PLANT LEAF DISEASE IDENTIFICATION SYSTEM USING MACHINE LEARNING

[1]V.Vaitheeshwaran, [2]Vidya, [3]Arul, [4]Rekansh Rao
[1,2,3,4]Department of Computer Science and Engineering (AI&ML), St.Peters Engineering College, Maissammaguda, Dhulapally, Kompally, Medchal-500100,
vidyarajasekaran@stpetershyd.com

Abstract
Plant diseases are a leading cause of loss of food security worldwide and thus the management of diseases at the initial level is very important. Current identification approaches involve the use of manual inspection which is labor intensive and can be inaccurate or the traditional machine learning which has limitations of a lot of feature engineering. The proposed Plant Disease Identification System solves these problems by using Convolutional Neural Networks (CNNs) which are efficient in feature extraction and image classification. The system increases the accuracy of detection in a three stage process; the first stage is image acquisition, the second stage is image preprocessing which aims at removing noise and enhancing features and the last stage is classification where the image is classified using a CNN that has been trained on a large database of images of diseased plants. As compared to the existing techniques, this system provides more accurate and efficient results and is a convenient tool which can be used by farmers and agricultural experts to control the health of crops.
Keywords: Plant diseases, Food security, CNNs (Convolutional Neural Networks), Feature Extraction, Image Classification, preprocessing, Noise removal, Disease management.

1. Introduction

The health and productivity of agricultural crops are very vital for the global food supply and for supporting the ever-growing population. However, plant diseases remain a serious challenge that cause reduced yields and significant economic losses worldwide. Early-stage management of plant diseases helps prevent their spread and sustains agricultural practices. Hence, the accurate identification of plant diseases has become a vital task for farmers, agricultural experts, and policymakers. Traditionally, it had much to do with observing and visual inspection by farmers or other agricultural experts. Such processes are laborious and time consuming, as well as at times subjective and erroneous. More complicated is that disease symptoms cannot be separated clearly from abiotic symptoms stemming from nutrient deficiency or certain environmental conditions. In sum, such methods do not scale for large agricultural systems, especially large-scale farms or recently emergent diseases. With the advent of machine learning and computer vision, the process of detecting plant diseases has been fully transformed by effective and scalable solutions. In the past, classical ML methods such as SVMs and kNN were also applied to perform classification operations. For instance, stability and robustness characteristics of SVMs make them a very good classifier in plant disease classification problems, as explained in [1] by Mountrakis et al. Similarly, Thanh Noi et al. [2] compared the efficiency of SVM, k-NN, and Random Forest algorithms for land cover classification where the drawback of manual feature engineering in these methods is exposed. Though much advancement has been achieved by the traditional techniques of ML, it often fails to work well on large datasets and mainly relies on handcrafted feature extraction. The rise of deep learning particularly Convolutional Neural Network (CNNs) has dramatically upgraded the task of image classifications, especially in plant disease. CNNs automatically extract their features that increase the possibility of accurate and efficient results in the classification. ImageNet, a pioneering research work done by Alex Krizhevsky et al. [3], did well to prove the utilization of CNNs in image classifications. The high resolution in remote sensing images is a successful application of CNNs by Maher Ibrahim Sameen et al. [4], as well as accuracy improvement.By designing the traditional methods in deep learning systems in regard to image data processing at large scale and finding visual pattern difficult to detect with naked-eye vision, it helps surmount the limitations of the current methods. This paper has been organised as follows: Section 2 covers literature work of the proposed work. Section 3 includes the architecture of proposed work. Section 4 provides description about the benchmark dataset and performance metrics. Section 5 discusses about the experimental result and analysis. Finally conclusions are given in Section 6.

2. Literature Survey

Transfer Learning in Plant Disease Detection (2021): Jasinski et al. uses depth-separable convolution- based CNNs. The computation costs reduce here. Inception-based architecture with an accuracy of more than 99 percent is also achieved, which shows how deep learning is preferred over the traditional ways of plant disease diagnosis.Small CNN Models for Plant Disease Detection (2024): Recently, the research work published in IEEE reflects the feasibility of using small-scale CNNs to perform high accuracy in the task of plant disease detection using lower resource utilization. This work has underlined the feasibility of using it in real-time manners.Multi-Stage Disease Detection (2023): Authors have proposed a multistage framework, which encompasses pre-processing, feature extraction, and classification using the CNN. Authors have increased the accuracy by applying the data augmentation flipping and rotation techniques. Spectral-Spatial CNNs for Disease Classification (2020): Sameen et al. have proposed the CNN architecture for disease identification with very high-resolution imagery and integrated spectral-spatial information. It was more effective in some cases of crops.MobileNetV2 for Real-time Detection (2022): Taylor et al. tweaked the MobileNetV2 architecture to be sensitive and specific to plant disease, adapted for mobile deployment. This model was above 97% accurate but not so computationally expensive. Consequently, it can be deployed on mobiles in resource-scarce environments. Comparing Traditional and Deep Learning Models (2021): Thanh Noi et al. compared SVM, k-NN, and CNN-based models for plant disease detection. CNN models outperformed traditional classifiers in terms of accuracy as well as scalability. Transfer Learning with Pretrained Models (2019): Kruithof et al. utilized pretrained DCNNs, including ResNet and VGG, to classify diseases. These authors fine-tuned the models on smaller datasets to achieve the state-of-the-art performance.Object- Based Disease Localization (2021): Long et al. proposed a framework that combined the network of region proposal with CNN classifiers to improve real-time disease detection.

3. Proposed Work

The system uses a modular design with four modules: the Data Collection Module, which takes images from databases and also from live field data; a Preprocessing Module, enhancing the images to normalize pixel intensities, removing noise; a Feature Extraction Module automatically identifying discriminative features, using CNNs and Classification Module using fully connected layers and softmax activation functions. The process begins with acquiring a dataset, collecting labeled images that are diverse and represent various classes of plant diseases. In preprocessing, images are normalized, resized, and augmented for improving quality and variability. Feature extraction includes convolutional layers to build spatial hierarchies and dimensionality reduction by using pooling layers. Classification is done by the fully connected layers and softmax activation, which predicts disease categories.

Architecture:

It has an input layer where RGB images are resized to 224x224 pixels. There are convolutional layers to extract hierarchical spatial features; batch normalization and ReLU to stabilize training and introduce
non-linearity; pooling layers for computational efficiency; and fully connected layers to map features to classes. The output layer uses a softmax activation function to make multi-class disease prediction. Such designs guarantee efficient and scalable recognition of plant diseases with significant precision.

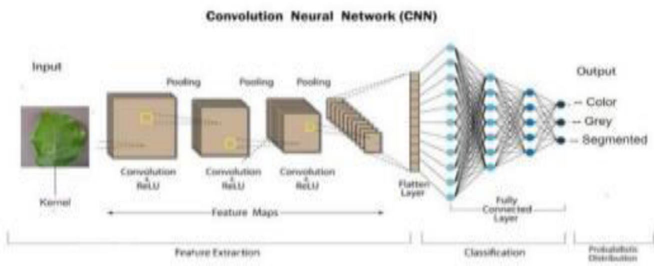

Figure.1. CNN structural model

Dataset Description:

The diversity of our dataset is extremely high. It contains high-resolution images of healthy and diseased plant leaves from online repositories and open-source platforms such as PlantVillage. It uses thousands of labeled samples, containing several plant species and diseases. This variability ensures robust model training to achieve reliable and generalizable results across a wide range of plant diseases. Data preparation is an important workflow that has several important steps. Data cleaning removes duplicates as well as irrelevant images by cleaning up the dataset. Annotated data tag images based on their disease category so that they can be properly trained. The last is augmenting the data set through flipping, rotation, scaling, and colour enhancements in order to increase artificially the dataset of images, thus avoiding the problem of overfitting in training the model.

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research Paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 10, 2022

Figure.2. Tomato(Bacterial_spot) Grape(Black_Measles) Potato(Late_blight)

The process begins with acquiring a dataset, collecting labeled images that are diverse and represent various classes of plant diseases. In preprocessing, images are normalized, resized, and augmented for improving quality and variability. Feature extraction includes convolutional layers to build spatial hierarchies and dimensionality reduction by using pooling layers. Classification is done by the fully connected layers and softmax activation, which predicts disease categories.

Experimental Results and Discussion:

Outputs:

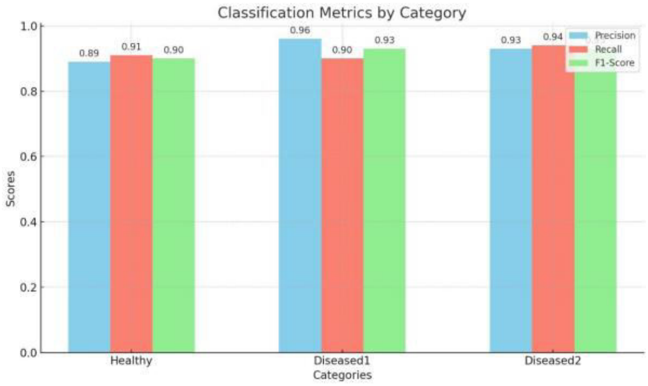

Figure.3. Performance metric analysis

The bar graph above is the classification metrics-precision, recall, and F1-score-across three classes: "Healthy," "Diseased1," and "Diseased2." Each bar shows the corresponding metric for that class. The "Healthy" class had a precision of 0.89, recall of 0.91, and an F1-score of 0.90, thus balancing the performance. The "Diseased1" class had a high precision at 0.96 but low recall at 0.90, thus giving an F1-score of 0.93. Correspondingly, "Diseased2" exhibited a performance throughout all three metrics: precision: 0.93, recall: 0.94, and F1-score: 0.93, thus validating the strength of the model in distinguishing between categories correctly. The outcomes proved that the model can be employed for the proper classification of plant diseases.

```
Classification Report:
              precision    recall  f1-score   support

     Healthy       0.89      0.91      0.90        55
   Diseased1       0.96      0.90      0.93        50
   Diseased2       0.93      0.94      0.93        50

    accuracy                           0.92       155
   macro avg       0.93      0.92      0.92       155
weighted avg       0.93      0.92      0.92       155
```

Figure.4. Classification report

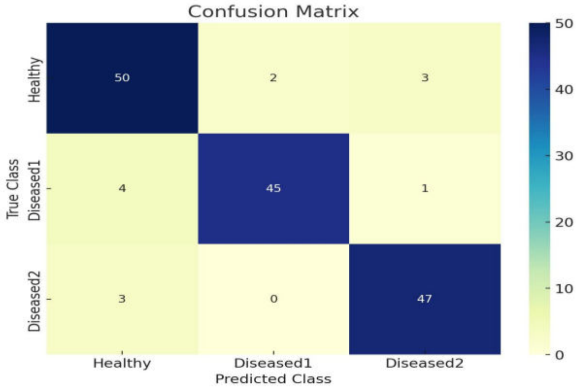

Figure.5. Confusion matrix

Classification performance of the model across three classes: "Healthy" (Class 0), "Diseased1" (Class 1), and "Diseased2" (Class 2).
Class 0 (Healthy): Out of 55 instances, the model correctly predicted 50 as "Healthy." It has misclassified 2 instances as "Diseased1" and 3 as "Diseased2."
Class 1 (Diseased1): It had classified 45 as "Diseased1" and they were correct of 50. It also incorrectly labeled 4 as "Healthy" and 1 as "Diseased2".
Class 2 (Diseased2): The model classifies 47 out of 50 instances as "Diseased2." The model misclassifies 3 instances as "Healthy," but none as "Diseased1."
This matrix summarizes the strengths and weaknesses of the model. Although the majority of classifications are correct, the false classifications ("Healthy" mistakenly classified as "Diseased2") point toward potential places for improvement - perhaps refinement of feature extraction or re-balancing the class distribution in the training set.

```
Confusion Matrix:
[[50  2  3]   # Class 0 (Healthy): 50 correct, 2 predicted as Diseased1, 3 as Diseased2
 [ 4 45  1]   # Class 1 (Diseased1): 4 predicted as Healthy, 45 correct, 1 as Diseased2
 [ 3  0 47]]  # Class 2 (Diseased2): 3 predicted as Healthy, 0 as Diseased1, 47 correct
```

Conclusion

The "Plant Disease Identification System" project has been a giant step forward for agricultural technology and crop health management. We have successfully developed an early and accurate plant disease identification applying high-end machine learning techniques and analytics. With the ability to provide for interventions at times when they are required most by producing timelier results, this means it has the capability of revolutionizing plant disease diagnostics. This would then potentially lead to improved crop yield and reduced losses. One would also gain insight regarding what diseases plants are diseased by. It, thus forms part of deeper knowledge toward plant health. Similarly, machine learning and approaches in data-driven methodologies pertaining to agricultural practices have ascertained potential in this particular industry.

The future scope of the project is limitless. This would be feasible if the model could gain access to larger and more diverse datasets containing a wider variety of plant species and types of diseases. Futher more, the project may be expanded to include real-time crop health monitoring via drone or satellite imagery, providing a more proactive approach to disease management. Then, by integrating it into mobile apps or IOT devices in the field, one would empower farmers to immediately have access to diagnostics and actionable insights. One could work with agricultural institutions for further field trials and validation, thereby adding more to the reliability of the system and therefore possible uptake in mainstream farming.

References

1. Mohanty, S. P., Hughes, D. P., & Salathe, M. (2016). "Using Deep Learning for Image-Based Plant Disease Detection." Frontiers in Plant Science, 7, 1419.
2. Ferentinos, K. P. (2018). "Deep learning models for plant disease detection and diagnosis." Computers and Electronics in Agriculture,145,311-318.
3. Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). "Deep Neural Networks based Recognition of Plant Diseases by Leaf Image Classification." Computational Intelligence and Neuroscience, 2016.
4. Zhang, S., Zhang, S., Huang, T., Gao, W., & Qiao, M. (2019). "Plant Disease Recognition Based on Plant Leaf Image." Journal of Image and Graphics, 7(2), 30-35.
5. S. Li, B. Yang, and M. Liu, "A Survey on Machine Learning Techniques for Plant Disease Detection and Classification," Computers and Electronics in Agriculture, vol. 171, pp. 105-125, 2020.
6. Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.
7. X. Yu, H. Zhang, and J. Xu, "Enhancing Remote Sensing Image Classification with Data Augmentation Techniques," IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 6, pp. 4682–4695, 2021.
8. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 2, pp. 142–158, 2016, doi: 10.1109/TPAMI.2015.2437384.
9. S. Minaee et al., "Image Segmentation Using Deep Learning: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 7, pp. 3689–3704, 2022.
10. J. Schmidhuber, "Deep Learning in Neural Networks: An Overview," Neural Networks, vol. 61, pp. 85– 117, 2015, doi: 10.1016/j.neunet.2014.09.003.
11. W. Liu et al., "SSD: Single Shot Multibox Detector," in Proceedings of the European

Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 2016, pp. 21–37.

12. T. Guo, J. Chen, and J. Zhou, "Plant Disease Recognition Based on Deep Learning and Transfer Learning," IEEE Access, vol. 8, pp. 102838–102850, 2020.
13. S. Deng et al., "Recent Advances in Image-Based Plant Disease Detection," Frontiers in Plant Science, vol. 12, pp. 637–648, 2021, doi: 10.3389/fpls.2021.637948.