

## Plant Leaf Disease Detection and Classification Using Machine Learning

P Kiran Kumar Reddy  
Computer Science and Engineering  
Lovely Professional University  
Phagwara, Punjab  
kiran.12014385@lpu.in

Perumalla Sivagopi  
Computer Science and Engineering  
Lovely Professional University  
Phagwara, Punjab  
siva.12014666@lpu.in

Vallabhaneni Prudhvi Sai  
Computer Science and Engineering  
Lovely Professional University  
Phagwara, Punjab  
vallabhaneni.12002918@lpu.in

Shaik Nawaz Shareef  
Computer Science and Engineering  
Lovely Professional University  
Phagwara, Punjab  
shaik.12019237@lpu.in

Leela Sasi Kanth  
Computer Science and Engineering  
Lovely Professional University  
Phagwara, Punjab  
suroju.12008346@lpu.in

Ravi Durga Prasad  
Computer Science and Engineering  
Lovely Professional University  
Phagwara, Punjab  
ravi.12010940@lpu.in

**Abstract—** A viable solution to problems in agricultural management has been the identification and categorization of plant leaf diseases through machine learning techniques. For early intervention and efficient disease control, this work focuses on developing a system that can automatically detect and classify plant leaf illnesses. The technology examines digital photos of plant leaves to detect symptoms that may be signs of several diseases by using machine learning algorithms like CNN and Random Forest. A collection of captioned photos representing various plant species and disease types makes up the training and testing dataset. In order to help farmers make quick decisions and allocate resources, the system uses machine learning algorithms to reliably classify photos into illness categories. Through its potential answer, this research advances agricultural technology.

**Keywords—** Plant leaf disease, Detection, Classification, Machine learning, Agricultural management, Digital image analysis, Crop protection.

### I. INTRODUCTION

The foundation of the world's food production and subsistence system, agriculture provides for the dietary needs of billions of people globally. Plant diseases, on the other hand, are a major problem for farmers and agricultural communities because they can negatively affect crop output, food security, and economic stability. Plant diseases are a persistent challenge to agricultural productivity because they cause significant yield losses and quality degradation in crops and are caused by pathogens such fungus, bacteria, viruses, and pests.

Global economic stability, food security, and agricultural production are all seriously threatened by plant diseases. For crop protection and disease management to be successful, these diseases must be promptly and accurately detected. Conventional disease detection techniques frequently depend on farmers or

agricultural specialists visually inspecting the affected area, which can be labor-intensive, time-consuming, and prone to human error. Furthermore, the illness can have already seriously harmed the crop by the time symptoms start to show. Farmers and agricultural specialists frequently use physical observation and visual inspection of plants as part of traditional techniques of disease identification and management. These methods are time-consuming, labor-intensive, and subjective by nature. Furthermore, they might not always be successful in detecting illnesses in the earliest stages, when intervention efforts might be most helpful. Therefore, automated and effective plant disease detection and classification technologies are desperately needed in order to supplement current techniques and offer prompt and precise diagnoses.



**Figure 1:** Marssonina leaf spot

Automated methods for plant disease identification and categorization have been made possible by recent developments in machine learning, computer vision, and digital imaging technology. These devices use machine learning algorithms to examine digital photos of plant leaves and recognise distinctive

signs linked to different illnesses. These algorithms can identify minute patterns and traits indicative of particular diseases by learning from enormous datasets of labelled photos, making disease diagnosis precise and effective.

Research in this area has advanced significantly thanks to the availability of extensive datasets with annotated photos of both healthy and damaged plants. Researchers can train and assess machine learning models for plant disease detection across a variety of agricultural contexts thanks to these datasets, which cover a broad range of plant species, disease kinds, and environmental factors.

The discipline of computer vision and machine learning has seen significant breakthroughs recently, and as a result, there is increasing interest in creating automated systems for the identification and categorization of plant diseases. These systems use digital image analysis techniques to identify typical symptoms associated with different diseases by analysing photographs of plant leaves. These systems may learn to identify patterns in the photos and make highly accurate distinctions between healthy and unhealthy plants by utilising machine learning algorithms.

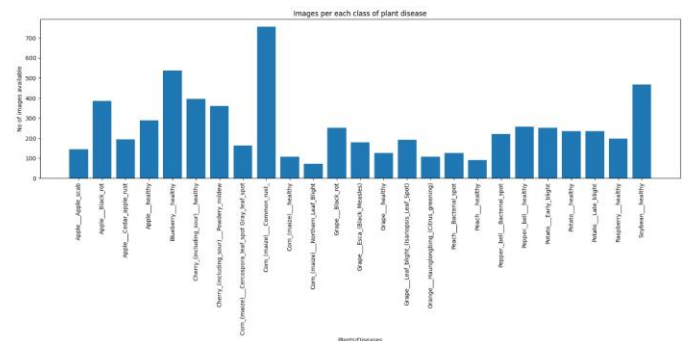


**Figure 2:** Images of tea leaves having diseases

Machine learning models for plant disease identification have been made easier to construct and assess by the availability of large-scale datasets with labelled photos of both healthy and diseased plants. With the help of these datasets, which cover a variety of plant species and disease kinds, researchers can develop models that can identify a broad range of plant diseases in a variety of crops.

By applying machine learning approaches, we hope to contribute to the current efforts in agricultural technology by creating a system for the identification and categorization of plant leaf diseases. Our method aims to automate disease diagnosis by utilising digital imaging and machine learning algorithms. This would enable farmers to receive accurate and timely information that will help with crop protection and disease management. We work to address the problems caused by plant diseases and improve agricultural sustainability and resilience in the face of changing environmental and

socioeconomic constraints by integrating cutting-edge technologies and interdisciplinary approaches.



**Figure 3:** This bar graph shows the number of images available in a plant disease dataset for various plant species and diseases.

Our method attempts to classify digital photos of plant leaves into appropriate disease groups and recognise symptoms suggestive of different diseases. Our goal is to detect diseases more accurately and efficiently by using machine learning models. This will provide farmers with the knowledge and tools they need to safeguard their crops and manage diseases.

## II. LITERATURE SURVEY

The combination of machine learning algorithms with image processing techniques has shown promise in recent years for the early diagnosis and categorization of plant diseases, which could lead to an increase in agricultural productivity. Our thorough investigation explores this intersection and provides information on how to identify and categorise plant leaves impacted by different illnesses. We demonstrate the effectiveness of image processing and machine learning in precisely identifying and classifying plant illnesses through an examination of prevalent infections and investigation scenarios. In addition, this study clarifies the difficulties and potential paths ahead in this area, offering practitioners and researchers insightful direction. All things considered, our research is an invaluable tool for comprehending the uses of machine learning and image processing for the identification and categorization of plant diseases.[1]

In addition to making a substantial contribution to the Gross Domestic Product (GDP) of countries, the agricultural sector is a vital source of food for human consumption. Nonetheless, manual farming practices are common in many impoverished nations, which makes it difficult to identify plant diseases in a timely manner. Farmers who discover diseases later in life suffer significant financial losses, which has an effect on the state and national economies. Efficient detection and classification of diseases is hampered by issues such uneven background during picture collection, segmentation, and classification. The proper application of control mechanisms can occur once diseases have been correctly identified through the use of symptoms and

characteristic analysis. This overview explores the field of plant diseases, including classical methods, machine learning techniques, and deep learning approaches for disease detection and classification. The results show that whereas conventional approaches and certain machine learning strategies like deep learning methods demonstrate superior performance in disease identification and classification.[2]

With over 19,000 fungi known to infect crops, plant diseases represent a serious threat to agricultural crops worldwide, affecting crop productivity and economic value. To lessen these impacts, leaf diseases must be diagnosed as soon as possible. Disease identification techniques that are automated are crucial because they minimise labour costs and facilitate early intervention. This work proposes an image segmentation-based computerised system for plant leaf disease identification and classification. The suitability of several deep learning algorithms and disease classification techniques for identifying a range of plant diseases is reviewed. Utilising automated identification techniques, like the ones described in this paper, agriculture can protect crop yields and increase productivity—two things that are essential to sustainability and economic growth.[3]

Hibiscus is a potent herb that is well-known in Ayurveda for its extraordinary therapeutic qualities. It is rich in vitamin C, flavonoids, amino acids, mucilage fibre, and antioxidants. Because of its wide range of nutrients, it can help with weight loss, cancer treatment, bacterial infections, fever management, and blood pressure regulation. Additionally, hibiscus shows promise in treating heart and nerve disorders as well as decreasing body temperature. Automatic identification of leaf disease is critical in agriculture, and image processing is a widely used method for this. This chapter presents a methodology that combines feature extraction with concurrent k-means clustering for sick leaf identification. Subsequently, a reweighted KNN linear classification method employs the retrieved features to precisely identify and classify different kinds of sick leaves. This novel strategy has the potential to improve the accuracy and efficiency of plant disease identification and treatment.[4]

This paper addresses the critical issue of low agricultural productivity caused by plant diseases, emphasizing the challenges faced by farmers in disease detection and control. Focusing on maize plant disease detection, supervised machine learning techniques including Naive Bayes, Decision Tree, K-Nearest Neighbor, Support Vector Machine, and Random Forest are employed using plant images. Through comparative analysis, the Random Forest algorithm emerges with the highest accuracy of 79.23%, outperforming other classification techniques. These trained models offer farmers a valuable tool for early disease detection and classification, enabling proactive measures to mitigate losses and optimize crop management strategies. [5]

The project's main goal is to improve cannabis plant disease detection and control by utilising machine learning techniques. Building a strong model for precise and quick disease diagnosis is the goal, and it will be accomplished by examining traits including texture, growth patterns, and leaf alterations. The goal of this project is to address the financial consequences of

cannabis ailments, which can have a big influence on productivity. The goal of the project is to enhance the early diagnosis and control of diseases in cannabis farming by utilising a variety of datasets of photographs of cannabis plants. The ultimate objective is to develop a productive and efficient technique that helps cultivators sustain the well-being and yield of their cannabis harvests, supporting the plant's many uses in both commercial and therapeutic contexts.[6]

Plant leaf disease detection using computer vision and machine learning is crucial for enhancing crop yield. Despite past efforts, achieving high detection accuracy remains challenging. To address this, we propose a novel hybrid approach. Firstly, image enhancement and conversion techniques mitigate illumination and noise issues. Next, we develop a combined feature extraction method utilizing GLCM, Complex Gabor filter, Curvelet, and image moments. Finally, a Neuro-Fuzzy Logic classifier is trained with these features. Implemented in MATLAB using PlantVillage Database, the approach achieves over 90% detection accuracy across two test cases. This comprehensive method offers improved classification accuracy, making it beneficial for UG/PG students undertaking project-based learning.[7]

Identifying plant diseases accurately is essential to preventing reductions in agricultural yield. Examining external signs is crucial and necessitates continuous plant health monitoring. Disease auditing requires specific tools, a significant investment of time, and labour. Image processing compares leaf pictures with datasets to expedite detection. Computers are taught to distinguish between healthy and diseased leaves through machine learning. This method provides a scalable option for the diagnosis of prevalent diseases. Machine learning makes efficient agricultural monitoring possible by utilising large datasets. This paper offers greenhouse producers an easy-to-use yet efficient way to identify plant diseases.[8]

The early diagnosis of rice blast disease, a major global concern in agriculture, is made possible by the machine learning algorithm proposed in this study. The system interprets photos of rice leaves with and without disease, automatically detecting signs of each. Features are retrieved from both healthy and damaged leaf portions using a 300 image dataset that has been split into training and testing sets. With training phase results showing 99% accuracy for blast-infected photos and 100% accuracy for healthy ones, the suggested approach achieves great accuracy. In testing, the algorithm's efficacy in identifying rice blast disease is demonstrated by the accuracy, which is steady at 90% for infected photos and 86% for healthy ones.[9]

In nations such as India, where agriculture accounts for about 58% of rural incomes, it is critical to maintain the productivity and health of important crops like tomatoes. To avoid significant losses, tomato plant disease detection and categorization are essential. This research tackles illness diagnosis through four stages: pre-processing, leaf segmentation, feature extraction, and classification. It does this by utilising cutting-edge technology including image processing. Preprocessing removes noise, and segmentation separates the impacted leaf sections. Regression



and disease classification are made easier by the robust supervised learning technique known as the k-nearest neighbours (KNN) algorithm. In the end, farmers get timely treatment recommendations that take into account the concerns found. This paper provides farmers with an essential tool to quickly evaluate and minimise hazards to tomato plants by effectively identifying disease symptoms based on colour, shape, and texture from photographs.[10]

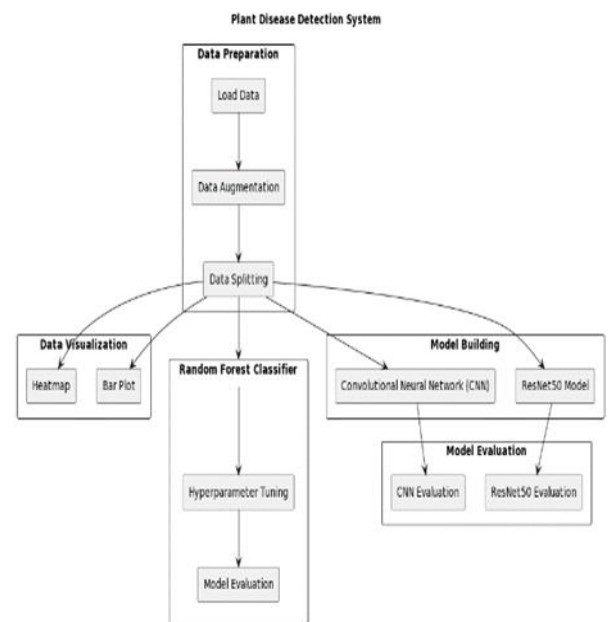
This work offers a comprehensive assessment of the literature with an emphasis on developments in machine learning (ML) and deep learning (DL) models for plant disease detection and classification systems. From 2017 to 2020, around 45 peer-reviewed publications were gathered from databases such as Scopus and Web of Science. The analysis is arranged in well-structured tables that offer insights into several ML and DL models that are applied to the categorization of plant diseases. Along with processing techniques like image segmentation and feature extraction, cutting-edge algorithms like Support Vector Machine (SVM), Neural Network (NN), K-Nearest Neighbour (KNN), and well-known DL models like AlexNet, GoogLeNet, and VGGNet are investigated. Standardised metrics are given, such as dataset size, number of diseases taken into account, kind of classifier, and accuracy of classification. This thorough analysis is an invaluable tool for scientists looking to use data-driven methods for identifying plant diseases, which could result in the creation of mobile applications that increase agricultural output.[11]

The identification and categorization of hemp illnesses, which have a substantial influence on hemp production and economic viability, are studied in this research using artificial intelligence and machine learning approaches. For this goal, it suggests three deep learning ensemble models and one SVM-based machine learning model. Hemp Powdery Mildew, Hemp Leaf Spot, Hemp Bud Rot, and Hemp Nutrient Deficiency are among the ailments that are being targeted. By utilising transfer learning in conjunction with pre-trained deep learning models, the study presents comparative evaluation results that demonstrate up to 98% accuracy across several models. These results show how machine learning has a great deal of potential in smart agriculture, especially in managing hemp diseases and providing opportunities for enhanced crop health and production optimisation.[12]

## Architecture:

Architectural framework that outlines a methodical approach to handling the intricacies involved in plant disease diagnosis. Fundamentally, the framework focuses on the important phases of model building, training, evaluation, and integration with conventional machine learning methods. In order to protect agricultural crops and improve overall output and sustainability, this all-encompassing strategy guarantees the system's efficacy and dependability in precisely recognising and diagnosing a range of plant diseases. This helps with prompt interventions

and management techniques. Obtaining and comprehending the dataset is the focus of careful attention in the early phases of data preparation and exploration. The architecture establishes the foundation for well-informed decision-making about model construction and training techniques by loading photos from directories and performing exploratory data analysis, including class distribution analysis and image count per class visualization. Additionally, preprocessing and data augmentation methods are used to improve the diversity and quality of the dataset, which helps the models develop reliable representations of plant illnesses from the available imagery.



**Figure 4:** Architecture of Plant disease Detection System

A crucial part of the architecture is the model creation step, where two different strategies are used: using a pre-trained ResNet50 model and creating a custom CNN architecture. The advantage of transfer learning is that the model can accelerate learning and increase classification accuracy by using the knowledge from training on the ImageNet dataset, thanks to the use of ResNet50 as the basis architecture. Concurrently, the customised CNN architecture offers adaptability in adjusting model parameters to the distinct features of the plant disease classification assignment, guaranteeing alignment with particular dataset prerequisites and goals.

After the model is built, the architecture moves to the training and validation phases, where strict training procedures are put in place to maximise the model's generalisation and performance. The models go through iterative training cycles using validation datasets and data generators, where model parameters are changed to maximise accuracy and minimise loss. The models' predictive powers are improved and their capacity to correctly identify plant diseases under a variety of circumstances is increased through this recurrent training and validation

procedure. In addition, by including conventional machine learning methods—such as a Random Forest Classifier for comparison analysis—the architecture expands its potential. The architecture uses methods for hyperparameter tweaking, like grid search cross-validation, to find the best model configurations in terms of complexity, interpretability, and performance. To summarise, the code snippet's architectural framework is a comprehensive strategy for plant disease identification that incorporates deep learning methods, data-driven approaches, and conventional machine learning concepts. The framework enables practitioners and academics to create resilient, scalable, and understandable solutions for reducing the effects of plant diseases on agricultural ecosystems and promoting sustainable food production methods by combining these elements in a synergistic way.

### III. METHODOLOGY

#### Random Forest Classifier:

**Overview:** This section gives a general description of the Random Forest Classifier algorithm and explains how it operates as an ensemble learning technique. It demonstrates how Random Forest Classifier builds several decision trees throughout training and combines the predictions from each to provide the final classifications.

**Application to the Project:** It explains Random Forest Classifier's function in the project and highlights that it is a conventional machine learning technique that can be used to compare with deep learning models. This offers a standard against which to compare how well deep learning systems categorise plant diseases.

**Hyperparameter Tuning:** The paragraph explains how to edit the Random Forest Classifier's default hyperparameters, first initialising them and then fine-tuning them using grid search cross-validation. The precise hyperparameters being adjusted and their effect on classification accuracy are mentioned.

#### Convolutional Neural Networks(CNN):

**Overview:** Convolutional neural networks, or CNNs, are a class of deep neural networks with a focus on image identification and classification problems. They are introduced in this paragraph. It gives a high-level explanation of how convolutional and pooling layers enable CNNs to learn hierarchical representations of picture information.

**Use in the Project:** It describes how CNN-based architectures are used in the project, mentioning in particular the use of a bespoke CNN architecture and a pre-trained ResNet50 model. The process of training both designs on the plant disease dataset and assessing their classification accuracy is explained in the paragraph.

**Training and Evaluation:** The training and evaluation procedure for the CNN architectures is covered in the final paragraph of this section. The utilisation of image data generators for training and the assembly of models with suitable optimizers, loss functions, and assessment metrics are mentioned. It also emphasises that in order to appropriately measure the models' performance, they are trained and assessed on both training and validation datasets.

A well-known deep convolutional neural network architecture called ResNet50 is the pretrained model used in the plant disease detection research. Having been pre-trained on a massive dataset, usually ImageNet, with millions of labelled photos in various categories, ResNet50 is well-known for its efficiency in image identification tasks. Convolutional, pooling, and fully connected layers are among the 50 layers in its architecture, which is intended to learn hierarchical representations of picture information. ResNet50 functions as a pretrained model in the context of the plant disease detection project, meaning that it has already received substantial training on the ImageNet dataset to identify a wide range of objects and attributes in images.

Transfer learning is used to modify the pretrained ResNet50 model for the particular purpose of classifying plant diseases. This is adding further layers to the fundamental design, like a Global Average Pooling layer and a Dense layer, in order to fine-tune the model. Using the information and features obtained during pre-training on ImageNet, these extra layers are subsequently trained on the plant disease dataset to collect characteristics appropriate to the task at hand. The plant disease detection project gains various advantages from the use of a pretrained model such as ResNet50. The model can efficiently extract pertinent features from input photos and classify plant diseases with high accuracy by utilising the pre-trained weights and learned features of ResNet50. Pretrained models are an important tool in the detection system's creation since they allow practitioners to attain strong performance even with a small amount of training data. In conclusion, the plant disease detection study finds that the pretrained ResNet50 model is an effective tool for feature extraction and transfer learning. With the help of transfer learning techniques and its pretrained status, it is easier to adapt learned features to the particular task of classifying plant illnesses. This leads to an effective and precise detection system that can precisely identify and diagnose a wide range of plant diseases.

### IV. CONCLUSION

Distinct trends are revealed by comparing the performance of the Random Forest, custom CNN, and ResNet50 models in the plant disease detection project based on the accuracy parameters that were provided. ResNet50 performs moderately, with a training accuracy of 28.7% and a test accuracy of 17.09%. Despite achieving a significantly greater accuracy during training, the model's lower test accuracy indicates that it has

difficulty generalising to new data. This points to a possible overfitting scenario in which the model may have learned useful patterns that apply to new cases by memorization of the training data.

The performance of the customised CNN model, on the other hand, is noticeably better, with test accuracy of 83.36% and training accuracy of 98.6%. These accuracy scores show excellent generalisation to new data and strong learning capabilities. The customised CNN model performs exceptionally well on both training and test datasets by efficiently learning complex patterns and features pertinent to the categorization of plant diseases. Because accurate disease diagnosis is crucial in real-world applications, its excellent accuracy on the test dataset highlights its dependability and appropriateness. The Random Forest model, on the other hand, displays a different performance profile.

Recalling the training data is an area where the model seems to flourish, with a near-perfect training accuracy of 99%. Its test accuracy is noticeably lower, though, indicating a substantial discrepancy between test performance and instruction. This disparity indicates that the model may have been overfit to the training set, as evidenced by its poor performance on unknown cases and inability to generalise. Because of its poor generalisation abilities, the Random Forest model might not be the best choice for real-world plant disease detection tasks.

The comparison analysis concludes by demonstrating how much better the bespoke CNN model is at precisely identifying plant illnesses. Because of its strong generalisation capabilities and outstanding performance on training and test datasets, it is the recommended option for plant disease detection applications. The best and most dependable method for classifying plant diseases accurately and consistently is the bespoke CNN model, while Random Forest suffers from overfitting and poor generalisation, and ResNet50 performs moderately.

## V. FUTURE ENHANCEMENTS

Future developments in the field of plant disease detection can be made to increase the system's efficacy. First off, adding more photos of plant diseases from a variety of settings to the dataset can improve the model's generalisation capacity. Rotation and colour jittering are examples of advanced data augmentation techniques that can further diversify the dataset and increase the robustness of the model.

Examining model architectural optimisations, such as changes and cutting-edge methods like network pruning, can improve interpretability and efficiency. By integrating predictions from several models, ensemble learning techniques have the potential to improve overall performance and reduce overfitting. Optimising model performance can be achieved by fine-tuning hyperparameters using methods like grid search. Users' trust is increased when explainability of the model is emphasised using methods such as attention maps.

Potential for real-time illness monitoring is provided by developments in real-time deployment and IoT device integration. Working together with subject matter specialists can improve the model's comprehension of plant diseases and adapt it to the requirements of agriculture.

## VI. REFERENCES

- [1] Appalanaidu, M.V., & Kumaravelan, G. (2021). Plant Leaf Disease Detection and Classification Using Machine Learning Approaches: A Review. *Innovations in Computer Science and Engineering*.
- [2] Jayswal, H., & Chaudhari, J.P. (2020). Plant Leaf Disease Detection and Classification using Conventional Machine Learning and Deep Learning.
- [3] Bhosale, Y.H., Zanwar, S.R., Ali, S.S., Vaidya, N.S., Auti, R.A., & Patil, D.H. (2023). Multi-Plant and Multi-Crop Leaf Disease Detection and Classification using Deep Neural Networks, Machine Learning, Image Processing with Precision Agriculture - A Review. 2023 International Conference on Computer Communication and Informatics (ICCCI), 1-7.
- [4] Usharani, B. (2022). House Plant Leaf Disease Detection and Classification Using Machine Learning. *Advances in Systems Analysis, Software Engineering, and High Performance Computing*.
- [5] Panigrahi, K.P., Das, H., Sahoo, A.K., & Moharana, S.C. (2020). Maize Leaf Disease Detection and Classification Using Machine Learning Algorithms.
- [6] Arivazhagan, M. (2023). Hemp Disease Detection and Classification using Machine Learning. *INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*.
- [7] Rao, A., & Kulkarni, S.B. (2020). RETRACTED: A Hybrid Approach for Plant Leaf Disease Detection and Classification Using Digital Image Processing Methods. *International Journal of Electrical Engineering & Education*, 60, 3428 - 3446.
- [8] Kalaiselvi, K., Kumaresan, D., Thilak, D., & Devi, D. (2023). Multiple plant leaf disease detection using efficient Machine learning classification techniques. 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), 1-4.
- [9] S, R. (2018). Rice Blast Disease Detection and Classification Using Machine Learning Algorithm. 2018 2nd International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE), 255-259.
- [10] Geetha, G.P., Samundeswari, S., Saranya, G., Meenakshi, K., & Nithya, M. (2020). Plant Leaf Disease Classification

- and Detection System Using Machine Learning. Journal of Physics: Conference Series, 1712.
- [11] Applalanaidu, M.V., & Kumaravelan, G. (2021). A Review of Machine Learning Approaches in Plant Leaf Disease Detection and Classification. 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 716-724.
- [12] Bose, B., Priya, J., Welekar, S., & Gao, Z. (2020). Hemp Disease Detection and Classification Using Machine Learning and Deep Learning. 2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), 762-769.
- [13] <https://shorturl.at/jOX07>