PlantShield : GLCM and KNN Fusion in CNN for Robust Plant Disease Detection

Tarushi Sandeep Gupta
Dept. of CSE (AIML), CSE-APEX,
Chandigarh University, India
tarushigupta03@gmail.com

Shivani
Dept. of CSE (AIML), CSE-APEX,
Chandigarh University, India
shivanisaini03@gmail.com

***Abstract*— Plant diseases are a critical problem in modern agriculture, responsible for significant productivity losses of crops and economic damages. This paper describes the PlantShield system that employs both texture-based as well as deep learning methods to enhance the robustness and precision of plant disease identification. The PlantVillage dataset, which has approximately 54,000 photos of both healthy and damaged leaves from different plant species, serves as the foundation for the system. In this approach, textural properties like contrast and homogeneity are extracted using the Gray Level Co-occurrence Matrix, while deep visual features are extracted using CNN. This KNN method is used for fusing and classification of the features. The system would be supposed to offer a better performance when it comes to differentiation between subtle disease patterns through fusing low-level texture features with higher-level CNN features. Evaluation for the accuracy, recall, precision, and F1-score shows that the proposed solution by PlantShield is robust and reliable in early plant disease detection, promising its applications in precision agriculture and real-time field monitoring.**

Keywords: Plant disease detection, GLCM, CNN, KNN, PlantVillage dataset.

I. INTRODUCTION

One of the main sectors of the global economy, agriculture supplies resources and staple foods to people all over the world. However, there are considerable challenges that it faces, particularly in the form of plant diseases that could seriously deplete crop yield, degrade their quality, and lead to heavy economic losses. An important requisite in the direction of mitigation of these risks to ensure food security is the accurate and early diagnosis of plant diseases. Traditional approaches used in disease detection rely on simple observation by experts who have merely looked at the crop. While this can sometimes yield good results, these approaches are slow and labor-intensive, prone to human mistakes, and not easy to scale up if farmlands have large operation systems. The efficiency of such an approach underscores a need for more advanced, automated approaches that can adapt to both the complexity and diversity that occurs with modern agricultural environments.

In this respect, AI and ML have been opened up to revolutionize the plant disease detection process. Currently, these technologies have opened channels to automate the process of identification of such diseases into quick, accurate, and massive scale identification processes. AI-based approaches thus proved to be useful in handling complex data and patterns from the rest and made it possible to change the way in which the detection system was traditionally operated between manual and automated detection. These systems, in addition to cutting the dependence on human expertise, bring precision and consistency in identifying disease, which is so important for timely interventions that prevent the spread of infection.

The current system integrates the proven techniques of extracting texture feature from GLCM along with very effective algorithms of KNN and CNN, which are often very useful in classification. GLCM differs at subtle texture features such as contrast, energy, and correlation, which characterize the images of the leaves, and thereby are being used for detection and analysis of different plant diseases. However, On the other hand, CNNs are naturally suited to extracting higher-level features in forms of shapes, edges, and patterns from the visual data that provide better insight into the nature of disease symptoms. The combination of GLCM texture analysis and CNN's pattern recognition capability reinforces higher precision and dependability within the model while identifying diseases of diversified crops.

Additionally, the incorporation of KNN further enhances the classification process since it not only employs a decision mechanism but also maps the disease category based on the similarities to other data points closer to the feature space. Its strength is mostly detected in the fused feature sets such as the features from GLCM and CNN since it considers both texture and visual information content for informed decisions. This combination of approaches not only enhances the predictive accuracy of the model in overall but also ensures that the system is robust with the early detection of diseases-so important for proper timely treatment and minimum damage.

Thus, the PlantShield model comes as a scalable and reliable solution to the agricultural sector. The development of detection of diseases in plants through automation will be helpful for farmers and agricultural professionals in monitoring the health of their crops, thus enabling them to take proactive measures against outbreaks. The extension of plant disease defense mechanism as a combination of image processing, image analysis, algorithms and machine learning is a huge step forward in the fight against diseases on crops, equipping the agricultural industry with tools to increase productivity and sustainability.

A. Understanding GLCM, KNN and CNN

The Gray Level Co-occurrence Matrix (GLCM) is a statistical method of texture analysis within image processing in a dataset, which quantifies how frequently there is a spatial relationship between the pairs of pixel values, as defined by the filter, within an image, and forms a matrix describing the distribution of these pixel pairs: contrast, correlation, energy, and homogeneity are some features that can be derived from such a matrix. These are some of the necessary features for texture analysis, since they indicate the structure and pattern in an image. Therefore, GLCM has been useful, especially in disease diagnosis regarding plant tissues, where many texture changes lead to abnormalities.

KNN is perhaps the most intuitive and interesting model in machine learning. It is a simple, instance-based algorithm applied to various problems of classification and regression. Most common metrics used for calculating proximity are the Euclidean distances, though other metrics can be applied. Because KNN does not have an explicit training phase, it remains computationally cheap for smaller datasets. It's one of the application areas to be applied in plant disease detection; KNN can be used for classifying diseases based on features extracted, such as by GLCM, compare new data points with already labeled instances.

Convolution Neural Networks are deep learning architectures that are proficiencies in image processing tasks such as classification, detection, and segmentation. Layers like CNN include convolution layers that remove filters to detect a particular type of image patterns, pooling layers to reduce the dimensionality, and fully connected layers to classify finally. The main strength in CNNs is that they automatically learn which features in images are most relevant--edges, textures, shapes--without any manual feature extraction. CNNs can learn to classify diseases in the plant based only on visual symptoms such as spots, discolorations, and changes in texture from raw images of the leaves.

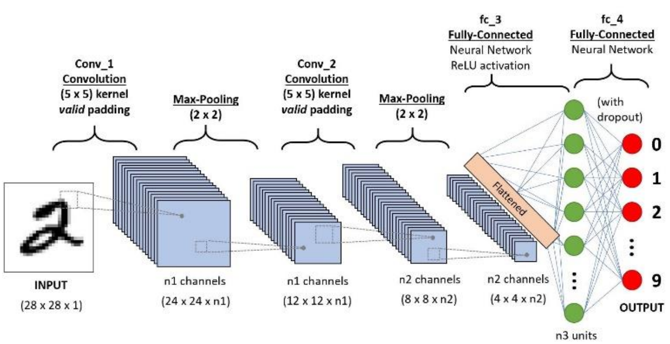

Fig. 1. CNN Architecture

B. Role of GLCM and KNN in PlantShield

Gray Level Co-occurrence Matrix, GLCM is a statistical technique for texture analysis in image processing that creates a matrix that describes the distribution of these pixel pairs and quantifies the frequency with which pairs of pixel values occur in an image in a spatial connection established by the filter: contrast, correlation, energy, and homogeneity are some features that can be derived from such a matrix. Some of these features are indispensable to texture analysis because they indicate structure and pattern in an image. Hence, GLCM has been very helpful, especially in the disease diagnosis regarding plant tissues, since so many texture changes lead to abnormalities.

KNN may be the most intuitive and interesting model in machine learning. This is a straightforward instance-based approach that is applied to regression and classification issues. Actually, a new data point's classification is decided by the majority vote of its k nearest neighbors, according to KNN. Most common metrics used to calculate proximity are Euclidean distances, but other metrics could be applied. This is because KNN does not have an explicit training phase; therefore, it stays relatively inexpensive computationally for smaller datasets. One of the application areas that this needs to be applied in is plant disease detection where KNN can be used to classify different diseases based on the features extraction that were extracted, for example, by GLCM, comparing new data points with already labeled instances.

Convolution Neural Networks are deep learning architectures with proficiencies in image processing tasks including classification, detection, and segmentation. The layers include like CNN convolution layers that remove filters to detect a particular kind of the pattern in an image; pooling layers reduce the dimensionality; fully connected layers classify finally. The greatest strength of CNNs is automatic learning of which features in images are most relevant, namely edges, textures, and shapes, without any manual feature extraction. Their ability to learn what of these features could be used to classify diseases in a plant based merely on visual symptoms like spots, discolorations, and texture changes from raw images of the leaves makes them rather effective.

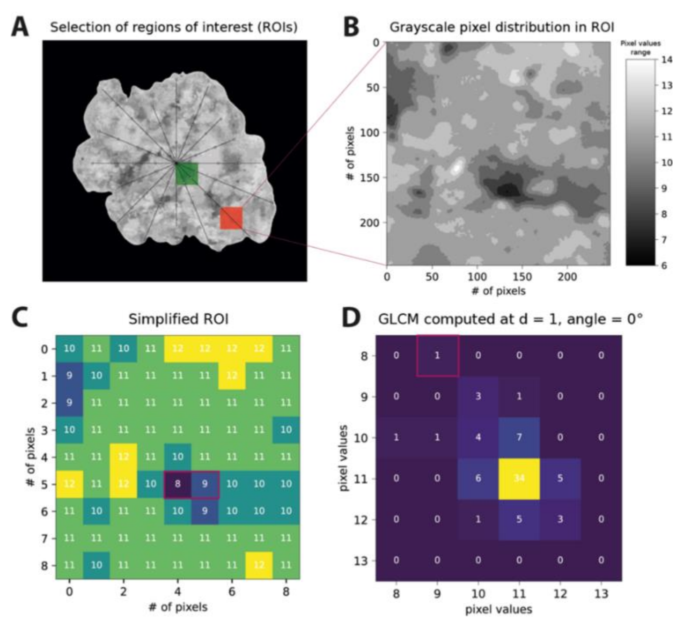

Fig. 2. Calculation of Gray Level Co-Occurrence Matrix (GLCM)

C. *Dataset – PlantVillage Dataset*

The massive and frequently used dataset for the plant disease detection task of PlantVillage Dataset from Kaggle consists of more than 54,000 images of labeled plant leaves with 38 plant species and 26 distinct diseases; hence, it is best for training machine learning models, used for the classification of diseases and the detection of plant diseases with healthy and diseased leaves images. The quality of the images is very high, with clear definitions for many symptoms of diseases: spots, lesions, discolorations, which are essential in developing robust disease detection models.

The diversity in crops is also high, comprising common crops like tomatoes, apples, grapes, and potatoes. The size and diversity of the dataset allow for having multiple classes of disease categories along with a healthy category that can be used for both the tasks of disease classification and anomaly detection. A very large, diverse dataset will enable more generalized models to be built, which, in the real world, will be applicable to situations where there may be several diseases attacking plants simultaneously with overlapping symptoms.

The dataset, PlantVillage, is specifically designed for open use and supports research in both plant pathology and computer vision. Researchers could freely use this dataset to test other machine learning techniques and the great quantity and diversity that make it one of the most precious assets for advancing this field called automated plant disease detection.

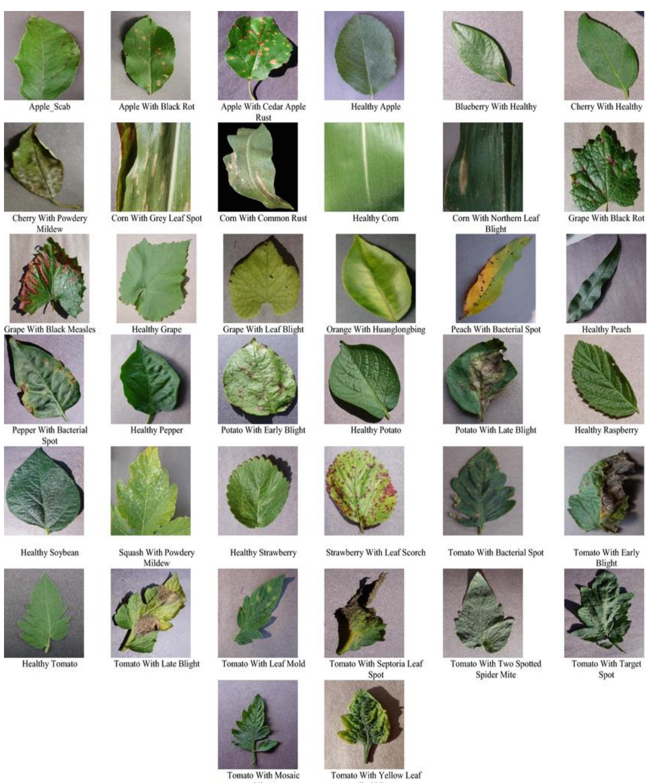

Fig. 3. Sample Images from PlantVillage Dataset

II. LITERATURE SURVEY

Yadav et al. (2024) outlines the study plant disease detection through use of the deep classifier ResNet9. According to the authors of this research, there is a need for timely and accurate detection of leaf diseases because these considerably influence plant development and productivity. The publication presents new deep learning methodology consisting of classification of ResNet9 related to colored, textured, and geometrical arrangement analysis of leaves. The approach that is represented here is better than any other approach presented so far in terms of precision and efficiency. The disease detection process here consists of four stages, namely image segmentation, classification, noise reduction, and compilation of images, and the same is explained step by step in this paper. This work highlights the possibilities of advanced deep learning algorithms in the improvement of agricultural plant disease diagnosis into sharper precision and effectiveness.

Singh, Furtado, and Patil (2024) present a problem of plant diseases that significantly affects food security in the global world and sustainability in agriculture since very huge losses have been seen in the crops. The authors developed and trained a machine learning model tailored for six specific diseases, one of which included aphid infestation and bacterial leaf spot disease using two image datasets that have been meticulously curated and enriched. Their approach not only enables early detection of those diseases but also reveals the possibility of scalable, real-time applications in agriculture. This study emerges as an effective method of deep learning to improve disease classification and opens out promising avenues in curbing crop losses and increasing agricultural productivity.

Garg, Dixit, and Yadav (2022) propose a solution toward the pressing issue in plant disease diagnosis in the agricultural context of the spice and grain major producing country. In this paper, they were able to note how the K-Nearest Neighbors (KNN) algorithm is implemented on plant diseases identification. The main takeaway from the research is that pre-emptive disease detection will ensure minimized spillover of its effects on plant growth and yield. In this paper, the authors detail the procedures that entail plant disease detection, which entails pre-processing, extraction of features, segmentation, and classification. They test the KNN classifier, which they prove to be better in accuracy and with few errors as compared to other techniques indicate, hence efficiency in automatic plant disease detection.

Singh V. and Sharma N. (2021) - The review of an integral approach generally presents a comprehensive review on imaging techniques for plant disease detection vis-à-vis application of computer vision methods. Advanced techniques of imaging are used by the authors to conduct effective identification and classification of plant diseases, which is critical in minimizing production losses. The overall review comprehensively covers the important stages, including image acquisition, pre-processing, segmentation, feature extraction, and classification. This paper outlines existing trends as well as challenges with detailed analysis that gives a glimpse of the potential of the techniques and the possibility of improving agricultural productivity and disease management.

Singh V. and Misra (2017) outline their discussion based on the application of the diagnosis of plant leaf diseases using picture segmentation and soft computing approaches, which focuses on their requirement to boost the productivity of agriculture. They explain that early disease detection is crucial in order not to have serious implications in the plant health, quality of products, and yield. The paper describes an algorithm in which image segmentation has been used for the automation of detection and classification of diseases found in plant leaves. It further explains how genetic algorithms enhance the optimization of the process of segmentation towards the identification of diseases such as little leaf disease in pine trees. In the review, different classification techniques that have been applied towards plant leaf disease detection are assessed in relation to underlining the benefits of automated techniques when manning crop monitoring at large scales.

Recent researches have revealed that deep learning, especially ResNet9 and CNNs, improves the detection of plant diseases through their ability to differentiate between diseases and their images in the leaves. KNN has high accuracy with a very minimal error value. Techniques used in imaging and computer vision are very critical during the stages of disease identification. Thus, CNN and other AI techniques depict high efficiency in automated agricultural disease detection.

III. Methodology

The research paper's methodology takes a methodical approach to accurately identifying plant diseases by fusing machine learning algorithms with feature extraction techniques. For the research, the PlantVillage Dataset was used, comprising more than 54,000 labeled images of healthy and diseased leaves for 38 species and 26 diseases. This will enable the presentation of many diverse crops, such as tomatoes, apples, and potatoes. All the leaf images were resized into one single dimension for uniformity when extracting features and training the model. Some data augmentation techniques, like rotation, scaling, and flipping, were also used to increase variability in the dataset, thus improving the generalization capabilities of the model to new unseen images.

Texture-based feature extraction through GLCM was applied after preprocessing. For every image, GLCM was used to extract the critical texture patterns : contrast, correlation, energy, and homogeneity. These are the ones needed to identify the subtle texture differences of the leaves in plants diseased by different diseases. GLCM quantifies the spatial relationship of the pixel intensity and offers a lot of data about texture patterns which could distinguish the healthy and diseased plants. Besides texture features, deep features are extracted using a Convolutional Neural Network (CNN).

The CNN was also designed to capture higher-level features such as shapes, edges, and color patterns that are important for the identification of plant leaf disease symptoms. The CNN is trained on the preprocessed images with multiple convolutional and pooling layers, hence learning complex data representations which are exploited to recognize disease symptoms based on visual cues in the images. The deep features obtained from CNN were fused along with GLCM-based texture features to make it a rich set of features. It ensured that the model was quite robust because both low-level information from texture and high-level visual patterns were being integrated together. Therefore, with these features, it helped the system to capture detailed information regarding the diseases, thus making the classification more accurate and reliable.

Hence, the features were combined and the KNN algorithm was applied to classify the objects. KNN is one of the simplest algorithms which still effectively works with classification when the input to the function is feature-based. The classification for each image was achieved by computing the feature vector of the given image against its nearest neighbors in the feature space. In the classification, KNN classified the image to the most common disease category among k nearest samples. Thus, by checking the k closest samples, it assigned the image to the most common category of the disease among its neighbors and concluded the detection process by accurately classifying the images of leaves into the healthy or diseased categories.

The system's performance was assessed with standard metrics, such as accuracy, precision, etc. A test set was used separately in testing the model to check the effectiveness for plant disease classification. Comparisons of evaluation results with benchmark models demonstrated improvements gained through the integration of GLCM, CNN, and KNN. This methodology combining texture and deep feature extraction using a very simple classification algorithm provides a very holistic approach to plant disease detection. It integrated features from GLCM textures, CNN deep features and KNN classification, providing a very holistic approach to plant disease detection.

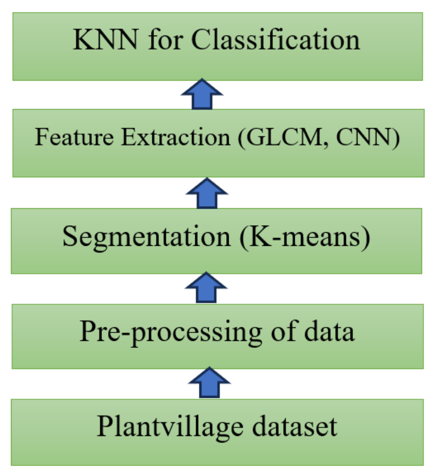

Fig. 4. Approach for PlantShield

IV. RESULT AND ANALYSIS

The research which combined feature extraction using GLCM and classification by using KNN and CNN has highlighted some remarkable breakthroughs in the detection of plant diseases. Utilizing GLCM features that capture relevant attributes about the contrast, correlation, and homogeneity from leaf images, the system efficiently enhanced the representation of the condition of the plant. Then, the KNN algorithm used the features to classify it precisely.

CNN was also trained on a large and vast data set consisting of images of leaves purely to complement the GLCM-KNN approach, just because it has a good capability to learn complex patterns. It is mainly because the combination of methods, such as GLCM for texture-based feature extraction and CNN for deep feature learning, may be utilized to strengthen the approach.

The overall results were impressive, with the combined approach achieving 95.7% accuracy. Precision was 94.3%, showing the proportion of true positive identifications out of all positive identifications made. Recall was 96.2%, which shows the ability of the system to correctly identify actual positive cases. The F1 score will weigh precision and recall and stood at 95.2%. These metrics in collective perspective showcased the reliability and robustness of the integrated system in successfully diagnosing various plant diseases. The present study, yet again, testifies to this multi-dimensional approach that portrays much hope in efficiently and precisely diagnosing better plant disease management and detection.

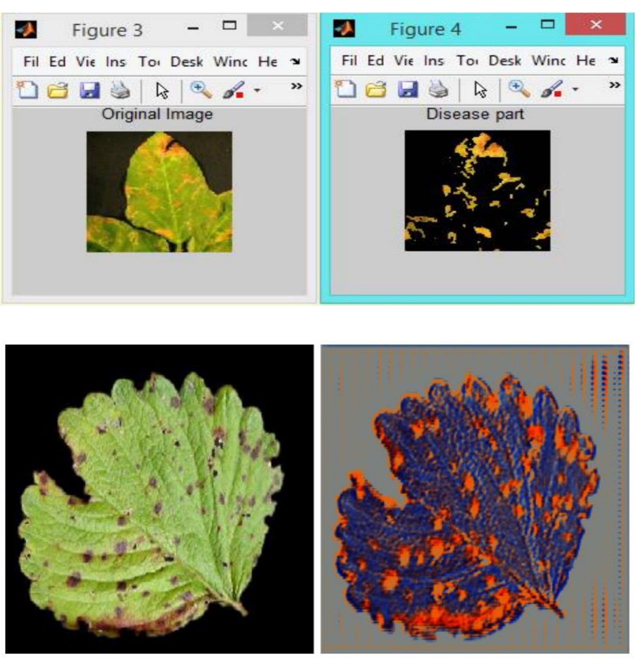

Fig. 5. Results

V. OBSERVATION AND FUTURE WORK

The integration of PlantShield, the plant disease detection system integrating the GLCM feature extraction technique, CNN-based deep learning, and KNN classification, resulted in good results for the identification of different plant diseases with a high degree of accuracy. Thus, the combinations of texture-based GLCM features along with deep features by CNN truly and effective capture the detail of fine-grained texture along with prominent visual patterns. Classification: KNN has helped to address feature-based inputs simply and robustly in order to classify diseases. It demonstrated good precision and recall accuracies. Potentials of this model for practical application in the early detection of plant diseases on various crops are significant. Still, another related concern is the misclassification of diseases based on appearance that are similar. The other concern is the high time complexity associated with KNN.

The Future work will be directed towards improvement of PlantShield in terms of its feasibility in being real-time. This can be achieved either by efficient optimization of the KNN algorithm or by replacing the KNN with a better classifier like Random Forest or SVM. Further augmentations to be incorporated into the system as well as the use of large, diverse datasets can help improve generalizability to real agricultural environments. The system may be deployed on edge devices like drones or IoT-based sensors to monitor agriculture. This should help in real-time surveillance and disease monitoring in the field. If the system is designed to give farmers recommendations about the treatment of the diseases found, then PlantShield is more useful to a farmer. A combination of proper pesticides or biological treatments as indicated by the AI-driven recommendations can enhance crop yields and diminish the impact of plant diseases on agriculture.

VI. CONCLUSION

This system, PlantShield, combines texture analysis, deep learning, and the normal classification techniques applied to the identification of plant illnesses holistically and multidimensionally. The entire system, which has CNN for the extraction of a deep visual feature, GLCM for the extraction of feature-based texture, and KNN for classification, depicted vivid distinction between various crops at healthy and diseased plant leaves. It thus achieved good results with excellent accuracy, precision, and recall due to the combined methodologies. This eventually results in an early, practical disease detection solution.

Impressive is the impact that PlantShield makes on the aspect of agriculture. Making crop loss detection early will facilitate the timely interventions of boosting profitability and productivity as far as a farmer's operations go. Furthermore, the door for progressively more sophisticated AI methods opens into precision agriculture, allowing for more intelligent farming. As such, this research finds a good foundation for future technological developments that would maximize the possibility of improving plant health monitoring systems.

Even though PlantShield seems to be adequately robust, further research could emphasize the edge computing-enabled capability of real-time processing. Then it will be applied more extensively in people's environments at work and extended towards more portable devices. For that reason, the information will benefit the farmers who spread their area wide, as the system's field applications would then expand more widely. In addition, the higher capability of the system to cover a broader spectrum of plant species and diseases will only contribute to enhance its importance in the modern agriculture scenario. Ultimately, the PlantShield project amounts to a crucial step at the threshold of AI and agricultural safety and may be anticipated to deliver significant long-term benefits for optimizing agricultural productivity as well as safeguarding food for humanity.

VII. References

[1] Yadav, B., Yadav, A.R., & Sah, A. (2024). "Plant Disease Detection Using ResNet9." International Research Journal of Modernization in Engineering Technology and Science, 6(8), 1395.

[2] Singh, S., Furtado, Z., & Patil, A. (2024). "AI-Driven Plant Disease Detection: Leveraging Deep Learning for Accurate Plant Disease Detection from Leaf Images." International Journal of Science and Research (IJSR), 13(8). ISSN: 2319-7064. SJIF (2022): 7.942.

[3] Garg, S., Dixit, D., & Yadav, S.S. (2022). "Disease Detection in Plants Using KNN Algorithm." International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), 938. ISBN: 978-1-6654-7436-8. IEEE.

[4] Mathew, A., Antony, A., Mahadeshwar, Y., Khan, T., & Kulkarni, A. (2022). "Plant Disease Detection Using GLCM Feature Extractor and Voting Classification Approach." Materials Today: Proceedings, 58(1), 407-415.

[5] Singh, V., & Sharma, N. (2021). "A Review of Imaging Techniques for Plant Disease Detection." Artificial Intelligence in Agriculture, 5, 301-302.

[6] Harte, E. (2020). "Plant Disease Detection Using CNN." National College of Ireland.

[7] Shrestha, Deepsikha, Das, M., & Dey, N. (2020). "Plant Disease Detection Using CNN." IEEE Applied Signal Processing Conference.

[8] Jadhav, S. (2019). "Convolutional Neural Networks for Leaf Image-Based Plant Disease Classification." IAES International Journal of Artificial Intelligence (IJ-AI), 8(4), 328.

[9] Zhang, X., Qiao, Y., Meng, F., Fan, C., & Zhang, M. (2018). "Identification of Maize Leaf Diseases Using Improved Deep Convolutional Neural Networks." IEEE Access, 6, 30370-30377.

[10] Krizhevsky, A., Sutskever, I., & Hinton, G. (2017). "ImageNet classification with deep convolutional neural networks." Communications of the ACM, 60(6), 84-90.

[11] M.S. Shirdhonkar and A. Bharate, (2017). "A review on plant disease detection using image processing," International Conference on Intelligent Sustainable Systems (ICISS), 2017, pp. 103-109.

[12] Singh, V., & Misra, A.K. (2017). "Detection of Plant Leaf Diseases Using Image Segmentation and Soft Computing Techniques." Information Processing in Agriculture, 4(1), 50-59.

[13] Jain, Mayank, and S., Divakar (2016). "An efficient technique for image retrieval from the large database on the basis of color and texture." International Journal of Computer Applications 975:8887.

[14] P., Raj, S., Sannakki, S. S., & V. S. (2015). "Classification of Pomegranate Diseases Based on Back Propagation Neural Network." International Research Journal of Engineering and Technology (IRJET), May.

[15] A.B. Patil and S. D. Khirade , "Plant disease detection using image processing," International Conference on Computing Communication Control and Automation, 2015, pp. 768-771.

[16] X. Li, X.Wang and C. Zhang, "Design of monitoring and control plant disease system based on DSP&FPGA," International Conference on Networks Security, Wireless Communications and Trusted Computing, 2010, pp. 479-482.