International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Image Based Diagnosis for Plant leaf Disease Detection Using an Efficient CNN Approach

Pasham Divya[1], K. Jayasri [2]

[1]Student, School of Informatics, Department of MCA, Aurora Deemed University, Hyderabad
[2]Assistant Professor, School of Engineering, Department of CSE, Aurora Deemed University, Hyderabad
[1]pashamdivya.7@gmail.com, [2]jayasri@aurora.edu.in

ABSTRACT:

Diseases in plants cause significant loss in crop yield, quality, and farmer income, making early and accurate diagnosis essential. Traditional methods relying on expert visual inspection are time-consuming, error-prone, and inefficient for large-scale agriculture. This proposes an image-based diagnosis system that employs a Convolutional Neural Network (CNN) architecture, VGG19, for detecting and classifying plant leaf diseases across 38 different disease classes.

The system involves data preprocessing techniques like augmentation, normalization, and resizing to improve training efficiency. It achieves an average accuracy of 98.51%, demonstrating the capability of CNN models for precise disease detection. Evaluation metrics such as precision, recall, F1-score, and accuracy provide comprehensive performance analysis. The proposed model enhances early detection, supports farmers in timely disease management, and contributes to sustainable agriculture.

Keywords: Plant Disease Detection, CNN, VGG19, Deep Learning, Image Processing, Convolutional Neural Network.

I. INTRODUCTION

Agriculture is a vital sector that significantly influences the economic stability and livelihood of millions of people worldwide, particularly in countries like India where a large portion of the population depends on farming. Plant diseases pose a severe threat to crop productivity, leading to considerable losses in yield and quality. Reports indicate that nearly 70% of the population in India relies on agriculture, contributing around 17% to the national GDP, highlighting the importance of maintaining healthy crops to ensure food security and economic growth.

Traditionally, plant disease identification has been performed through manual visual inspection by experts, which is time-consuming, labour-intensive, and prone to human error. Moreover, such methods are often not scalable for large agricultural fields and can result in delays in diagnosis, leading to significant crop damage. Farmers also face challenges in adopting multiple disease control methods without proper and timely detection.

Recent advances in Deep Learning and Computer Vision have revolutionized the field of automated disease detection. Convolutional Neural Networks (CNNs) have proven highly effective in image classification tasks due to their ability to extract and learn hierarchical features from images. Among CNN architectures, VGG19 has shown promising results in identifying complex patterns, making it an excellent choice for plant leaf disease detection.

The proposed work aims to develop an image-based disease diagnosis system using a customized VGG19 CNN architecture, designed to classify 38 disease types from images of plant leaves. The system uses advanced preprocessing techniques and evaluates performance using multiple metrics, including accuracy, precision, recall, and F1-score, achieving an impressive 98.51% accuracy. This approach not only automates disease detection but also supports farmers in making informed decisions, ultimately improving crop quality and yield.

II. LITERATURE REVIEW

Image-based plant disease detection systems have evolved from traditional image processing methods to advanced deep learning techniques that leverage Convolutional Neural Networks (CNNs). Current research highlights significant improvements in accuracy and automation but also points out limitations related to dataset diversity, real-time adaptability, and field deployment.

Demilie [1] presented a comparative study of machine learning (ML) and deep learning (DL) techniques such as CNN, KNN, and SVM for plant disease detection. The study concluded that CNN models performed best, especially with large datasets, achieving high detection accuracy. However, real-time adaptability and multilingual support for farmers were not addressed.

Bhargava et al. [2] proposed a computer vision-based system for detecting and diagnosing plant leaf diseases using CNN and few-shot learning techniques. The model achieved high classification accuracy under lab conditions but lacked adaptability to real-world scenarios and multilingual support.

Li et al. [3] reviewed deep learning methods such as CNN and transfer learning for plant protection and found that these approaches outperformed traditional techniques. However, the study identified limitations such as the need for larger and more diverse datasets and challenges in supporting low-resource devices.

Aboelenin et al. [4] introduced a hybrid model combining CNN and Vision Transformers (ViT) for enhanced classification accuracy, achieving results above 99%. Despite its superior performance, the complexity of the architecture makes it less practical for small-scale farmers due to hardware requirements.

Krishna et al. [5] emphasized the importance of generalization by using multiple datasets with CNNs and EfficientNet architectures. The approach improved performance across diverse images but lacked user-friendly interfaces and farmer-centric features.

Gawande et al. [6] provided a foundation by reviewing traditional image processing techniques like segmentation and thresholding. While these methods showed accuracy under controlled conditions, they lack deep learning advancements, limiting real-world application.

Ingole et al. [7] reviewed basic image processing algorithms such as K-means and ANN, noting their potential but also their inability to deliver real-time solutions suitable for field use.

While each study contributes valuable insights, they collectively reveal the need for a task-specific deep learning model one that supports multi-class disease detection, provides high accuracy (98.51%), and is adaptable for practical agricultural environments. The proposed work addresses these gaps by developing a VGG19-based CNN architecture with robust preprocessing, improved generalization, and comprehensive performance evaluation.

III. METHODOLOGY

A. Existing Methodology

Existing approaches for plant leaf disease detection primarily rely on traditional image processing techniques or shallow machine learning models. These methods involve manual feature extraction followed by classification using algorithms like SVM, KNN, or ANN.

Key Features of Existing Methods:

- Segmentation & Feature Extraction: Images are segmented to isolate diseased areas, and features such as texture, color, and shape are manually extracted.
- Classification Models: Algorithms like SVM or ANN are used for classification.
- Pretrained CNN Models: Some approaches use pretrained architectures such as AlexNet or VGG16 for transfer learning to improve accuracy.

Limitations:

- Require manual feature engineering, which is labor-intensive and less accurate.
- Limited generalization due to small and less diverse datasets.
- Lack of real-time adaptability and user-friendly field deployment.
- Evaluation mostly based on accuracy, ignoring other metrics like precision, recall, and F1-score.

B. Proposed Methodology

The proposed methodology introduces an image-based disease diagnosis system using a customized VGG19 CNN architecture optimized for detecting and classifying 38 different plant diseases. It emphasizes data preprocessing, model optimization, and multi-metric evaluation to enhance detection accuracy.

1. System Architecture

- Custom CNN Design: A VGG19-based model with 16 convolutional layers, 5 max-pooling layers, and 3 fully connected layers is used.
- Global Average Pooling (GAP): Added to improve generalization and reduce overfitting.
- Task-Specific Training: Unlike pretrained models, this CNN is trained from scratch to focus on disease-specific features.

2. Data Preprocessing

- Data Augmentation: Rotation, flipping, resizing, and cropping are applied to increase dataset diversity and avoid overfitting.
- Normalization: Ensures consistent pixel value ranges across all images.
- Resizing: Images are resized to 224×224 pixels to match VGG19 input requirements.

- Data Splitting: Dataset is divided into training (80%), validation (20%), and testing sets, with stratified sampling to maintain class balance.

3. Functional Modules

- Image Acquisition Module: Captures plant leaf images for diagnosis.
- Disease Detection Module: Uses the CNN model to classify input images into one of the 38 disease categories.
- Evaluation Module: Computes performance metrics such as precision, recall, F1-score, and accuracy.
- Visualization Module: Displays detection results, including predicted disease type and confidence score.

4. Performance Evaluation

The model is trained for multiple epochs using categorical cross-entropy loss and Adam optimizer. Evaluation metrics include:

- Precision – Accuracy of positive predictions.
- Recall – Ability to detect all diseased leaves correctly.
- F1-score – Balance between precision and recall.
- Overall Accuracy – Percentage of correctly classified images.

IV. SYSTEM DESIGN AND ARCHITECTURE

The system is designed using a multi-layer architecture to ensure modularity, scalability, and accuracy in plant disease detection. The design focuses on efficient image preprocessing, feature extraction through CNN layers, and classification using a customized VGG19 model.

A. Architectural Diagram

1. Presentation Layer
   - Provides a user-friendly interface for uploading plant leaf images.
   - Displays disease predictions, confidence levels, and performance metrics.
   - Supports visualization of results using processed images and classification labels.
2. Application Layer
   - Hosts the VGG19 Convolutional Neural Network for disease detection.
   - Handles data preprocessing (augmentation, normalization, resizing) and model inference.
   - Includes Global Average Pooling (GAP) to enhance generalization and minimize overfitting.
   - Uses categorical cross-entropy as the loss function and Adam optimizer for training.
3. Data Layer
   - Maintains labelling of healthy and diseased leaf images for 38 disease classes.
   - Supports structured data management for efficient model retraining and performance updates.

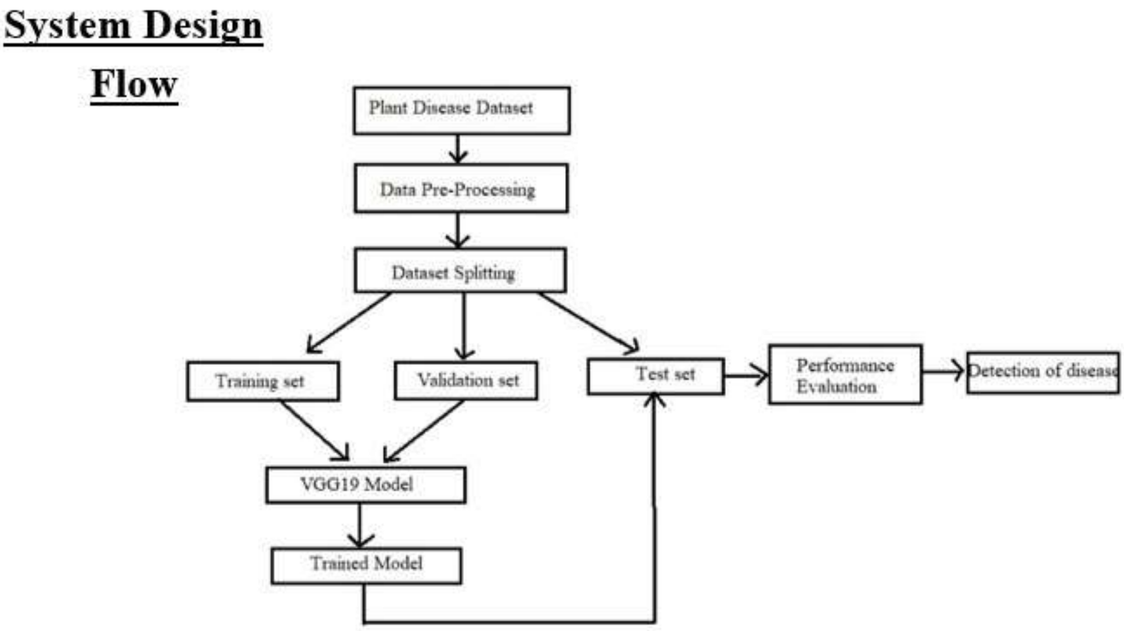

Fig. 1: System Diagram Flow

B. Security and Scalability Considerations

 Security: Data is processed locally, preventing unauthorized access to plant images and predictions.

 Scalability: Additional disease types, datasets, or model improvements (e.g., ResNet, EfficientNet) can be integrated without altering the core architecture.

 Extensibility: Can be upgraded for real-time field diagnosis**,** mobile application deployment, or IoT-based agriculture systems**.**

C. Data Flow Diagram (DFD):

Data Flow Diagrams (DFDs) describe the processes of how the transfer of data takes place from the input till prediction of the corresponding output.

DFD – Level 0

The DFD Level 0 depicts the users to input the image of the plant leaves. The system in turn detects and recognizes the plant leaf disease.

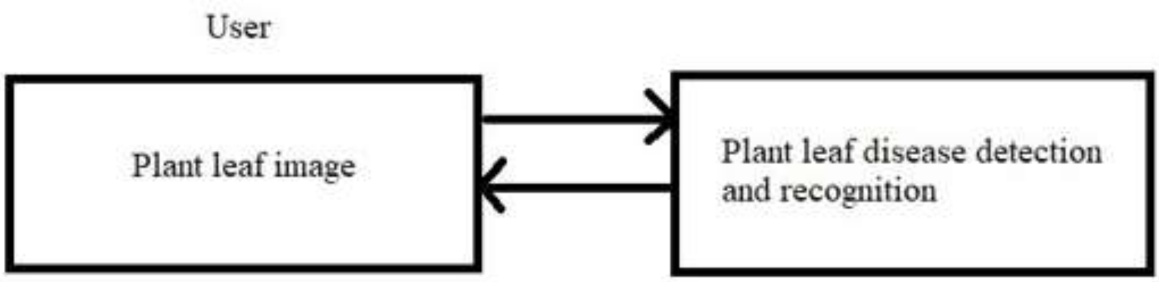

Fig. 2: Data Flow Diagram (Level 0)

DFD – Level 1

The below Figure displays the DFD Level 1, where the CNN model takes the image from the training dataset and then CNN model predicts the type of disease of the leaf.

Fig. 3: Data Flow Diagram (Level 1)

DFD – Level 2

DFD Level 2 goes one step deeper into parts of 1-level DFD. It can be used to plan or record the specific/necessary detail about the system's functioning.

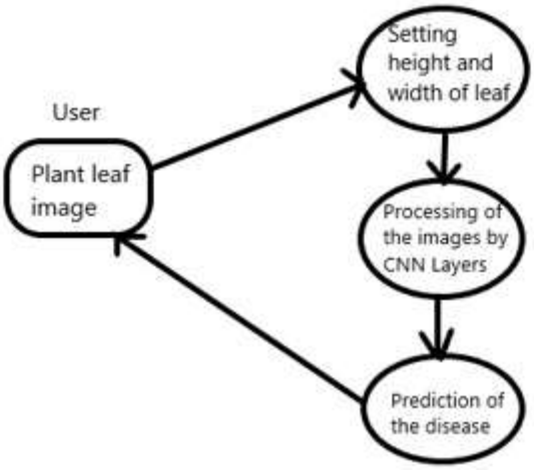

Fig. 4: Data Flow Diagram (Level 2)

V.IMPLEMENTATION

A. Frontend Implementation

The frontend provides a user-friendly interface for uploading plant leaf images and displaying disease predictions.

- Upload Screen: Users can select images of plant leaves for analysis.
- Prediction Display: The system shows the predicted disease name and corresponding confidence level.
- Visualization of Results: Sample results include detection of diseases such as *Tomato Yellow Leaf Curl Virus*, *Cedar Apple Rust*, and *Potato Early Blight*.

B. Model Training (Proposed VGG19 Architecture)

- Architecture:
  - The model is based on VGG19 architecture, which consists of 16 convolutional layers, 5 max-pooling layers, and 3 fully connected layers.
  - Global Average Pooling (GAP) is integrated to enhance interpretability and generalization.
- Training Process:
  - Dataset images are resized to 224×224 pixels.
  - Preprocessing includes augmentation (rotation, flipping, resizing, cropping), normalization, and shuffling to ensure consistency and prevent overfitting.
  - Model trained using categorical cross-entropy loss and Adam optimizer for ~50 epochs with a batch size of 22.

C. Dataset Implementation

- Dataset Composition:
  - Contains images of 38 plant diseases across 14 plant species.
- Data Splitting:

- Training Set: 80% of dataset for model learning.
- Validation Set: 20% of dataset for tuning and overfitting checks.
- Test Set: 33 images used for final performance evaluation.

D. Evaluation Metrics

The model was evaluated using four performance metrics:

1. Precision: Accuracy of positive predictions.
2. Recall: Ability to correctly identify diseased leaves.
3. F1-Score: Harmonic mean of precision and recall for balanced performance.
4. Accuracy: Overall correctness of predictions.

Final results indicate that the VGG19 model achieved an accuracy of 98.51%, outperforming VGG11, VGG13, and VGG16.

DATASET

PLANT DISEASE DATASET

Title	Description
Image size	256*256 pixel
number of plant species	14
number of plant leaf disease classes	38
number of images	87908 rgb images

Fig. 5: Data Set

Dataset description of plant disease dataset with related information

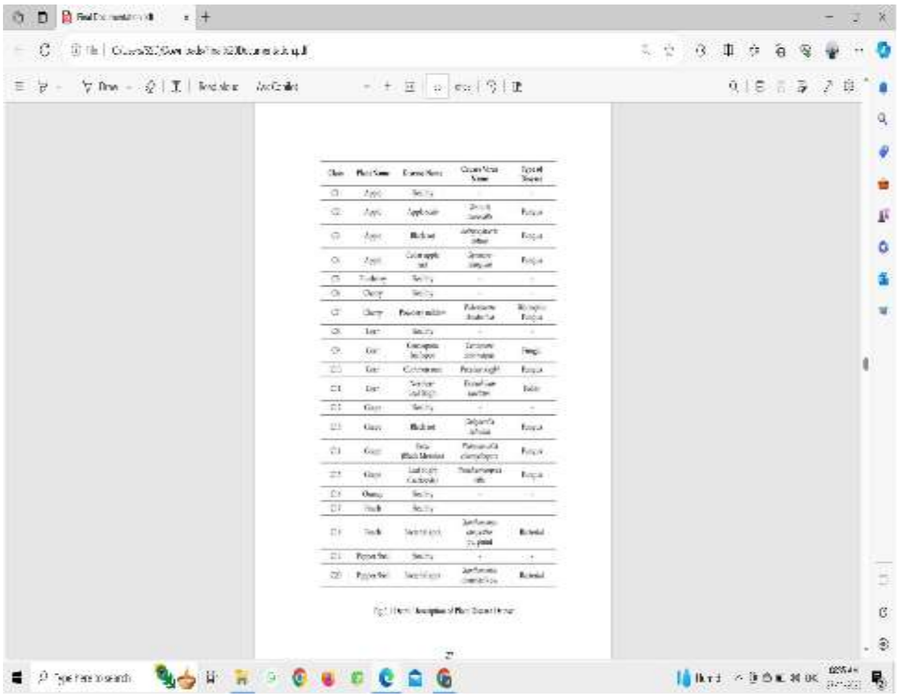

Fig. 6: Dataset description of plant disease

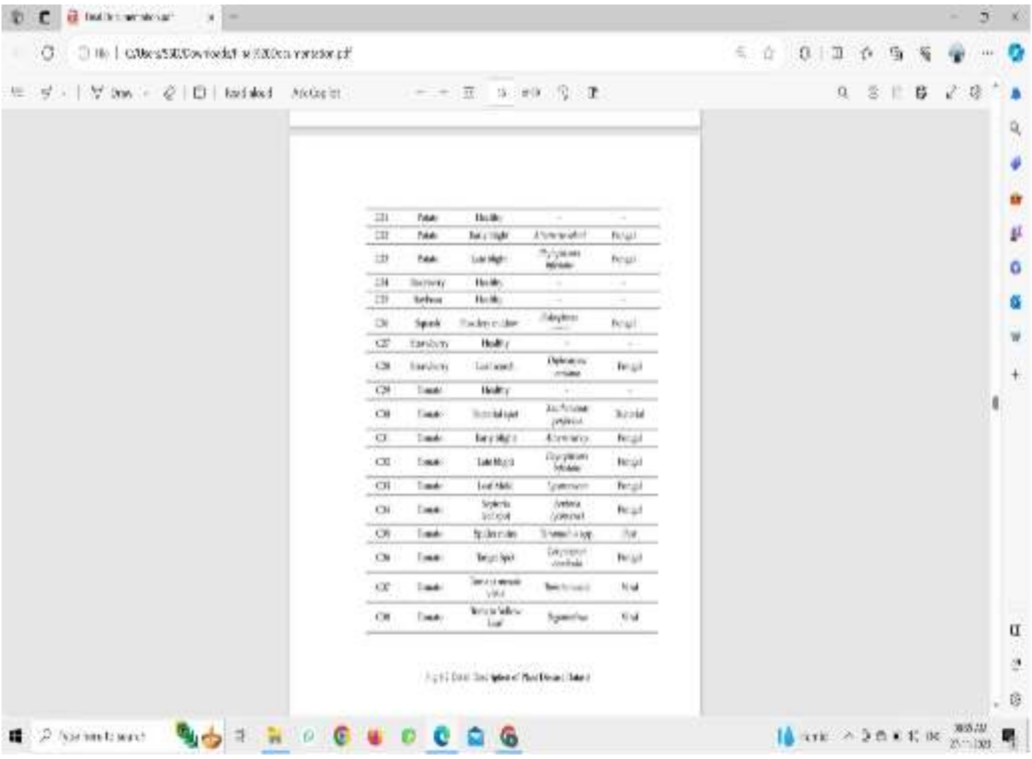

Fig. 7: Dataset description of plant disease

VI. TECHNOLOGY AND STACK OVERVIEW

The plant leaf disease detection system is developed using a combination of deep learning frameworks, programming tools, and image-processing techniques to achieve high accuracy and efficient performance.

A. Convolutional Neural Networks (CNNs)

- CNNs are the backbone of the model, capable of learning hierarchical image features for accurate disease classification.
- The proposed system uses a VGG19-based CNN architecture, optimized for detecting 38 plant disease classes.

B. VGG19 Architecture

- Composed of 16 convolutional layers, 5 max-pooling layers, and 3 fully connected layers.
- Global Average Pooling (GAP) integrated for reduced overfitting and improved interpretability.
- Trained using categorical cross-entropy loss and Adam optimizer.

C. Programming Language

- **Python** is used for implementation due to its vast ecosystem of libraries for deep learning and computer vision.

D. Libraries and Frameworks

- **TensorFlow / Keras**: For designing, training, and validating the CNN model.
- **OpenCV**: For image preprocessing tasks such as resizing, normalization, and augmentation.
- **NumPy & Pandas**: For numerical computations and dataset handling.
- **Matplotlib & Seaborn**: For plotting accuracy, loss curves, and visualizing model performance.

E. Dataset Management

- Dataset structured into training, validation, and testing sets.
- Images standardized to 224×224 pixels to align with VGG19 requirements.

F. Development Environment

- Implemented in Jupyter Notebook / Python IDE.
- Training conducted on systems with GPU support to accelerate computation.

VII. RESULTS

	precision	recall	f1-score	support
Apple___Apple_scab	1.00	0.99	0.99	504
Apple___Black_rot	1.00	1.00	1.00	497
Apple___Cedar_apple_rust	0.99	0.99	0.99	440
Apple___healthy	0.98	1.00	0.99	502
Blueberry___healthy	0.99	1.00	0.99	454
Cherry_(including_sour)___Powdery_mildew	1.00	0.99	1.00	421
Cherry_(including_sour)___healthy	1.00	0.99	0.99	456
Corn_(maize)___Cercospora_leaf_spot Gray_leaf_spot	0.99	0.85	0.91	410
Corn_(maize)___Common_rust_	1.00	0.99	0.99	477
Corn_(maize)___Northern_Leaf_Blight	0.88	1.00	0.93	477
Corn_(maize)___healthy	1.00	1.00	1.00	465
Grape___Black_rot	0.99	1.00	0.99	472
Grape___Esca_(Black_Measles)	1.00	0.99	0.99	480
Grape___Leaf_blight_(Isariopsis_Leaf_Spot)	1.00	0.99	1.00	430
Grape___healthy	1.00	1.00	1.00	423
Orange___Haunglongbing_(Citrus_greening)	1.00	0.99	1.00	503
Peach___Bacterial_spot	1.00	0.99	0.99	459
Peach___healthy	0.99	0.99	0.99	432
Pepper,_bell___Bacterial_spot	1.00	1.00	1.00	478
Pepper,_bell___healthy	0.99	0.99	0.99	497
Potato___Early_blight	1.00	1.00	1.00	485
Potato___Late_blight	0.97	1.00	0.98	485
Potato___healthy	0.97	0.99	0.98	456
Raspberry___healthy	0.99	1.00	1.00	445
Soybean___healthy	1.00	0.98	0.99	505
Squash___Powdery_mildew	1.00	1.00	1.00	434
Strawberry___Leaf_scorch	1.00	1.00	1.00	444
Strawberry___healthy	1.00	1.00	1.00	456
Tomato___Bacterial_spot	0.97	0.98	0.98	425
Tomato___Early_blight	0.97	0.95	0.96	480
Tomato___Late_blight	0.95	0.95	0.95	463
Tomato___Leaf_Mold	0.99	1.00	0.99	470
Tomato___Septoria_leaf_spot	0.99	0.96	0.97	436
Tomato___Spider_mites Two-spotted_spider_mite	0.94	1.00	0.96	435
Tomato___Target_Spot	0.98	0.93	0.95	457
Tomato___Tomato_Yellow_Leaf_Curl_Virus	0.99	0.98	0.99	490
Tomato___Tomato_mosaic_virus	0.98	1.00	0.99	448
Tomato___healthy	0.99	1.00	0.99	481
accuracy			0.99	17572
macro avg	0.99	0.98	0.98	17572
weighted avg	0.99	0.99	0.99	17572

Fig. 8: Performance Analysis

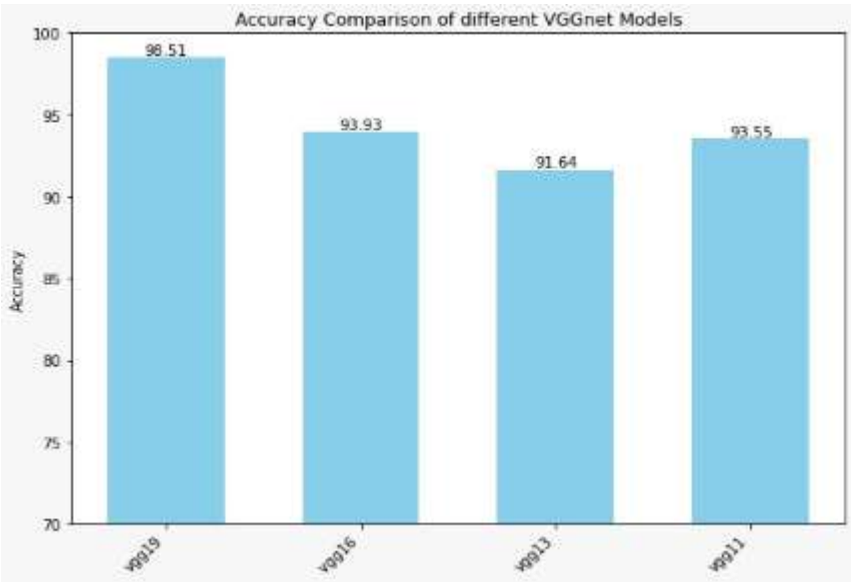

Fig. 9: Accuracy comparision of different VGGNet models

VIII. DISCUSSION

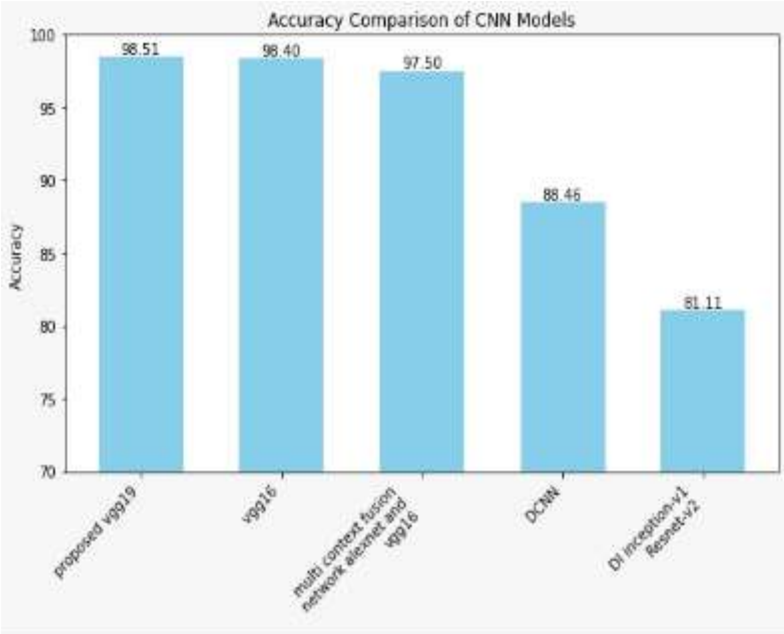

Fig. 10: Comparison of Proposed VGG19 with Existing CNN models

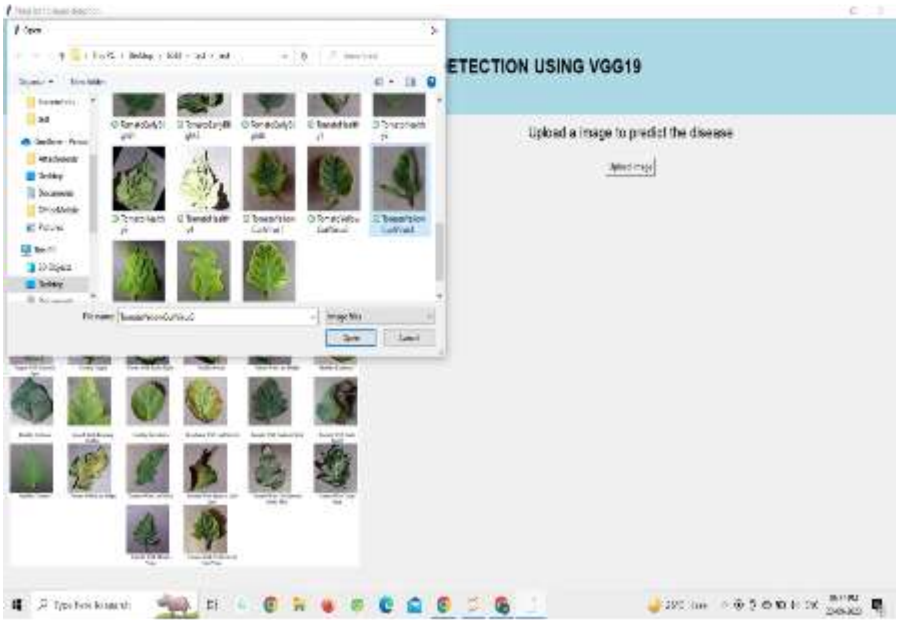

Fig. 11: Upload an image to detect the disease

Fig. 12: Tomato Yellow Leaf Curl Virus

Fig. 13: Cedar Apple Rust

Fig. 14: Potato Early Blight

VIII. DISCUSSION

The proposed CNN-based plant leaf disease detection system effectively addresses the challenge of early identification and classification of plant diseases, which directly impacts crop yield and farmers' economic stability. The model was trained on a diverse dataset of 38 disease classes and healthy leaves, achieving high accuracy (98.51%) with the VGG19 architecture.

The approach successfully utilized key deep learning techniques such as data augmentation, normalization, and custom CNN architecture design to enhance feature extraction and reduce overfitting. Evaluation metrics including precision, recall, F1-score, and accuracy validated the reliability of the system across multiple performance aspects.

Compared to traditional manual inspection, the system demonstrated faster, more accurate, and scalable disease detection. However, limitations such as dependence on image quality, lack of real-time field validation, and dataset diversity were noted. Future improvements could include incorporating drone-based aerial imagery, larger datasets, and deployment in mobile applications for real-time diagnosis.

Overall, the potential of deep learning techniques like VGG19 in agricultural applications, offering a practical and effective solution to enhance plant disease management and improve crop productivity.

IX. CONCLUSION

The CNN-based plant leaf disease detection system demonstrates the effective application of deep learning techniques for agricultural problem-solving. By leveraging a custom VGG19 architecture, the model achieved a high accuracy of **98.51%**, confirming its capability to identify and classify plant diseases across multiple categories.

Compared to traditional manual inspection, the system provides faster diagnosis, reduces human error, and ensures consistency in disease detection. Incorporating preprocessing methods like data augmentation and normalization further enhanced the model's generalization to diverse image samples.

While the results are promising, future work could focus on expanding the dataset, integrating real-time field image testing, and developing mobile or drone-assisted applications to make the system more practical and accessible for farmers.

In summary, this project provides a reliable and scalable framework for plant disease detection, helping improve crop productivity and supporting precision agriculture.

REFERENCES

1.Demilie, W. B. (2024). Plant disease detection and classification techniques: a comparative study of the performances. Journal of Big Data, 11(5).

2. Bhargava, A., Shukla, A., Goswami, O. P., Alsharif, M. H., Uthansakul, P., & Uthansakul, M. (2024). Plant leaf disease detection, classification, and diagnosis using computer vision and artificial intelligence: A review. IEEE Access, 12, 37443–37454.

3. Li, L., Zhang, S., & Wang, B. (2021). Plant disease detection and classification by deep learning—A review. IEEE Access, 9, 56683–56694.

4. Aboelenin, S., Elbasheer, F. A., Eltoukhy, M. M., El-Hady, W. M., & Hosny, K. M. (2025). A hybrid framework for plant leaf disease detection and classification using convolutional neural networks and vision transformer. Complex & Intelligent Systems, 11, 142.

5. Krishna, M. S., Machado, P., Otuka, R. I., Yahaya, S. W., Neves dos Santos, F., & Ihianle, I. K. (2025). Plant leaf disease detection using deep learning: A multi-dataset approach. J, 8(1), 4.

6. Gavhale, K. R., & Gawande, U. (2014). An overview of the research on plant leaves disease detection using image processing techniques. IOSR Journal of Computer Engineering, 16(1), 10–16.

7. Ingole, K. (2021). Review on leaf disease detection using image processing techniques. International Research Journal of Engineering and Technology (IRJET), 4(4), 2044 2049.