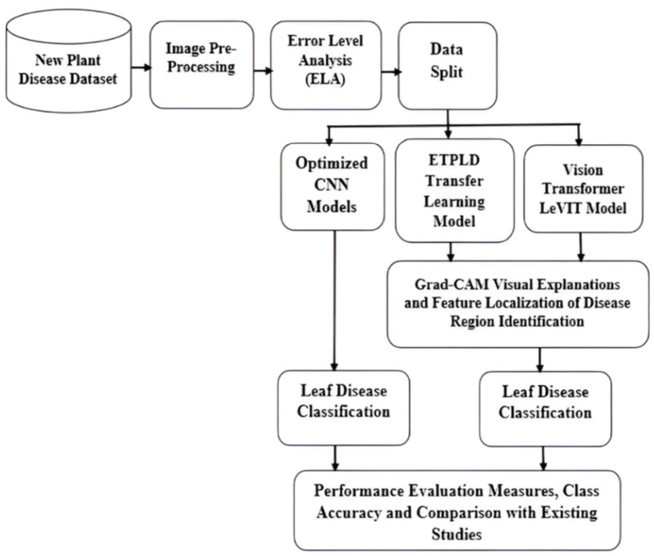

Fig. 1. Proposed method.

A. Preprocessing and Augmentation

The preprocessing of images from the New Plant Disease Dataset [17] enhances feature extraction and classification precision. To avoid overfitting and ensure reliable results, the CNN model requires multiple iterations and a large image dataset. Separating diseased leaves from background information in images before feeding them to deep learning models improves the identification and classification of leaf diseases. The sequential image augmentation model included RandomFlip (50% probability), RandomRotation (0.2 radians), RandomZoom (0-20%), RandomHeight/RandomWidth (80-120% resizing), Rescaling ([0, 1] scaling), and Resizing (224×224 standardization) to improve model generalization [18].

B. Error Level Analysis (ELA)

Error Level Analysis (ELA) is usually performed [19] to examine the performance of the model. ELA is a forensic image analysis method that detects digital image tampering and identifies altered areas. It works by using multiple compressions to introduce different error levels in an image. To highlight these inconsistencies, the ELA calculates the error difference between the original image and a compressed version. Cloned, retouched, and superimposed regions have higher error rates. ELA, or pixel error level, indicates the difference in intensity values between the original and the compressed image. *I_original* represents the pixel's intensity in the original picture and *I_compressed* represents it in the compressed form.

The dataset was split into 80% for training, 10% for validation, and 10% for testing. Sequential augmentation reduced overfitting. ETPLDNet and LEViT models predicted labels with Adam optimization.

C. Optimized CNN Model for Detecting and Classifying Leaf Diseases

The misdiagnosis of leaf diseases and the challenges of disease diversity and environmental factors underscore the importance of early detection and treatment. CNNs have advanced image-based recognition by enabling built-in feature selection and eliminating the need for extensive image preprocessing [20]. However, obtaining large datasets for such problems remains challenging.

CNNs typically consist of input, hidden, and output layers, including convolution, normalization, pooling, and fully-connected layers that generate classification feature maps using filters [20]. The Rectified Linear Unit (ReLU) activation function is commonly used in image processing. An optimized CNN model with a modified layered architecture has shown accurate leaf disease classification by capturing complex image features [9]. This model employs hierarchical levels for classification tasks, utilizes multiple optimizers with varied learning rates for enhanced performance, and includes max pooling, normalization, and three convolutional 2D blocks. Hyperparameter optimization is used to analyze learning rate and optimizer selections, leading to improved leaf disease classification and prediction.

The preprocessed input image with a resolution of 256×256 is fed into this model. The optimized CNN model consists of three convolutional (Conv2D) layers, each followed by max pooling and batch normalization layers. The first Conv2D layer has 128 filters, while the second and third have 256 filters each. After these layers, the output is flattened into a 1D array using a flattened layer, which is then passed through two fully connected (dense) layers: one with 512 neurons and the other with 256 neurons, both followed by dropout layers to prevent overfitting. The final output layer is a dense layer with 38 neurons for classification. Additionally, dynamic hyperparameter tuning, such as learning rate at 0.0001 and the Adam optimizer, enhance the model's ability to adapt and improve performance, setting it apart from standard CNN models with fixed architectures and static hyperparameter settings.

D. Multi-Class Accuracy in Leaf Disease Classification Using the ETPLDNET Model

The ETPLDNet model is a potent deep transfer learning XAI system that is specifically engineered to accurately identify and diagnose diseases in plant leaves. Grad-CAM is employed to produce it following the merge of ResNet50 with extra thick and dropout layers. Figure 2 illustrates the fundamental design of ETPLDNet.

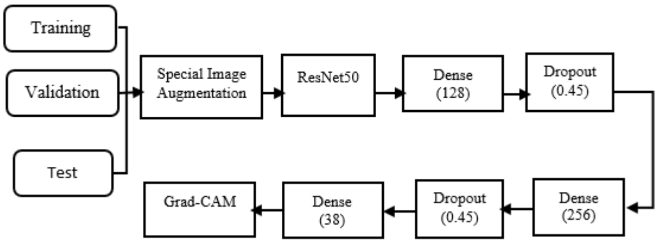

Fig. 2. ETPLDNet architecture.

Transfer learning in leaf disease classification addresses challenges such as limited labeled data and the complexity of diverse plant diseases. Transfer learning leverages pre-trained models, such as ResNet50, to improve performance, reduce manual labeling, and accelerate training, while maintaining high accuracy. This approach enables efficient, scalable solutions for real-world agricultural applications with minimal data and computational effort.

The 55-layer ETPLDNet model, based on ResNet50, predicts and classifies leaf diseases [21]. ResNet50, a Residual Network with 50 layers, revolutionized deep neural network training by addressing gradient drops and accuracy loss with skip connections that improve gradient flow and facilitate learning complex structures [21]. ETPLDNet leverages ResNet50 by using its deep residual learning architecture, which includes skip connections to overcome vanishing gradients and efficiently learn complex features. To enhance its performance for leaf disease detection, ETPLDNet incorporates additional dense layers (128 and 256 units) and a dropout layer (45% dropout rate) to prevent overfitting and improve the model's capacity to capture intricate patterns. It also integrates Grad-CAM, an XAI technique, to provide visual explanations of the model's decisions, making the classification process more transparent and interpretable.

Furthermore, the model employs hyperparameter optimization, adjusting learning rates and batch sizes to maximize its performance, achieving superior accuracy across 38 disease classes and 14 plant species compared to the standard ResNet50 model. These enhancements make ETPLDNet more accurate, interpretable, and suitable for practical agricultural use. The model is trained using mini-batch stochastic gradient descent. ETPLDNet's dense layers combine features non-linearly to handle 38 classes and capture complex leaf disease identification and categorization patterns [21].

E. Multi-Class Accuracy in Leaf Disease Classification and Detection Using LEViT

Transformers are used mainly for NLP [11]. ViTs were created to classify images with transformers. Language transformers extract words from sentences, whereas ViTs create patches [22]. A misaligned patch would distort the image, so its position matters. A 224×224 image dimension was chosen, which can be divided into 16×16 patches, resulting in 256 patches. Each patch corresponds to a grid of 14×14 pixels within the original image. The advanced vision transformer LEViT detects and classifies plant leaf diseases using XAI, based on Grad-CAM, dense layers, and ViT_ImageNet-21k. Figure 3 illustrates the architecture of the LEViT model.

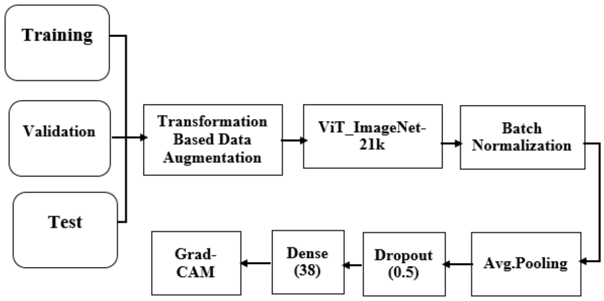

Fig. 3. LEViT architecture with added layers.

Training, validation, and testing data are provided to the Transformation-based data augmentation module to be altered. Then, the extra data are examined by an already trained ViT model, called ViT_ImageNet-21k, using the large ImageNet-21k dataset. Batch normalization stabilizes activation patterns, average pooling condenses feature maps, and 0.5 probability dropout stops overfitting. A layer of 38 units predicts classes. Finally, Grad-CAM uses gradient-weighted class activation mapping to explain the model's output. The output is separated into 38 classes using the 12-segment LEViT model encoding. In this LEViT model, images are converted into segments and processed using transformer blocks to take advantage of the vision transformer mechanism.

F. Explainability and Interpretability using Grad-CAM

CNN researchers propose deeper topologies for higher-level visuals. Convolutional layers save spatial data, reconciling spatial and semantic features. Grad-CAM [5] uses sinuous gradients for localization maps, prioritizing the choice of the output layer based on class score gradients [5]. The expanded Grad-CAM process uses the dataset to train the leaf disease classification deep learning model. The last convolutional layer outputs and projects class scores from the training model. Gradients and weight feature maps are updated to generate relevance ratings, creating a heatmap that showcases the model's decision-making process for the predicted classes.

G. Optimizers

Three optimizers were used to compare the results of the models [16]. Adam updates weights using AdaGrad, managing per-parameter learning rates, RMSProp adjusts learning rates based on running averages, and SGD optimizes loss functions iteratively, minimizing the objective function by updating weights iteratively based on the learning rate η, as described in (1)-(4). The RMSProp optimizer is defined by:

$$(w,t) \coloneqq (w,t-1) + (1-\gamma)\left(\nabla Q_i(w)\right)^2 \quad (1)$$

The learning parameters are updated as follows:

$$w \coloneqq w - \frac{n}{\sqrt{v(w,t)}} \nabla Q_i(w) \quad (2)$$

The SGD optimizer is defined as:

$$Q(W) = \frac{1}{n} \sum_{i-n}^{n} Q_i(W) \quad (3)$$

$$w \coloneqq w - \eta \nabla Q(w) \quad (4)$$

where η is the learning rate.

H. Evaluation Metrics

The evaluation metrics used for the models are presented in (5)-(8) [16]:

$$Accuracy = \frac{Total\ Number\ of\ Correct\ Predictions}{Total\ Predictions} \quad (5)$$

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (6)$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (7)$$

$$F1 - score = \frac{2*(Precision*Recall)}{Precision+Recall} \quad (8)$$

III. RESULTS AND ANALYSIS

Python was used to implement the proposed models along with TensorFlow [17] and Keras for the transfer learning models. The Adam optimizer was used for training. The experimental analysis was carried out on a GPU-enabled P-100 server with an i5 processor and 8 GB RAM.

A. Error Level Analysis (ELA)

ELA is used to detect image tampering in forensic analysis, identifying compressed or altered image regions. Comparing error levels across picture portions helps ELA find significant changes. Plant disease images were evaluated using ELA. ELA provides images with different q compression values. Visually evaluating photos at different compression settings helps find the best image processing setting and how compression affects quality. ELA analysis adjusts training compression for damaged dataset images, ensuring image processing and data integrity. Figure 4 shows ELA analysis on the Apple___Apple_scab leaf image.

B. Evaluation of the Optimized CNN on the New Plant Disease Dataset

Multiple methods with early stopping were used to train the plant disease model. The proposed method uses CNN to classify diseased and healthy leaves. CNN's accuracy was 99.02%. The CNN model was best at distinguishing between healthy and bacterial leaves. Figure 5 shows the performance of the optimized CNN model. Since dynamic drift was implemented in the learning rate through the reduced learning rate feature, there is a variation in performance accuracy and loss at epoch 5. This improves the accuracy of the model and makes it perform better than competing models. The average multiclass accuracy achieved was 99.02% on the 38 classes of the dataset.

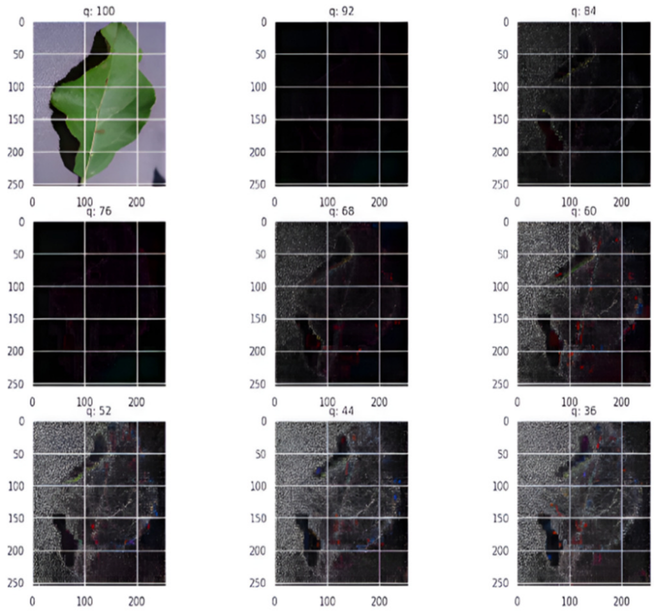

Fig. 4. ELA analysis on the Apple___Apple_scab leaf image.

(a)

(b)

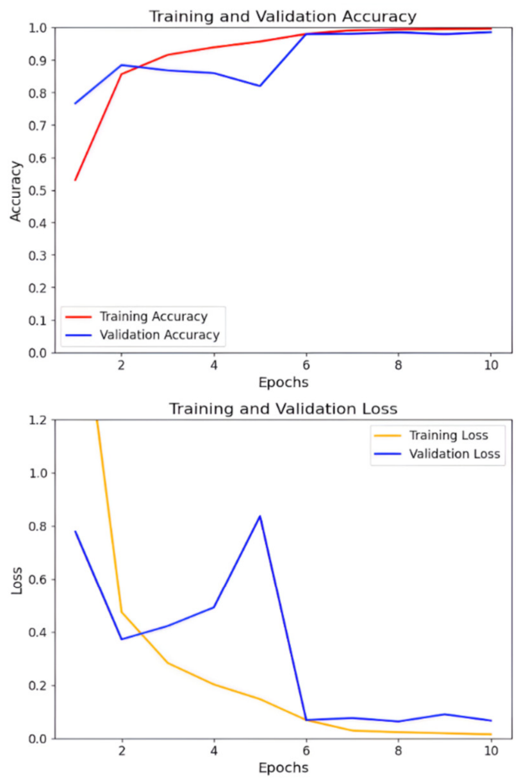

Fig. 5. The effectiveness of the fine-tuned CNN model.

TABLE I. PERFORMANCE COMPARISON OF CNN MODELS

Model	Accuracy (%)
CNN model with size 224×224	95.33
CNN model with size 210×210	91.09
CNN model with size 256×256	98.27
Optimized CNN model with 62×62	99.58

Table I compares the sequential CNN performance. Comparing model architectures and settings shows large multiclass accuracy differences. CNN model accuracy depends on input size: 224×224, 210×210, and 256×256 achieved 95.33%, 91.09%, and 98.27%, respectively. The optimized CNN model with 62×62 outperformed the others with 99.58% multiclass accuracy.

C. ETPLDNet Performance on New Plant Disease Dataset

Figure 6 shows that the ETPLDNet model classified leaf diseases well on the New Plant Disease Dataset. The ETPLDNet model performed well, with 92.53% training accuracy and 0.2294% loss after 100 epochs. The model had an average multiclass accuracy of 97.33% and a loss of 0.0814. These impressive results demonstrate the ETPLDNet model's capacity to reliably classify leaf diseases across classes. The model's accuracy and low loss values show its ability to capture rich characteristics and patterns in leaf disease images for accurate predictions. Through automated leaf disease identification and categorization, the ETPLDNet model can improve plant health management and agricultural productivity.

Figure 6 shows that the differences in training and validation datasets result in lower accuracy and higher loss. The model may overfit the training data after many epochs and iterations. More than 100 training epochs reduce the discrepancy between the training and validation metrics, indicating better model generalization. For fewer epochs, tiny data changes and overfitting may explain the training-validation parameter discrepancy, including larger training than validation loss. The prevalence of this problem in machine learning highlights validation accuracy as a generalization indicator. The proposed ETPLDNet model surpassed previous studies with 92.53% accuracy. These results show that the ETPLDNet model can effectively diagnose plant diseases and can help farmers detect them more accurately. However, more evaluation measures and comparative research on diverse datasets are needed to fully establish its advantage.

D. LEViT Performance on the New Plant Disease Dataset

The performance of the LEViT model was evaluated over 50 training epochs, as shown in Figure 7. The model has a validation accuracy of 96.19% and a training accuracy of 95.22% after 50 epochs. The validation loss is substantially smaller at 0.1023 than the training loss of 0.1023. These results demonstrate the model's capacity to generalize knowledge from training to unseen data. To reduce overfitting and improve model generalizability, hyperparameters such as learning rate and regularization techniques should be adjusted to obtain an ideal loss-to-training accuracy trade-off.

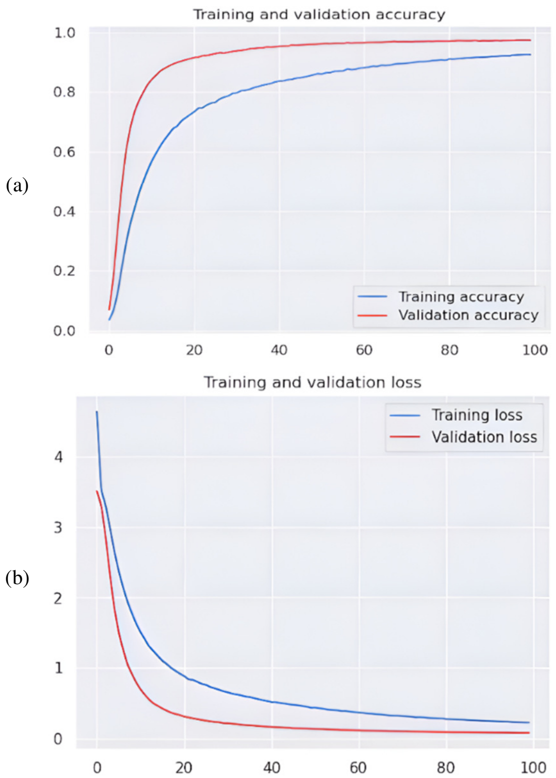

Fig. 6. Analysis of ETPLDNet's performance on the dataset.

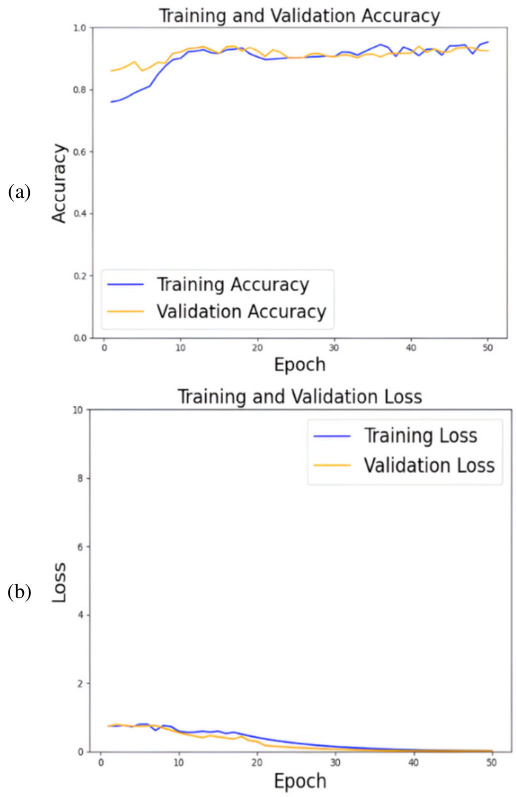

Fig. 7. Performance of the LEViT model.

LEViT, a revolutionary model for the diagnosis and classification of leaf diseases, surpasses the competition with an accuracy of 95.22% as a result of its enhanced vision transformer architecture. LEViT beat ResNet50 [24], VGG-16 [25], InceptionV3 [25], CNN, and ETPLDNet [23] in the detection of plant diseases. LEViT is a promising new instrument for precision agriculture and plant pathology disease management due to its high efficacy.

TABLE II. PERFORMANCE ANALYSIS OF DEEP LEARNING MODELS ON NEW PLANT DISEASE DATASET.

Model	Accuracy(%)
ResNet50 [24]	91.81
VGG-16 [25]	90.40
InceptionV3 [25]	87.84
CNN	91.09
ETPLDNet	92.53
LEViT [16]	95.22

E. Grad-CAM Analysis

Grad-CAM visualized ETPLDNet and LEViT neural networks for leaf disease classification. Model projections were strengthened by heat maps of ETPLDNet's decision-making and LEViT's diseased region identification. The ETPLDNet Grad-CAM model can accurately detect plant disease zones, improving plant health and agricultural productivity through targeted treatments and management, as shown in Figure 8. Grad-CAM with LEViT identifies critical regions in the leaf image for accurate prediction. Discriminative properties aid accurate predictions. Figure 9 shows that a fine-tuned LEViT model with Grad-CAM can detect leaf disease locations.

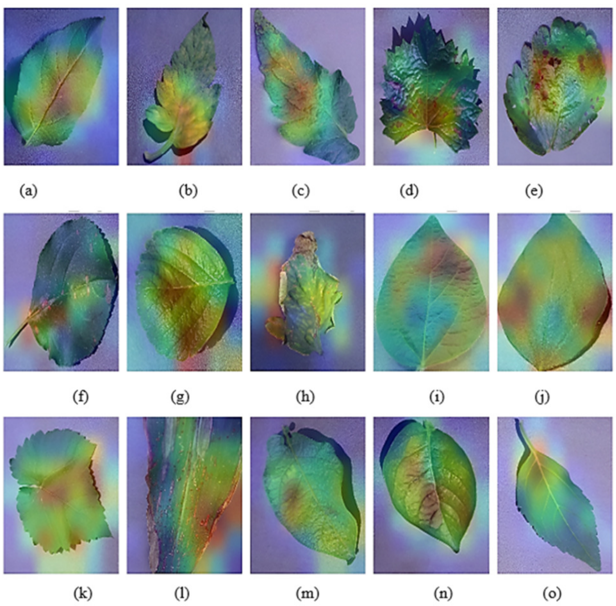

Fig. 8. ETPLDNet model with Grad-CAM for accurately predicting disease regions on leaves (correct predictions): (a) Apple_healthy, (b) Tomato Leaf Mold, (c) Tomato Septoria leaf spot, (d) Grape Esca (Black Measles), (e) Strawberry Leaf scorch, (f) Apple Black rot, (g) Strawberry healthy, (h) Tomato Bacterial spot, (i) Soybean_healthy, (j) Soybean_healthy, (k) Grape healthy, (l) Corn (maize) Northern Leaf blight, (m) Potato Late blight, (n) Potato healthy, (o) Cherry (including sour) healthy.

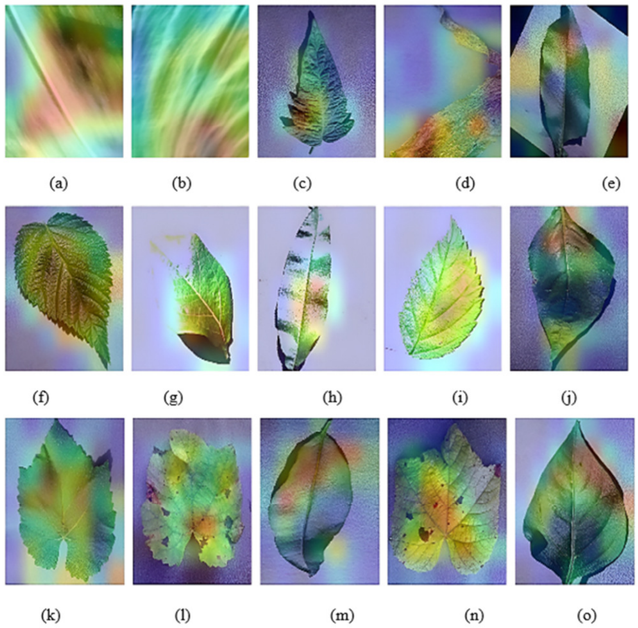

Fig. 9. LEViT model with Grad-CAM for accurately predicting disease regions on leaves (correct predictions if not mentioned otherwise: (a): Corn (maize) healthy, (b) Corn (maize) healthy, (c) Tomato Spider Mites (Two-Spotted Spider Mite), (d) True: Corn_(maize) Northern Leaf Blight - Predicted: Corn (maize) Cercospora leaf spot Greay leaf spot, (e) Peach healthy, (f) Raspberry healthy, (g) Blueberry healthy, (h) Peach healthy, (i) Raspberry healthy, (j) True: Pepper bell Bacterial spot - Predicted: Pepper bell healthy, (k) Grape healthy, (l) Grape Leaf blight (isariopsis_Leaf_Spot), (m) True: Peach healthy - Predicted: Apple healthy, (n) Grape Leaf blight (isariopsis Leaf Spot), (o) Pepper bell healthy.

IV. DISCUSSION

The ETPLDNet model achieved 92.53% training accuracy and a loss of 0.2294% after 100 iterations, with a multiclass accuracy of 97.33% and a loss of 0.0814, indicating its ability to capture rich characteristics for accurate predictions [23]. Compared to ResNet50 [24], VGG-16 [25], and InceptionV3 [25], ETPLDNet shows superiority, albeit marginally less accurate than DenseNet121. This study compared the performance of the LEViT model with ResNet50 [24], VGG-16 [25], and InceptionV3 [25], showing that LEViT achieved superior accuracy at 95.22%, outperforming ResNet50 (91.81%), VGG-16 (87.99%), and InceptionV3 (91.84%). LEViT's ViT architecture with self-attention mechanisms enhances feature extraction and generalization, while XAI integration improves interpretability. These attributes make LEViT the most effective model for leaf disease classification. The LEViT model achieved 95.22% training accuracy with 0.1038% loss after 50 iterations, and a 92.33% test accuracy on the same dataset, showcasing its potential for reliable plant disease detection and management beyond models such as ResNet50 [24], VGG-16 [25], InceptionV3 [25], CNN, and ETPLDNet. LEViT consistently outperformed these architectures, positioning itself as a superior method for precise plant disease identification in agriculture.

The ETPLDNet, LEViT, and CNN models can be implemented on different datasets for leaf disease detection using data augmentation, preprocessing, and transfer learning.

A comparison study across datasets can reveal insights into their generalization and robustness for real-world agricultural use.

V. CONCLUSION AND SUMMARY

This study showcased significant advances in plant disease detection using deep learning models, particularly CNN- and ViT-based approaches. These models, such as ETPLDNet and LEViT, demonstrated impressive accuracy rates, outperforming traditional sequential CNNs and transfer learning models while also offering interpretability through XAI Grad-CAM. These models improved leaf disease detection accuracy and interpretability compared to previous works. Previous models such as VGG-16 and InceptionV3 achieved accuracies of 87.99% and 91.84%, respectively, but lacked transparency in decision-making and struggled with generalization across diverse datasets. The optimized CNN model achieved exceptional performance with ResNet50 and VGG-16. ETPLDNet excelled in leaf disease identification, surpassing ResNet50 and VGG-16, while LEViT showed substantial improvements with accuracy reaching 95.22%, 96.19%, and 92.33%. Future research should explore XAI transfer learning on diverse datasets to improve generalization and use ensemble learning and ViT to improve model performance and transparency.

REFERENCES

[1] R. G. Dawod and C. Dobre, "Upper and Lower Leaf Side Detection with Machine Learning Methods," *Sensors*, vol. 22, no. 7, Jan. 2022, Art. no. 2696, https://doi.org/10.3390/s22072696.

[2] L. V. Madden, G. Hughes, and F. van den Bosch, *The Study of Plant Disease Epidemics*. American Phytopathological Society, 2007.

[3] G. Bhadur and R. Rani, "Agricultural Crops Disease Identification and Classification through Leaf Images using Machine Learning and Deep Learning Technique: A Review," in *Proceedings of the International Conference on Innovative Computing & Communications (ICICC) 2020*, Mar. 2020, https://doi.org/10.2139/ssrn.3564973.

[4] A. R. B. Patil *et al.*, "A Literature Review on Detection of Plant Diseases," *European Journal of Molecular & Clinical Medicine*, vol. 7, no. 7, pp. 1605–1614, 2020.

[5] Y. M. Abd Algani, O. J. Marquez Caro, L. M. Robladillo Bravo, C. Kaur, M. S. Al Ansari, and B. Kiran Bala, "Leaf disease identification and classification using optimized deep learning," *Measurement: Sensors*, vol. 25, Feb. 2023, Art. no. 100643, https://doi.org/10.1016/j.measen.2022.100643.

[6] J. Liu and X. Wang, "Plant diseases and pests detection based on deep learning: a review," *Plant Methods*, vol. 17, no. 1, Feb. 2021, Art. no. 22, https://doi.org/10.1186/s13007-021-00722-9.

[7] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," *Frontiers in Plant Science*, vol. 7, Sep. 2016, https://doi.org/10.3389/fpls.2016.01419.

[8] J. Boulent, S. Foucher, J. Théau, and P.-L. St-Charles, "Convolutional Neural Networks for the Automatic Identification of Plant Diseases," *Frontiers in Plant Science*, vol. 10, Jul. 2019, https://doi.org/10.3389/fpls.2019.00941.

[9] R. Sharma *et al.*, "Plant Disease Diagnosis and Image Classification Using Deep Learning," *Computers, Materials & Continua*, vol. 71, no. 2, pp. 2125–2140, 2022, https://doi.org/10.32604/cmc.2022.020017.

[10] A. Ahmad, D. Saraswat, and A. El Gamal, "A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools," *Smart Agricultural Technology*, vol. 3, Feb. 2023, Art. no. 100083, https://doi.org/10.1016/j.atech.2022.100083.

[11] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." arXiv, 2020, https://doi.org/10.48550/ARXIV.2010.11929.

[12] T. Lin, Y. Wang, X. Liu, and X. Qiu, "A survey of transformers," *AI Open*, vol. 3, pp. 111–132, Jan. 2022, https://doi.org/10.1016/j.aiopen.2022.10.001.

[13] H. Wu *et al.*, "CvT: Introducing Convolutions to Vision Transformers," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, Canada, Oct. 2021, pp. 22–31, https://doi.org/10.1109/ICCV48922.2021.00009.

[14] P. S. Thakur, P. Khanna, T. Sheorey, and A. Ojha, "Explainable vision transformer enabled convolutional neural network for plant disease identification: PlantXViT." arXiv, Jul. 16, 2022, https://doi.org/10.48550/arXiv.2207.07919.

[15] S. Alqethami, B. Almtanni, W. Alzhrani, and M. Alghamdi, "Disease Detection in Apple Leaves Using Image Processing Techniques," *Engineering, Technology & Applied Science Research*, vol. 12, no. 2, pp. 8335–8341, Apr. 2022, https://doi.org/10.48084/etasr.4721.

[16] Bh. Prashanthi, A. V. P. Krishna, and Ch. M. Rao, "LEViT- Leaf Disease identification and classification using an enhanced Vision transformers(ViT) model," *Multimedia Tools and Applications*, Aug. 2024, https://doi.org/10.1007/s11042-024-19866-6.

[17] S. Bhattarai, "New Plant Diseases Dataset." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset.

[18] "Data augmentation," *TensorFlow*. https://www.tensorflow.org/tutorials/images/data_augmentation.

[19] J. Meshram, "jayant1211/Image-Tampering-Detection-using-ELA-and-Metadata-Analysis." Oct. 30, 2024, [Online]. Available: https://github.com/jayant1211/Image-Tampering-Detection-using-ELA-and-Metadata-Analysis.

[20] W. S. Jeon and S.-Y. Rhee, "Plant Leaf Recognition Using a Convolution Neural Network," *The International Journal of Fuzzy Logic and Intelligent Systems*, vol. 17, no. 1, pp. 26–34, Mar. 2017, https://doi.org/10.5391/IJFIS.2017.17.1.26.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, https://doi.org/10.1109/CVPR.2016.90.

[22] "Vision Transformers (ViT) Explained," *Pinecone*, https://www.pinecone.io/learn/series/image-search/vision-transformers/.

[23] E. Daglarli, "Explainable Artificial Intelligence (xAI) Approaches and Deep Meta-Learning Models for Cyber-Physical Systems," in *Artificial Intelligence Paradigms for Smart Cyber-Physical Systems*, IGI Global Scientific Publishing, 2021, pp. 42–67.

[24] K. Balavani, D. Sriram, M. B. Shankar, and D. S. Charan, "An Optimized Plant Disease Classification System Based on Resnet-50 Architecture and Transfer Learning," in *2023 4th International Conference for Emerging Technology (INCET)*, Belgaum, India, May 2023, pp. 1–5, https://doi.org/10.1109/INCET57972.2023.10170368.

[25] G. Wang, Y. Sun, and J. Wang, "Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning," *Computational Intelligence and Neuroscience*, vol. 2017, no. 1, 2017, Art. no. 2917536, https://doi.org/10.1155/2017/2917536.

AUTHORS PROFILE

Bh. Prashanthi was admitted for a Ph.D. in KLU in 2020 and joined the Department of Computer Science and Engineering, GRIET, Hyderabad in 2016. She earned an M. Tech. (Computer Science) from RRS College of Engineering, Hyderabad 2013. Her area of research interest includes text mining, machine learning, and data science. She has 12 years of teaching experience. She has published articles in international conferences and journals.

Dr. Anne Venkata Praveen Krishna is an Associate Professor at Koneru Lakshmaih Education Foundation, deemed to be University. He is actively involved in research and publications in the areas of network security, intrusion detection, artificial intelligence, machine learning, and IOT. He has active research publications in reputed Springer, Elsevier, IEEE Explorer, SCI, and Scopus-based international journals. He was recognized by the Association of Computing Machinery (ACM) as one of the top brand ambassadors for 2016. He was also appointed as a Sectional Committee Member of the 105th Indian Science Congress for Information Science & Technology for 2017.

Dr. Ch. Mallikarjuna Rao is a professor at the Computer Science and Engineering Department at Gokaraju Rangaraju Institute of Engineering and Technology. He has a B. E. degree from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, obtained in 1998. He also earned an M. Tech. degree in Computer Science Engineering from JNTUA, Ananthapuramu, Andhra Pradesh in 2007. In 2016, he was awarded a Ph.D. from the same university (JNTUA, Ananthapuramu). With over 25 years of teaching experience, Dr. Rao has a distinguished academic career. He has actively published more than 30 research articles in reputed international journals and conferences and two patents. Currently, he is supervising four Ph. D scholars. His areas of research interest include data mining, big data, and machine learning.