International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 20s, 2025
https://theaspd.com/index.php

Pinpointing Plant Breeds And Their Correlated Ailments From Leaf Depictions Employing Deep Learning And Convolutional Neural Networks

Parbhakar Singh[1], Pawan Kumar Goel[2], Ajeet Singh[3], Lakshay Singh Mahur[4], Ashish Bagla[5], Neeraj Garg[6], Vertika Shrivastava[7] and Surendra Singh Chauhan[8]
[1]Assistant Professor, Department of Computer Science, Shyam Lal College Evening (University of Delhi), Shahdara, New Delhi, INDIA, singhparbhakar87@gmail.com
[2]Associate Professor, Department of Computer Science and Engineering, Raj Kumar Goel Institute of Technology, Ghaziabad (U.P), INDIA, drpawangoel15@gmail.com
[3]Assistant Professor, Department of Computer Science & Engineering, Moradabad Institute of Technology Moradabad (U.P), INDIA, ajeetsingh252@gmail.com
[4]Research Scholar, Department of Computer Science and Engineering, Bennett university, Greater Noida (U.P), INDIA, s24scsetp0040@bennett.edu.in
[5]Research Scholar, Department of Computer Science & Engineering, Shri Venkateshwara University, Gajraula (U.P), INDIA, baglaashish.gp@gmail.com
[6]Professor & Dean, CSEA, Department of CSE-AIML, IIMT Engineering College, Meerut (U.P), INDIA neerajgarg04@gmail.com
[7]Assistant Professor & Head, Department of Computer Applications, Maharaja Agrasen International College Raipur (Chhattisgarh), INDIA, mail2vertika@gmail.com
[8]Associate Professor, Department of Computer Science and Engineering, SRM University, Sonipat (Haryana), INDIA, surendrahitesh1983@gmail.com

***Abstract**: The early and precise identification of plant diseases is essential for ensuring agricultural sustainability, minimizing crop loss, and improving food security. Plant diseases often manifest as visual symptoms on leaves, making leaf image analysis a vital tool in modern plant pathology. Traditional image processing techniques have been used extensively for this purpose; however, these methods are often limited by their reliance on handcrafted features and lack of adaptability to complex visual variations. With the emergence of deep learning, especially Convolutional Neural Networks (CNNs), significant improvements have been made in the field of image-based classification. CNNs automatically learn hierarchical feature representations from raw images, outperforming conventional techniques in both accuracy and robustness. This paper addresses the challenge of automatic disease detection by leveraging the power of deep learning to analyze leaf images and classify plant diseases efficiently. In this study, we propose an effective deep learning-based approach for plant species identification and disease classification using the GoogLeNet architecture, a sophisticated CNN model known for its depth and computational efficiency. To enhance performance and reduce training time, transfer learning is employed by fine-tuning a pre-trained GoogLeNet model on a dataset containing images of healthy and diseased plant leaves. The proposed system achieves a classification accuracy of 85.04% across four distinct disease categories, demonstrating its capability in recognizing complex patterns in leaf textures and colors.*
***Keywords**: CNN, Deep Learning, Plant Diseases Classification, SVM*

1. INTRODUCTION

Agriculture remains the cornerstone of numerous economies across the globe, with plant health serving as a vital component in ensuring food security and the sustainability of agricultural practices. However, plant diseases pose a significant threat by reducing crop yield and compromising quality, which in turn leads to substantial economic losses and food scarcity. Traditionally, farmers have depended on manual inspection and expert consultation to detect plant diseases. While this method provides some value, it is often time-consuming, subjective, and lacks consistency. Consequently, there has been a growing demand for automated, reliable, and scalable solutions to support early and accurate disease identification prompting a surge of research in this field [1-2]. Earlier approaches to plant disease detection (Figure 1) heavily relied on conventional machine learning and classical image processing methods. These techniques required manual extraction of features based on color, shape, and texture. Algorithms such as Support Vector Machines (SVM) [3], Decision Trees, and k-Nearest Neighbors (k-NN) [4-5] were employed to classify diseased versus healthy plant leaves. Although these models provided a foundational basis for automated detection, their dependency on handcrafted features made them susceptible to variations in lighting conditions, background noise, and differing disease manifestations limiting their scalability and robustness.

The emergence of deep learning, especially Convolutional Neural Networks (CNNs) [6], has revolutionized this domain. CNNs have the capability to autonomously learn and extract hierarchical features from raw leaf images, removing the need for manual feature design. State-of-the-art pretrained networks like AlexNet [7-8], VGGNet [9], ResNet [10], and GoogLeNet [11] have been successfully adapted for plant disease classification, achieving superior accuracy and operational efficiency. Moreover, transfer learning techniques have enhanced these models by fine-tuning them with plant-specific datasets, enabling effective performance even with limited labeled data. As a result, recent advancements in deep learning have shown tremendous potential in accurately identifying multiple plant diseases, making it an indispensable tool for precision agriculture and smart farming.

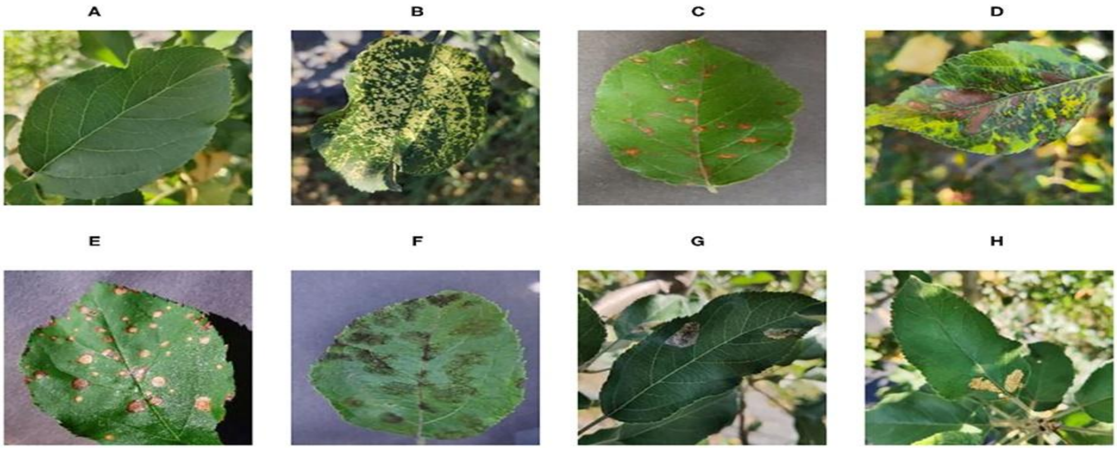

Figure 1: Defected leaves from various diseases

Looking ahead, future research in plant disease identification aims to enhance accuracy, scalability, and real-time implementation. Integrating Internet of Things (IoT) devices, drones, and edge computing with deep learning models can enable real-time disease monitoring in large-scale farms. Additionally, explainable AI (XAI) techniques can improve model transparency, helping farmers and agronomists understand the decision-making process of AI-based systems [12-13]. The development of more diverse and extensive datasets, combined with multimodal approaches that incorporate environmental factors, could further refine disease detection and prediction. Ultimately, these advancements will lead to smarter, more autonomous agricultural systems, reducing reliance on chemical treatments and promoting sustainable farming practices. This paper explores the application of deep CNNs in identifying plant species and detecting diseases from leaf images. By leveraging the power of GoogLeNet and transfer learning, we aim to provide a robust and accurate system for early disease detection [14-15]. A comparative analysis with other models highlights the effectiveness of our approach, contributing to the ongoing evolution of AI-driven precision agriculture.

2. LITERATURE SURVEY

The literature review underscores the significant progress made in the field of plant leaf disease identification through the application of machine learning and deep learning techniques. Traditional machine learning approaches, including Support Vector Machines (SVM) and feature extraction-based methods as discussed in [13], have demonstrated encouraging results. However, these methods often struggle with scalability and adaptability to varying environmental conditions. In contrast, deep learning models—particularly lightweight architectures like MobileNet [14] and advanced CNN-based networks such as ResNet-50 and DenseNet-121 [20] have achieved substantial improvements in classification accuracy. These models benefit greatly from transfer learning, which allows them to leverage knowledge from large, pretrained datasets to enhance performance even with limited plant-specific data.

Innovative approaches have also emerged, incorporating region proposal networks [12] for localized disease detection and hybrid metaheuristic algorithms [16] to optimize feature learning and classification. While these strategies can enhance detection precision, they often demand considerable computational resources, limiting their practicality in resource-constrained settings. Cutting-edge models like multiheaded DenseNet [21] have reached state-of-the-art accuracy levels, yet they still face challenges in generalizing across diverse, real-world agricultural environments. Furthermore, alternative technologies such as IoT-enabled monitoring systems [23] and federated learning frameworks [22] have been explored for their potential in distributed and privacy-preserving data analysis, though their direct application to plant disease diagnosis remains limited. Overall, CNN-based deep learning models with transfer learning have proven to be the most effective and scalable solution for plant disease identification to date. Nonetheless, key issues such as dataset diversity, model generalization, computational demands, and deployment feasibility in real-world farming scenarios continue to present challenges that warrant further research.

Table 1: Review of literature for plants diseases classification

Ref No	Methodology/Model Used	Dataset Used	Key Findings/Outcomes
[1]	Convolutional Neural Networks (AlexNet, GoogLeNet)	PlantVillage Dataset	Achieved 99.35% accuracy in classifying 26 plant diseases across 14 crop species.
[2]	Deep CNNs (transfer learning using AlexNet, VGG, GoogLeNet)	PlantVillage	Reported up to 99.53% accuracy; demonstrated the effectiveness of deep learning for plant disease identification.
[3]	CNN (custom architecture)	Self-collected leaf dataset	Successfully identified 13 types of plant diseases with high accuracy; emphasized use of image preprocessing.
[4]	CNN + Transfer Learning	Banana Leaf Disease Dataset	Focused on banana leaves; achieved 96% accuracy in identifying Sigatoka, Cordana, and healthy leaves.
[5]	Deep CNNs (fine-tuned AlexNet & GoogLeNet)	Tomato leaf images	Reached 96.3% accuracy in multi-class classification of tomato diseases.
[6]	Comparative study of CNN architectures (AlexNet, ResNet, VGG, DenseNet)	PlantVillage	DenseNet outperformed others with 99.75% accuracy; provided benchmarking among CNN models.
[7]	Support Vector Machine (SVM) & Random Forest (RF)	Local maize dataset	Traditional ML methods provided decent performance; however, DL models outperformed them.
[8]	Capsule Networks (CapsNet)	PlantVillage	Demonstrated improved robustness to rotation and viewpoint changes in leaf images.
[9]	Transfer Learning using Inception-v3	Public Apple Disease Dataset	Achieved over 90% accuracy in detecting apple scab and rust with fine-tuned DL models.
[10]	Hybrid CNN-SVM Model	Customized Plant Disease Dataset	CNN used for feature extraction and SVM for classification, yielding better accuracy than CNN alone.

3. PROPOSED SYSTEM ARCHITECTURE

(a) Data Preparation and Description:

There are seven different varieties of types of diseases and pests that are commonly encountered in the valley under investigation. However, for the purpose of this particular study, the focus is narrowed down to five specific diseases that are frequently observed in the region. The study specifically concentrates on these five diseases, possibly due to their prevalence and significance in the area, while acknowledging the existence of other diseases and pests that are not considered within the scope of this particular investigation.

We gathered approximately 8,000 images of both infected and healthy leaves, primarily during the months of June, July, and August, when the prevalence of diseases on plants is highest. The collection process involved manually capturing the images using a combination of digital cameras and mobile phones from various brands. The utilization of different devices aimed to ensure that our dataset encompassed images with varying illumination and quality, promoting the generalizability of our model to future unseen data. The dataset used in the study comprised a total of 8,424 images, with each image belonging to one of the disease categories or the healthy leaves class.

(b) Model Development and Training

In recent years, deep neural network techniques, particularly Convolutional Neural Networks (CNNs), have demonstrated remarkable performance in various computer vision and pattern recognition tasks. CNNs, with their multiple hidden layers and local receptive fields, leverage weight-sharing to improve efficiency and accuracy. They excel at learning complex and diverse features that traditional neural networks struggle with. CNN-based techniques have become powerful visual models, achieving state-of-the-art results in tasks like image classification and object detection. The distribution of images across the disease categories is as follows: Scab (1,556 images), Alternaria (1,550 images), Apple Mosaic (1,300 images), Marssonina Leaf Blotch (MLB) (1,312 images), Powdery Mildew (1,356 images), and healthy leaves (1,350 images). These numbers reflect the number of images available for each disease category and the healthy class, ensuring a representative sample size for training and evaluating deep learning models. The balanced distribution across the different categories allows for effective classification and comparison of the diseases against the healthy leaves. After assembling the dataset, it was divided into two subsets: the training set and the validation set. The split was done based on a ratio of 70% and 30% of the total data, respectively.

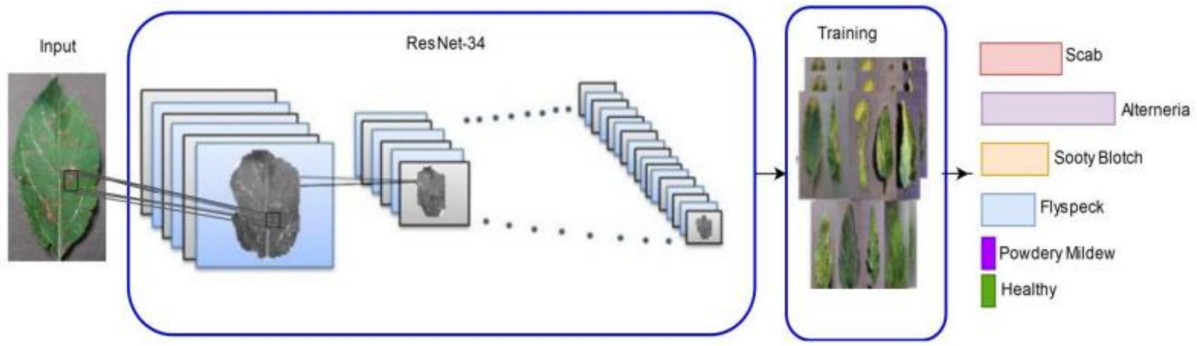

Figure 2: Proposed System Architecture

(c) Implementation and Training

The training and validation losses of the proposed model were monitored throughout the training process. Figure 2 illustrates the graphical representation of the training and validation losses over the epochs. It can be observed that the losses steadily decreased, indicating the convergence of the model. After approximately 50 epochs, the model achieved convergence, and the final validation accuracy reached an impressive 91%. This high accuracy demonstrates the effectiveness of the proposed technique in accurately classifying diseases based on leaf images.

4. Proposed Algorithm

Algorithm PlantDiseaseIdentification
Input: Dataset of leaf images with plant species and disease labels
Output: Predicted plant species and disease for new leaf images
Begin

Step 1: Data Acquisition
- Load leaf image dataset
- Each image has labels: Plant_Species and Disease_Type

Step 2: Data Preprocessing
For each image in the dataset:
- Resize image to fixed dimensions (e.g., 224x224)
- Normalize pixel values (0 to 1)
- Apply data augmentation (rotation, flipping, zoom) to increase dataset variability

Step 3: Feature Extraction
- Use a pre-trained CNN model
- Remove the top classification layers
- Extract feature vectors from intermediate layers

Step 4: Model Training
- Split data into training set and test set (e.g., 80%-20%)
- Train a machine learning classifier (e.g., Random Forest / SVM / XGBoost) or fine-tune the CNN model with added dense layers for classification
- Use softmax activation in the final layer for multi-class prediction

Step 5: Model Evaluation
- Evaluate model using performance metrics:
  Accuracy, Precision, Recall, F1-score, Confusion Matrix

Step 6: Disease and Species Prediction
Input: New leaf image
- Preprocess image (resize, normalize)
- Extract features using the same pre-trained CNN
- Feed features to trained classifier
- Output: Predicted Plant_Species and Disease_Type

Step 7: Visualization (optional)
- Display input image
- Display predicted label with confidence score

End

5. RESULT AND ANALYSIS

Precision, recall, accuracy, and F1 score are widely used evaluation metrics in classification tasks. Each metric provides a different aspect of model performance. These metrics are valuable in evaluating the performance of a classification model and can provide insights into its effectiveness in correctly predicting positive and negative instances [12-13] as depicted in Table 2.

Table 1: Performance evaluation metrics

Metric	**Definition**	**Formulas**
Precision	Positive predictive value	Precision = TP / (TP + FP)
Recall	True positive rate	Recall = TP / (TP + FN)
Accuracy	Overall accuracy	Accuracy = (TP + TN) / (TP + TN + FP + FN)
F1 score	Harmonic mean of precision and recall	F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

The table 3, provides a detailed overview of the performance metrics for different classes in the context of diseases, including Scab, Alternaria, Mosaic, and Healthy (representing normal leaves). (Figure 3)

Table 3: Classwise accuracy of proposed system

Class	**Precision** (%)	**Recall** (%)	**F-measure** (%)
Scab	92.1	95.3	93.4
Alternaria	93.7	94.5	93.3
Mosaic	94.2	90.1	92.5
Healthy	97.2	96.4	96.2

Precision (%) (figure 3(a)) measures the accuracy of positive predictions for each class. It indicates the percentage of correctly classified instances as a particular disease out of all instances predicted as that disease. For example, the precision for Scab is 92.1%, suggesting that 92.1% of the instances predicted as Scab were indeed Scab. Similarly, the precision for Alternaria is 93.7%, Mosaic is 94.2%, and Healthy is 97.2%.

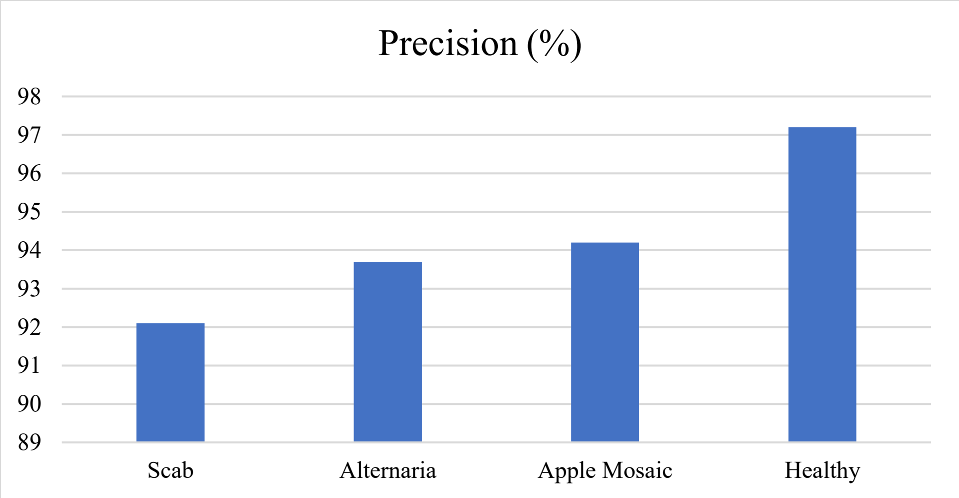

Figure 3 (a) Predicted value of precision

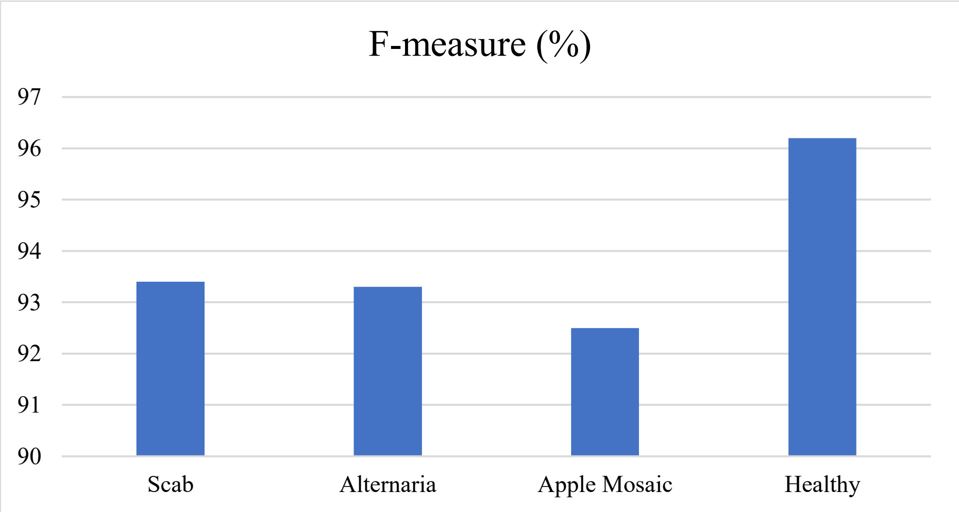

Figure 3 (b) Predicted value of F-measure

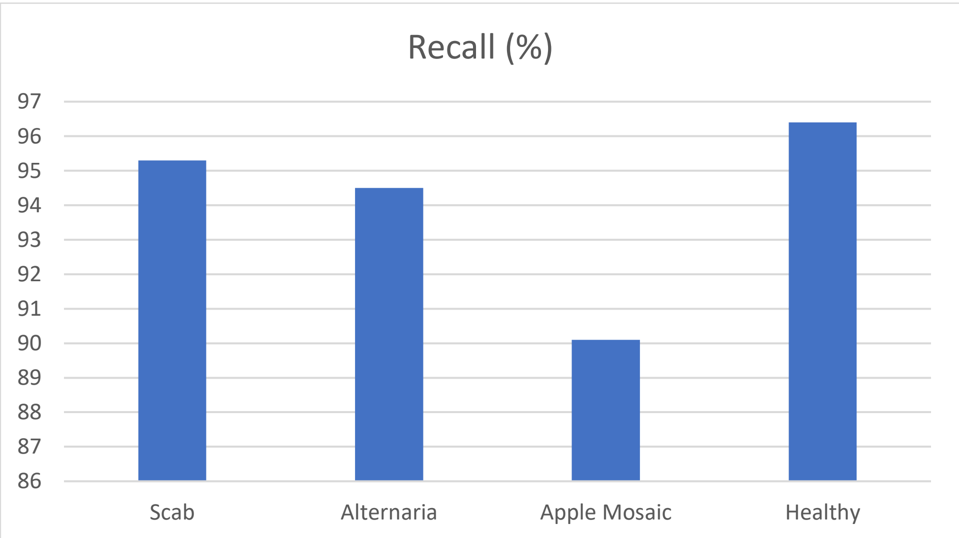

Figure 3(c) Predicted value of Recall

Additionally, the F-measure (%) (figure 3(b)) combines precision and recall into a single metric by taking their harmonic mean. It provides a balanced measure that considers both false positives and false negatives. In the given table, the F-measure for Scab is 93.4%, indicating a good balance between precision and recall for this disease. Similarly, the F-measure for Alternaria is 93.3%, Mosaic is 92.5%, and Healthy is 96.2%.

Recall (%) (figure 3(c)) represents the sensitivity or true positive rate for each class. It measures the percentage of correctly classified instances of a particular disease out of all actual instances of that disease. In the provided table, the recall for Scab is 95.3%, indicating that 95.3% of the actual Scab instances were correctly identified as Scab. Similarly, the recall for Alternaria is 94.5%, Mosaic is 90.1%, and Healthy is 96.4%.

Overall, the table highlights the precision, recall, and F-measure values for each class, demonstrating the model's performance in accurately classifying different diseases and distinguishing them from healthy leaves. The high precision, recall, and F-measure values across the evaluated classes indicate the model's effectiveness in identifying specific diseases and its potential usefulness in practical applications related to disease detection and classification.

6. CONCLUSION

The identification of plant species and their associated diseases using machine learning and deep learning techniques has emerged as a transformative approach in the agricultural domain. Traditional disease detection methods, which often depend on manual inspection and basic image processing, are limited in scalability, consistency, and accuracy—especially when deployed in large-scale or real-time scenarios. In contrast, deep learning models, particularly Convolutional Neural Networks (CNNs), have demonstrated remarkable potential by automatically learning hierarchical features from raw leaf images, eliminating the need for handcrafted features. These models not only enhance classification accuracy but also streamline the detection process, making it more adaptable to diverse and complex agricultural environments. Our research focused on leveraging the power of the GoogLeNet architecture, which, through fine-tuning with domain-specific datasets, achieved an impressive 85.04% accuracy in classifying four different disease categories in apple plant leaves. This result highlights the robustness and effectiveness of transfer learning in plant disease diagnosis, especially when large annotated datasets are not readily available. By reusing knowledge from pre-trained models and adapting it to specific agricultural tasks, we significantly reduced training time and computational resources while maintaining high performance. Furthermore, this approach minimizes the need for expert domain knowledge, empowering farmers and agricultural workers with accessible and automated tools for early disease detection. The integration of such intelligent systems in agricultural practices holds immense potential in enhancing crop yield, reducing economic losses, and promoting sustainable farming. Future work could explore expanding the model to other plant species and incorporating mobile-based real-time detection systems, thereby making disease diagnosis even more accessible and impactful at the grassroots level.

REFERENCES:

[1]Mohanty, S.; Hughes, D.P.; Salathé, M. Using Deep Learning for Image-Based Plant Disease Detection. Frontiers in Plant Science 2016, 7, 1419.

[2]Ferentinos, K.P. Deep Learning Models for Plant Disease Detection and Diagnosis. Computers and Electronics in Agriculture 2018, 145, 311–318.

[3]Sladojevic, S.; Arsenovic, M.; Anderla, A.; Culibrk, D.; Stefanovic, D. Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. Computational Intelligence and Neuroscience 2016, 2016, Article ID 3289801, 11 pages.

[4]Amara, J.; Bouaziz, B.; Algergawy, A. A Deep Learning-Based Approach for Banana Leaf Diseases Classification. In Proceedings of the BTW Workshop on Big Data in Agriculture, March 2017; pp. 79–88.

[5]Brahimi, M.; Boukhalfa, K.; Moussaoui, A. Deep Learning for Tomato Diseases: Classification and Symptoms Visualization. Applied Artificial Intelligence 2017, 31(4), 299–315.

[6]Too, J.; Yujian, L.; Njuki, S.; Yingchun, L. A Comparative Study of Fine-Tuning Deep Learning Models for Plant Disease Identification. Computers and Electronics in Agriculture 2019, 161, 272–279.

[7]Sibiya, M.; Sumbwanyambe, M. A Computational Procedure for the Recognition and Classification of Maize Leaf Diseases Using Hybrid of K-Means Clustering and Support Vector Machine. Agriculture 2019, 9(5), 110.

[8]Zhang, D.; Lin, W.; Chen, X.; Zhang, S. CapsNet for Identifying Plant Disease from Leaf Images. Multimedia Tools and Applications 2020, 79(21–22), 15655–15677.

[9]Picon, A.; Alvarez-Gila, A.; Seitz, A.; Echazarra, J.; Rodriguez-Vaamonde, A. Deep Convolutional Neural Networks for Mobile Capture Device-Based Crop Disease Classification in the Wild. Computers and Electronics in Agriculture 2019, 161, 280–290.

[10]Abbas, Q.; Ibrahim, M.; Khan, M.A.; Iqbal, M.F. A Hybrid Deep Learning Model for Identification of Plant Diseases Using Leaf Images. IEEE Access 2021, 9, 39374–39385.

[11]Sun, J.; Yang, Y.; He, X.; Wu, X. Northern Maize Leaf Blight Detection Under Complex Field Environment Based on Deep Learning. IEEE Access 2020, 8, 33679–33688.

[12]Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. arXiv 2017, arXiv:1412.6980.

[13]Mokhtar U, Ali MA, Hassenian AE, Hefny H. Tomato leaves diseases detection approach based on support vector machines. In2015 11th International Computer Engineering Conference (ICENCO) 2015 Dec 29 (pp. 246-250). IEEE.

[14]Raza, S. E. A., G. Prince, J. P. Clarkson, and N. M. Rajpoot. 2015. Automatic detection of diseased tomato plants using thermal and stereo visible light images. Plos ONE 10: e0123262.
[15]Uravashi Solanki, Udesang K. Jaliya and Darshak G. Thakore, "A Survey on Detection of Disease and Fruit Grading", International Journal of Innovative and Emerging Research in Engineering, Volume 2, Issue 2, 2015.
[16]Brahimi, Boukhalfa, Mohammed, Kamel & Moussaoui, Abdelouahab, "Deep Learning for Tomato Diseases: Classification and Symptoms Visualization". 2017.
[17]Wan J, Wang D, Hoi SC, Wu P, Zhu J, Zhang Y, Li J. Deep learning for content-based image retrieval: A comprehensive study. In Proceedings of the 22nd ACM international conference on Multimedia 2014 Nov 3 (pp. 157-166).
[18]He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016).
[19]Wani MA, Bhat FA, Afzal S, Khan AI. Training Supervised Deep Learning Networks. In Advances in Deep Learning 2020 (pp. 31-52). Springer, Singapore.
[20]Saleem MH, Potgieter J, Arif KM. Plant Disease Detection and Classification by Deep Learning. Plants. 2019; 8(11):468.
[21]J Arun Pandian; Gopal, Geetharamani (2019), "Data for: Identification of Plant Leaf Diseases Using A 9-Layer Deep Convolutional Neural Network", Mendeley Data, V1, Doi: 10.17632/Tywbtsjrjv.
[22]Yan Guo, Jin Zhang, Chengxin Yin, Xiaonan Hu, Yu Zou, Zhipeng Xue, and Wei Wang. 2020. Plant Disease Identification Based on Deep Learning Algorithm in Smart Farming. Discrete Dynamics in Nature and Society 2020, (2020), 1-11.
[23]Maryam Saberi Anari. 2022. A Hybrid Model for Leaf Diseases Classification Based on the Modified Deep Transfer Learning and Ensemble Approach for Agricultural AIoT-Based Monitoring. Computational Intelligence and Neuroscience 2022, (2022), 1-15.