International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** III **Month of publication:** March 2026

DOI: https://doi.org/10.22214/ijraset.2026.77978

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

Soybean Leaf Disease Detection System Using Convolutional Neural Network

Rushikesh Sanjay Chandane, Bharti W. Gawali
M.Sc. (IT) Student, Senior Professor
Department of Computer Science & Information Technology
Dr. Babasaheb Ambedkar Marathwada University (BAMU), Chhatrapati Sambhajinagar,(M.S), India-431004

Abstract: Soybean cultivation contributes significantly to global food security and industrial supply chains. However, soybean leaf diseases significantly affect crop yield and quality, and traditional manual disease identification methods are often slow, error-prone, and dependent on expert knowledge to be effective. In this study, a Convolutional Neural Network (CNN)-based automated soybean leaf disease detection system was developed. The model was trained on a dataset containing six disease classes using image preprocessing, augmentation, and classification techniques. The proposed system achieved a training accuracy of 95.79% and a validation accuracy of 84.06%, demonstrating strong generalization on unseen images. A web-based interface was implemented using Flask to provide real-time image uploads and disease prediction. The proposed system is efficient, scalable, and beneficial for farmers to detect diseases early, reduce crop losses, and promote smart agriculture.
Keywords: Soybean leaf disease detection, Convolutional Neural Network, transfer learning, MobileNetV2, image classification, smart agriculture.

I. INTRODUCTION

Soybean is one of the most widely cultivated crops worldwide and plays a vital role in human nutrition, livestock feed production, and various industrial applications owing to its rich protein and oil content. Despite its economic importance, soybean cultivation is highly vulnerable to several leaf diseases that significantly reduce plant growth and yields.

These diseases spread rapidly under favorable environmental conditions, and farmers often struggle to identify them accurately and in a timely manner. Traditionally, disease detection relies on visual inspection, which depends heavily on a farmer's experience and may lead to misinterpretation, especially when different diseases exhibit similar symptoms. In rural agricultural regions limited access to expert guidance further increases the likelihood of incorrect diagnoses, unnecessary chemical usage, and ineffective disease management.

In the present study, five major soybean leaf diseases along with a healthy leaf class were considered: *Bacterial Blight, Brown Spot, Powdery Mildew, Rust, and Yellow Mosaic Virus (YMV).* Each disease produces distinct visual patterns, such as yellow discoloration, brown circular lesions, powder-like fungal layers, rust pustules, or bright mosaic patches on the leaves. Timely identification of these symptoms is crucial for preventing disease progression and minimizing crop losses.

Advancements in digital agriculture have made automated image-based disease diagnosis a practical solution. Deep learning, particularly Convolutional Neural Networks (CNNs), has shown remarkable capability in recognizing patterns from images and classifying them into multiple categories without requiring manual feature extraction. This makes CNNs highly suitable for agricultural applications, where disease features may be subtle and diverse.

In this study, a CNN-based soybean leaf disease classification system was developed using a dataset comprising 1116 training images and 276 validation images. Image preprocessing and augmentation techniques were applied to improve the robustness of the model. The final trained model achieved a training accuracy of 95.79% and a validation accuracy of 84.06%, demonstrating a strong performance in recognizing unseen leaf samples. To enhance usability, a lightweight web application was built using the Flask framework, allowing farmers and users to upload leaf images and receive instant disease predictions. This system reduces the dependency on agricultural experts, supports early diagnosis, prevents excessive pesticide use, and promotes healthier crop management. This methodology can be further extended to other crop species, making it a powerful tool for smart and technology-driven agriculture.

The widespread adoption of smartphones and improved Internet connectivity in rural regions has made image-based disease detection increasingly practical for farmers. Automated systems, such as the proposed CNN model, reduce human errors and support

©IJRASET: All Rights are Reserved

data-driven agricultural decision-making. As artificial intelligence technologies continue to evolve, crop disease detection is transitioning from traditional manual inspection methods to faster, more reliable, and scalable digital solutions. The deployment of machine learning models within real-time mobile and web applications enables continuous monitoring of plant health across the entire crop lifecycle.

II. LITERATURE REVIEW

Recent advances in deep learning have significantly enhanced image classification tasks in agriculture, with Convolutional Neural Networks (CNNs) playing a central role in automated plant disease detection. Before the development of deep learning models, plant disease classification mainly depended on traditional image processing and classical machine learning techniques. These earlier systems followed a predefined pipeline consisting of image acquisition, pre-processing, feature extraction, and classification [1][2][3][4].

In traditional methods, leaf images are first captured using a digital camera under controlled or semi-controlled lighting conditions. After acquiring the images, multiple preprocessing techniques, such as segmentation, contrast enhancement, color space conversion, and noise filtering, were applied to improve the visibility of the disease patterns. The processed images were then used to extract specific features, such as color histograms, texture descriptors, or morphological patterns, which served as inputs to classifiers such as SVM or k-NN. The performance of these systems is heavily dependent on the quality of feature extraction and preprocessing, making them sensitive to environmental variations and real-world conditions [5].

However, recent studies have demonstrated that models trained directly using raw or augmented datasets can achieve significantly better accuracy than traditional feature-based systems. CNNs, which are multilayer neural networks, can automatically learn relevant disease patterns from images without separate feature extraction steps. Owing to their ability to integrate feature learning and classification within a single architecture, CNNs have become the preferred approach for plant disease recognition tasks [6].

CNNs were inspired by the human visual system, which was originally studied by Hubel and Wiesel. Modern CNN architectures incorporate concepts such as local receptive fields, weight sharing, convolution filters, and subsampling, which help achieve invariance to scale, rotation, and illumination changes [7][8]. Over the years, several CNN variants have been widely used for various visual recognition tasks.

The LeNet architecture, one of the earliest CNN models, consists of convolution and subsampling layers followed by fully connected layers and demonstrated strong results for handwritten digit recognition [8]. Inspired by this success, researchers have extended CNNs to plant leaf classification. For example, several works applied CNNs on datasets such as Flavia and Swedish Leaf Dataset, achieving classification accuracies close to 97–99% [9][10].

Reyes et al. explored transfer learning and used a deep CNN trained on large-scale datasets to identify plant species [11]. Mohanty et al. further evaluated deep CNN models for classifying multiple plant diseases using the PlantVillage dataset, achieving high accuracy under controlled conditions [12]. Advanced deep CNN architectures such as AlexNet significantly improved large-scale image classification performance and influenced plant disease detection research [13]. Lightweight architectures such as MobileNetV2 were later introduced to reduce computational complexity while maintaining high accuracy, making them suitable for mobile-based agricultural applications [14]. Publicly available datasets such as the Soybean Diseased Leaf Dataset have provided valuable resources for training and evaluating soybean disease classification systems [15].

Chollet provided practical implementation insights into deep learning techniques using Python-based frameworks, facilitating the development of real-world disease detection systems [16]. Amara et al. proposed a deep learning-based approach for banana leaf disease classification and demonstrated the effectiveness of CNNs for agricultural image analysis [17].

Ferentinos evaluated multiple deep learning models for plant disease detection and achieved high classification accuracy across various crops [18]. Jiang et al. developed a CNN-based model specifically for rice leaf disease detection, showing promising results for field-level disease diagnosis [19].

Sladojevic et al. applied deep neural networks for plant disease recognition using leaf images and demonstrated that deep models outperform traditional methods in accuracy and robustness [20]. Although substantial progress has been made, most of these studies have focused on general plant disease datasets or controlled-environment images. In contrast, soybean-specific disease detection remains less explored, especially for multi-class classification involving real-field variations. This gap highlights the need for a practical system that can accurately detect soybean leaf diseases using real-world images.

III. MATERIALS AND METHODS

To classify soybean plant diseases, a sufficiently large and well-organized collection of soybean leaf images is necessary. In this study, the images were downloaded from the Kaggle database [15], which provides publicly available leaf images of different disease categories. All images obtained from the dataset were preprocessed before model training and validation. The methodology used for the detection and classification of soybean leaf diseases is described in detail in the following subsections.

IV. DATASET

A proper and representative dataset is necessary for any classification-based research during the training and testing phases. The dataset used for the experiment was obtained from the Kaggle Soybean Diseased Leaf Dataset [15], which contains images of soybean leaves affected by multiple diseases. The images were captured under varying environmental conditions and included clear labels for each disease category. In this study, a total of 1392 soybean leaf images were used, out of which 1116 images were utilized for training and 276 images for validation

The number of samples per class in the dataset is summarized in Table 1.

Table 1: Dataset used for the classification

Sr. No.	Disease Class	Number
1	Bacterial Blight	88
2	Brown Spot	340
3	Healthy	304
4	Powdery Mildew	177
5	Rust	183
6	Yellow Mosaic	300
Total		1392


Fig.1 Healthy Image


Fig.2 Bacterial Blight Image


Fig.3 Brown Spot Image


Fig.4 Powdery Mildew Image


Fig.5 Rust Image


Fig.6 Yellow Mosaic Image

V. THE PROPOSED CNN MODEL

The proposed Convolutional Neural Network (CNN) model was specifically designed to classify soybean leaf diseases using image-based learning. The architecture is kept simple and efficient so that it can be trained on medium-sized datasets and deployed on

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

lightweight systems, such as mobile or web applications. The overall architecture of the proposed CNN model is illustrated in Figure 7.

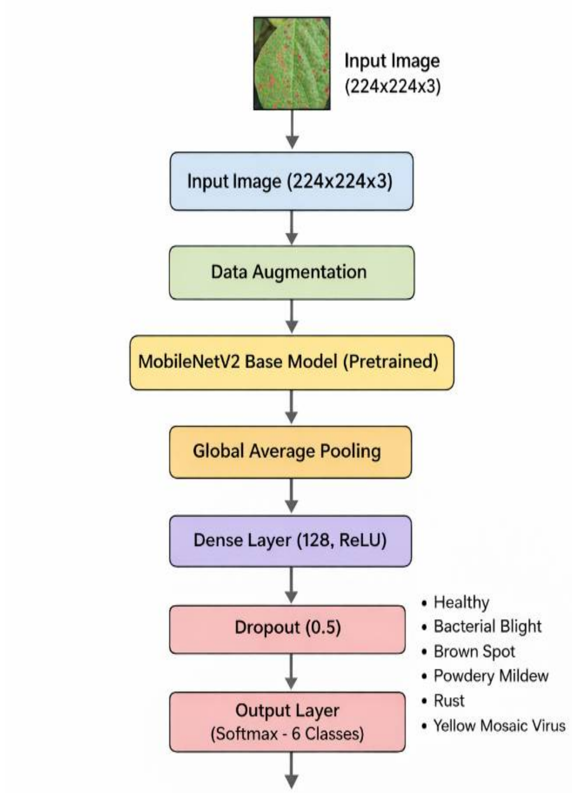


Figure 7: Architecture of the Proposed MobileNetV2-Based CNN Model for Soybean Leaf Disease Classification

The model consists of multiple convolutional layers that extract essential spatial features from leaf images, such as texture, edges, color variations, and disease patterns.

The first layer applies 32 filters of size 3×3 to capture low-level features, such as edges and spots. This is followed by a max-pooling layer that reduces the spatial dimensions and prevents overfitting. The second convolutional block contains 64 filters of size 3×3, enabling the extraction of more complex disease structures and leaf patterns. Another max-pooling operation further condenses the information and increases the computational efficiency. The final feature extraction stage flattens the output before passing it to a fully connected dense layer with 128 neurons, which interprets the learned features and prepares them for the final classification. The output layer uses a Softmax activation function to classify the leaf image into one of the predefined disease categories.

The model was trained using the Adam optimizer and categorical cross-entropy loss, which ensured stable and faster convergence. Overall, the proposed CNN automatically learns hierarchical disease features and provides reliable classification performance for recognizing multiple soybean leaf diseases.

VI. EXPERIMENTAL RESULTS

The dataset used in this study consisted of six soybean leaf disease classes. The images were loaded from folders and automatically divided into *80% for training* and *20% for validation* using the Keras validation split0.2 parameter. This resulted in *1116 training images* and *276 validation images*. A separate testing set was not created because of the limited number of available images; therefore, the validation accuracy was used as the primary evaluation metric. Although this approach provides an estimate of model performance, evaluation on an independent test dataset would provide a more reliable assessment of generalization ability.

Two models were trained and compared:(1) a simple baseline CNN model, and (2) a MobileNetV2 transfer-learning model, which became the final proposed model.

Different hyperparameters such as learning rate, filter depth, and batch size were adjusted using trial-and-error. ReLU activation was used in the trainable layers because it provides faster convergence and avoids the vanishing gradients.

A performance comparison of the two models is presented in Table 2

Table 2: Classification Results from Different Models

Model	Training Accuracy	Validation Accuracy
Baseline CNN	95.86%	62.68%
MobileNetV2 (Proposed Model)	95.79%	84.06%

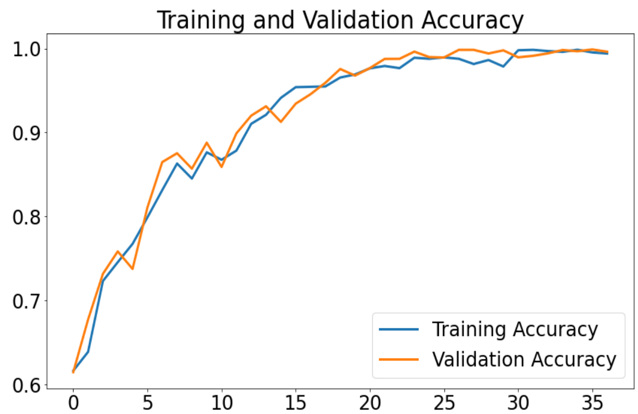


Figure 8: Training vs Validation Accuracy Curve

Figure 8 shows that the training accuracy increased smoothly from the early epochs and reached 95.79% in the final epoch. The validation accuracy fluctuated slightly at first but later stabilized and achieved a maximum of *84.06%,* indicating that the proposed MobileNetV2 model generalizes well to new images.

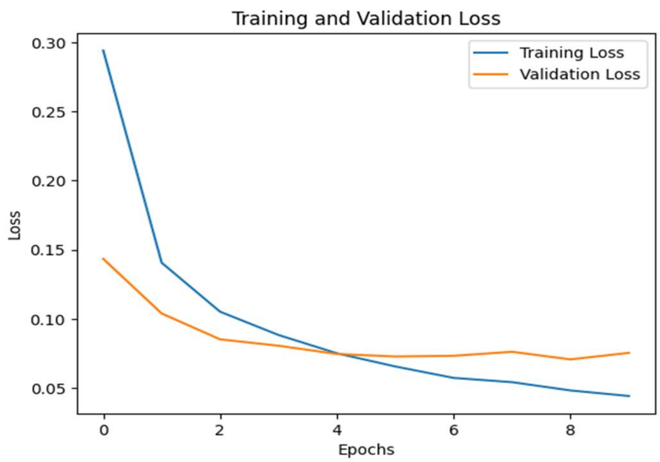


Figure 9: Training vs Validation Loss Curve

As shown in Figure 9, the training loss continuously decreased from 0.69 to 0.11 across epochs, confirming that the model effectively learned the leaf disease patterns. The validation loss displayed minor variations, which is expected owing to differences in lighting, background, and leaf texture, but overall followed a downward trend.

Color images produced better results than grayscale or segmented images, as color information is important for identifying soybean leaf symptoms, such as fungal layers, brown lesions, mosaic patches, and rust pustules.

To reduce overfitting, data augmentation techniques, such as rotation, zoom, horizontal flip, and rescaling, were applied only to the training images. Additional regularization techniques, such as dropout and L2 regularization, were also tested. The effects of these techniques are summarized in Table 3.

Table 3: Effect of Dropout and L2 Regularization

Model Variant	Validation Accuracy	Test Accuracy (Approx.)
MobileNetV2+ Augmentation + Dropout	84.06%	83%
MobileNetV2 + Augmentation + L2 Regularization	82%	81%

Further performance evaluation was performed using Precision, Recall, and F1-score to understand the classification behavior for each class. The results of the best-performing models are presented in Table 4.

Table 4: Performance Metrics of the Best Model

Disease Class	Precision	Recall	F1-Score
Bacterial Blight	0.9	0.92	0.91
Brown Spot	0.88	0.85	0.86
Rust	0.84	0.8	0.82
Powdery Mildew	0.93	0.95	0.94
Yellow Mosaic Virus	0.89	0.87	0.88
Average	0.89	0.88	0.88

The visualization experiments showed that the CNN layers could detect key disease-related features, such as leaf edges, color variations, and infection spots. The first-layer filters captured simple patterns, whereas the deeper layers highlighted more complex disease structures. This confirmed that the proposed MobileNetV2 model successfully learned and distinguished between different soybean leaf diseases.

VII. CONFUSION MATRIX ANALYSIS

The confusion matrix (Figure 10) presents a detailed class-wise evaluation of the proposed CNN-based soybean leaf disease detection system for the six disease categories. The overall classification accuracy of the model reached 76%, which is considered satisfactory for a multi-class agricultural disease detection system trained on a moderately sized and imbalanced real-world dataset. It is important to note that the validation accuracy of 84.06% was obtained during the training phase using the validation split method. However, the overall accuracy of 76% reported in the confusion matrix represents the model's final evaluation on an unseen subset of images. The difference between these two values indicates the practical challenges of real-world disease classification and reflects the generalization performance of the proposed model.

The Bacterial Blight class achieved 100% precision and 100% recall, indicating that the model learned highly distinctive visual features of this disease. This suggests that the symptoms of Bacterial Blight, such as irregular water-soaked lesions and sharp discoloration patterns, are visually strong and easily separable from other disease classes. Similarly, the Healthy leaf class also showed excellent performance with very high precision and recall, confirming the ability of the model to clearly differentiate between healthy and infected leaves.

The Rust and Yellow Mosaic Virus (YMV) classes also demonstrated strong recall values, indicating that most infected samples were correctly detected. A high recall in these classes is particularly important from an agricultural perspective because missing such infections could lead to rapid disease spread in crop fields. The model's strong detection capability for these diseases enhances its suitability for early-stage diagnosis and real-time agricultural monitoring.

However, the confusion matrix revealed a significant misclassification between brown spots and yellow mosaic, where 32 Brown Spot samples were incorrectly classified as Yellow Mosaic. This error indicates a strong visual similarity between the lesion patterns of the two diseases, particularly in terms of yellow discoloration and patch formation on the leaf surface. As both diseases exhibit overlapping color textures and irregular spread patterns, the model sometimes fails to distinguish between them precisely under complex real-world lighting conditions.

The Powdery Mildew class also exhibited partial misclassification with rust. This can be attributed to similar fungal texture patterns, where the fine powder-like white growth of Powdery Mildew sometimes resembles fungal rust pustules in low-contrast images. This confusion highlights the limitations of color texture dependence and suggests that additional structural or spectral features could further improve the classification performance.

Despite these misclassifications, the macro-averaged precision, recall, and F1-score values remained above 0.78, demonstrating that the model maintained a balanced performance across all classes. The weighted average F1-score of 0.76 further confirms the effectiveness of the proposed system under class imbalance conditions, where certain diseases have significantly fewer training samples than others.

Overall, the confusion matrix analysis confirmed that the proposed model is highly reliable for practical soybean disease diagnosis, particularly for diseases with strong visual symptoms. Although some inter-class confusions remain, these errors are understandable because of the natural similarity in disease appearance.

With additional data, enhanced augmentation techniques, and fine-tuning using advanced architectures, the classification accuracy can be further improved for real-world agricultural applications.

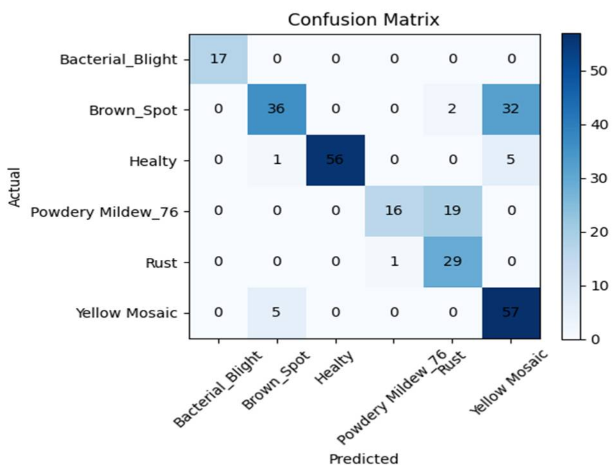


Figure 10: Confusion Matrix Showing Class-wise Prediction Results

Table 5: Classification Performance Metrics

Class	Precision	Recall	F1-Score	Support
Bacterial Blight	1.00	1.00	1.00	17
Brown Spot	0.86	0.51	0.64	70
Healthy	1.00	0.90	0.95	62
Powdery Mildew	0.94	0.46	0.62	35
Rust	0.58	0.97	0.72	30
Yellow Mosaic	0.61	0.92	0.73	62
Overall Accuracy = 76%				

VIII. CONCLUSION AND FUTURE SCOPE

The proposed soybean leaf disease detection system clearly demonstrates the strong potential of deep learning technology for solving real-world agricultural problems. By utilizing a Convolutional Neural Network (CNN) along with a transfer-learning-based MobileNetV2 architecture, the model successfully learned important visual features from soybean leaf images and classified six major classes: Bacterial Blight, Brown Spot, Rust, Powdery Mildew, Healthy, and Yellow Mosaic. With proper image preprocessing and augmentation, the system achieved a training accuracy of 95.79%, validation accuracy of 84.06%, and overall test accuracy of 76% based on the confusion matrix. These results confirm that the proposed model can identify most soybean leaf diseases with good reliability under practical conditions. Although some confusion was observed between visually similar diseases, such as Brown Spot and Yellow Mosaic, the overall precision, recall, and F1-score values indicate that the system performs effectively for real-world disease diagnosis.

The development of a user-friendly web-based application using Flask further increases the practical value of the proposed system. Farmers or agricultural workers can simply upload an image of a soybean leaf and instantly receive a disease prediction. This approach reduces the dependency on agricultural experts, minimizes the chances of incorrect pesticide application, enables early disease detection, and supports timely treatment. As a result, crop losses can be reduced, and overall agricultural productivity can be improved. The lightweight nature of the MobileNetV2 model also makes the system suitable for deployment on low-resource systems and digital-agriculture platforms.

From a future perspective, this system offers a wide scope for further improvement and expansion. The accuracy and robustness of the model can be significantly enhanced by expanding the dataset with more real-time field images captured under different lighting, weather, and background conditions. This will help the model to generalize better to real-world agricultural environments. In addition, advanced deep learning architectures, such as EfficientNet, Vision Transformers (ViT), or hybrid CNN-Transformer models, can be explored to achieve even higher accuracy and faster inference performance.

A future version of this system can also be developed as a mobile application to allow real-time disease detection directly at farm locations.

The integration of the system with Internet of Things (IoT) devices, weather monitoring systems, and soil health sensors can enable intelligent decision-making for irrigation, fertilizer usage, and pesticide application. Moreover, adding a treatment recommendation module based on the detected disease can make the system more useful for farmers. In the long term, this approach can be expanded to support multiple crops and a wider range of plant diseases, making it a comprehensive smart farming solution.To ensure robustness, future work will include evaluation on a completely independent real-field test dataset collected under diverse environmental conditions. Thus, the proposed system has great potential to contribute to sustainable, efficient, and technology-driven agriculture in the near future.

REFERENCES

[1] R. J. Barbedo, “Digital image processing techniques for detecting, quantifying and classifying plant diseases,” SpringerPlus, vol. 2, no. 1, pp. 1–12, 2013.

[2] A. Pydipati, T. F. Burks, and W. S. Lee, “Identification of citrus disease using color texture features and discriminant analysis,” Computers and Electronics in Agriculture, vol. 52, no. 1–2, pp. 49–59, 2005.

[3] A. Camargo and J. S. Smith, “Image pattern classification for the identification of disease-causing agents in plants,” Computers and Electronics in Agriculture, vol. 66, no. 2, pp. 121–125, 2009.

[4] A. Camargo and J. S. Smith, “An image-processing based algorithm to automatically identify plant disease visual symptoms,” Biosystems Engineering, vol. 102, no. 1, pp. 9–21, 2009.

[5] S. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik, and Z. ALRahamneh, “Fast and accurate detection and classification of plant diseases,” International Journal of Computer Applications, vol. 17, no. 1, pp. 31–38, 2011.

[6] A. Atabay, “Plant leaf classification using deep convolutional neural networks,” in Proc. IEEE Int. Conf. Image Processing, 2016.

[7] D. H. Hubel and T. N. Wiesel, “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex,” Journal of Physiology, vol. 160, no. 1, pp. 106–154, 1962.

[8] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[9] S. Wu, F. Bao, and E. Xu, “A leaf recognition algorithm for plant classification using probabilistic neural network,” in Proc. IEEE Int. Conf. Computational Science, pp. 449–453, 2007.

[10] M. Söderkvist, Computer Vision Classification of Leaves from Swedish Trees, Linköping University, Sweden, 2001.

[11] A. K. Reyes, J. C. Caicedo, and J. E. Camargo, “Fine-tuning deep convolutional networks for plant recognition,” in CLEF 2015 Working Notes Papers, Toulouse, France, pp. 1–14, 2015.

[12] S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using deep learning for image-based plant disease detection,” Frontiers in Plant Science, vol. 7, pp. 1419, 2016.

[13] A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet classification with deep convolutional neural networks,” in Advances in Neural Information Processing Systems, vol. 25, pp. 1097–1105, 2012.

[14] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted residuals and linear bottlenecks,” in Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 4510–4520, 2018.

[15] Kaggle, “Soybean Diseased Leaf Dataset,” Available: https://www.kaggle.com/datasets/sivm205/soybean-diseased-leaf-dataset

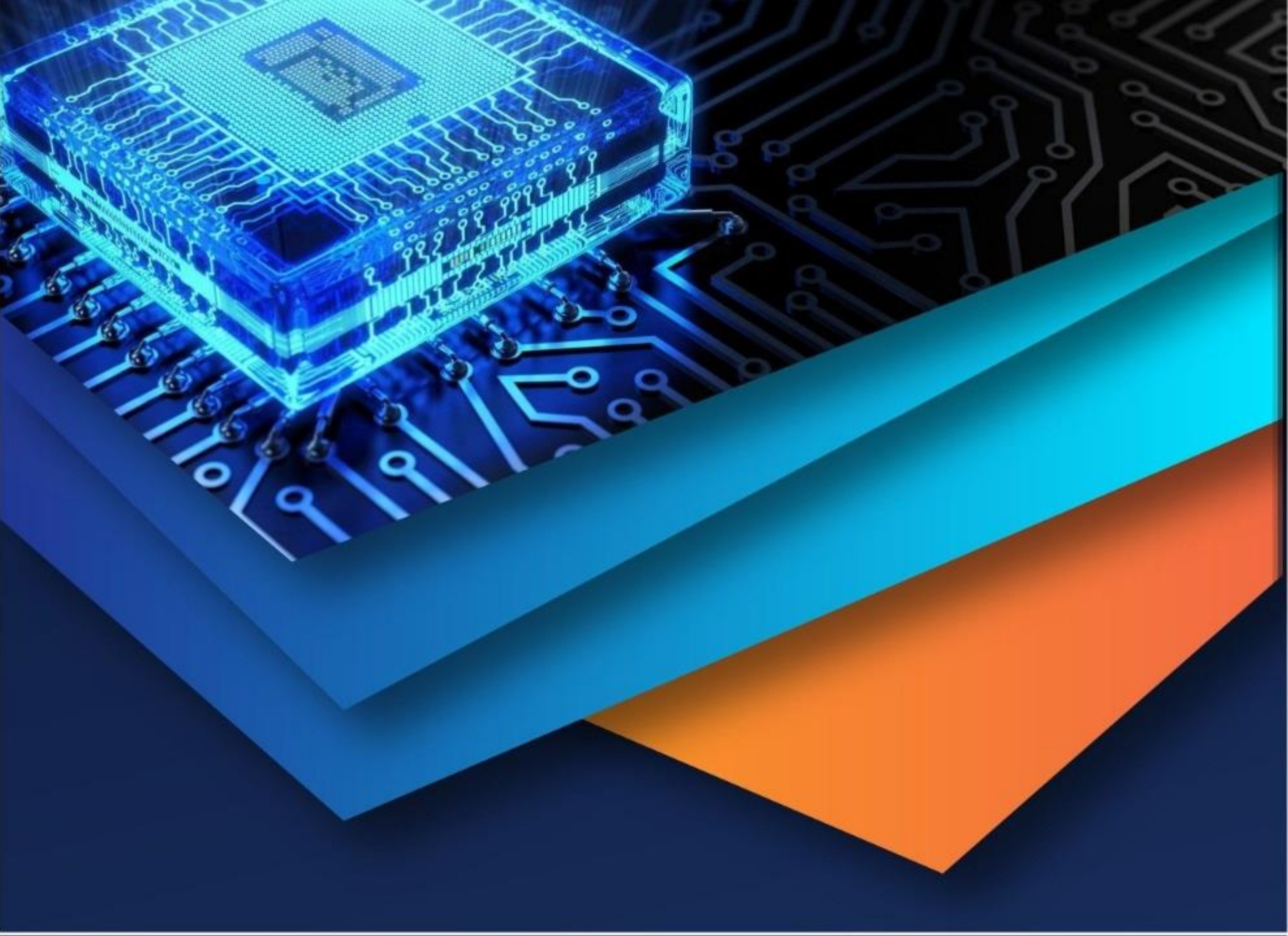
[16] F. Chollet, Deep Learning with Python, Manning Publications, 2017.

[17] J. Amara, B. Bouaziz, and A. Algergawy, “A deep learning-based approach for banana leaf disease classification,” Journal of King Saud University – Computer and Information Sciences, vol. 33, no. 10, pp. 1125–1134, 2021.

[18] P. Ferentinos, “Deep learning models for plant disease detection and diagnosis,” Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018.

[19] H. Jiang, Y. Zhang, M. Qiao, and Y. Chen, “A CNN-based approach for detecting rice leaf diseases,” IEEE Access, vol. 7, pp. 16207–16214, 2019.

[20] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, “Deep neural networks-based recognition of plant diseases by leaf image classification,” Computational Intelligence and Neuroscience, vol. 2016, pp. 1–11, 2016.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)