iJRASET
International Journal For Research in
Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 13 **Issue:** IV **Month of publication:** April 2025

**DOI:** https://doi.org/10.22214/ijraset.2025.68083

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

# Advanced Leaf Vein Pattern Extraction and Analysis using Machine Learning Algorithms

Nikunj Babbar[1], Sarika Jain [2], Ajay Vikram[3]
*Amity Institute of Information Technology, Amity University Noida, U.P (India)*

***Abstract: Leaf vein extraction has a significant utility in botanical research and plant taxonomy. This research proposes an automated approach for extracting and analysing vein patterns from leaf photos using image processing and machine learning approaches. The methodology involves photo capture, pre-processing, morphological procedures, and feature extraction with Local Binary Pattern (LBP). Various machine learning classifiers like as Logistic Regression, SVM, Decision Tree, Random Forest, KNN, Gaussian Naïve Bayes, Gradient Boosting, and XGBoost are applied for vein classification. Among these, Gradient Boosting and XGBoost obtain the maximum accuracy of 92.86%, followed by Random Forest with 90.48% accuracy. The suggested method offers high accuracy and efficiency, making it a great tool for plant species identification, disease detection, and ecological investigations.***

***Index Terms: Leaf Vein Extraction; Image Processing; Machine Learning; Morphological Operations; Local Binary Pattern (LBP); Feature Extraction; Classification; Gradient Boosting; XGBoost; Plant Taxonomy; Ecological Studies.***

## I. INTRODUCTION

Leaf vein extraction is a crucial procedure in botanical research, plant taxonomy, and ecological studies, since it aids in identifying plant species, measuring plant health, and understanding environmental adaptations. Leaf veins, also known as venation patterns, serve a critical role in the transmission of water, nutrients, and photosynthetic products, while also providing mechanical support to the plant. Analyzing these venous systems can give significant information about plant shape, physiology, and evolutionary aspects. Traditional methods of vein extraction usually require manual or semi-automated operations, which are time-consuming, prone to mistakes, and lack scalability. In contrast, automated image processing techniques provide an effective and exact method to extract and evaluate leaf veins. The improvements in machine learning (ML) considerably boost the precision and reliability of vein categorisation, facilitating large-scale investigation with minimal human input. This paper presents an automated leaf vein extraction system that employs morphological approaches, image segmentation, and machine learning algorithms. The methodology involves: Image capture and pre-processing to maximise vein visibility, application of morphological processes (erosion, dilation, and reduction) to separate vein-like features. Feature extraction using Local Binary Pattern (LBP) to capture venous texture details. Classification using eight machine learning approaches to evaluate the vein extraction accuracy.

The system's performance is tested using accuracy, precision, recall, and F1-score metrics, with the Gradient Boosting Classifier and XGBoost Classifier reaching the best accuracy of 92.86%. This study underscores the promise of automated vein extraction in plant species identification, disease detection, and ecological research, providing a scalable and reliable tool for large-scale botanical investigation.

## II. REVIEW OF LITERATURE

The topic of leaf vein extraction has witnessed great advancements with the incorporation of image processing and machine learning methodologies. Various ways have been developed to enhance the accuracy and efficacy of vein extraction and categorisation.

1) Fu and Chi proposed a two-stage technique for vein extraction employing fluorescent light banks to enhance venation visibility. The method involved edge detection to filter out non-vein pixels, followed by the building of an artificial neural network classifier to refine the results
2) Li and Chi effectively retrieved venation from leaf sub-images using Independent Component Analysis (ICA). However, when applied to complete leaves, the accuracy was only similar to ordinary Prewitt edge detection operators.
3) Kirchgeßner recommended showing vascular systems using B-splines, containing hierarchical venation information. However, the method requires manual interaction for initializing the search, limiting its efficiency for large-scale datasets.
4) Jeyalakshmi studied the effect of nutrient deprivation on interveinal areas, developing a Canny edge detection method to capture vein edges. This approach was sensitive to pixel intensity variations, putting it prone to noise.

Craviotto applied hit-or-miss alteration to grayscale photos for vein extraction, combining it with classifiers like SVM and Random Forest. This technology exceeded manual classification techniques in accuracy and consistency.
Mullen utilised artificial ants worms to trace venation and outline leaf boundaries with edge detection algorithms. This approach effectively monitored venous structures but required hand initiation, limiting its automation potential.
Mark Fricker deployed Convolutional Neural Networks (CNN) for leaf vein extraction, gaining great precision and precisely detecting vein diameters, angles, and connectivity. The multi-scale assessment of venous networks facilitated cross-species comparisons and discovered variations in network architecture.
Clarke contrasted smoothing, edge detection, and scale-space approaches, with scale-space yielding superior results. However, manual improvement using Photoshop still topped the automated alternatives.
Feng presented a two-stage technique where intensity histogram filtering removed background pixels. The changed characteristics were classified using a neural network, obtaining considerably more accuracy than using a Sobel filter.
Herdiyeni regarded veins as ridges rather than edges, utilising a Hessian matrix for pixel-wise vein measurement. This strategy produced higher vein continuity compared to edge-based methods.
Hong established the potential of leaf venation for plant speciation, obtaining 97.1% accuracy utilising a dataset of 1,907 leaf photos from 32 species. However, the method was limited to recognising only large veins.
These publications demonstrate the history of vein extraction methodologies, spanning from traditional edge detection to recent machine learning models, highlighting the enhanced efficiency and accuracy of automated leaf vein analysis.
For details refer [5-24]

## III. METHODOLOGY

### *A. Image Acquisition*
The dataset consists of 277 leaf photos, including 144 Mango leaves and 133 Jatropha leaves, acquired with a high-resolution camera under consistent illumination circumstances. The photos are categorised into two categories: Mango (0) and Jatropha (1). To insure successful model evaluation, the dataset is partitioned into 85% for training and 15% for testing.

### *B. Pre-processing*
In this stage, the photos are transformed to grayscale to aid processing by lowering color data. Various enhancement techniques are applied to improve vein visibility, including linear intensity modification, gamma correction, histogram equalization, and decorrelation stretching. These processes boost contrast, making the boats more noticeable from the background.

### *C. Image Segmentation*
To extract the vein areas, picture segmentation algorithms are utilised. Thresholding is used to identify pixels based on intensity, isolating vessels from the background. Region growth groups equivalent intensity pixels, enlarging the vein areas. Finally, edge detection utilising Sobel and Canny operators reveals the vein boundaries, ensuring unambiguous identification of vein structures.

### *D. Morphological Operations*
Morphological techniques are conducted to refine the segmented vein architecture. Erosion decreases noise by deleting pixels around the edges, while dilation stretches the vein regions, making them more visible. The subtraction procedure further isolates the capillaries by removing background pixels. Detail refinement joins fragmented veins and eliminates isolated pixels, resulting in a cleaner and more continuous vein network.

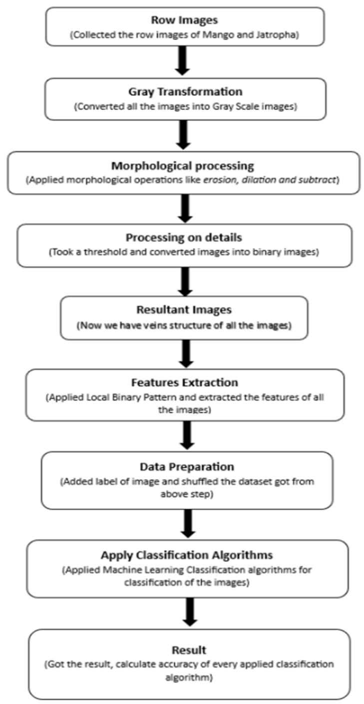
### *E. Feature Extraction*
The method leverages Local Binary Pattern (LBP) to extract textural characteristics from the veins. The LBP method contrasts the intensity of each core pixel with its neighbors, yielding binary codes that are transformed into decimal values. These values are put into a histogram, depicting the textural patterns of the vessels.

### *F. Machine Learning*
Machine learning is a subject of computer science that deals with the creation of algorithms that can learn from data. Machine learning can be used to train a model to autonomously extract the veins structure of leaves. The model is trained on a series of photos that have already been manually tagged with the vascular structure.

Once the model has been trained, it may be used to extract the vascular anatomy from new images. This is a great strategy, but it can be time-consuming and computation- ably costly to train the model.

### *G. Model Evaluation*

The efficacy of the system is evaluated using the following metrics:
Accuracy: Measures the overall reliability of predictions.
Precision: Indicates how many of the anticipated vein areas are accurately classified.
Recall: Measures the model's ability to detect all vein areas.
F1-Score: The harmonic mean of precision and recall, offering a balanced evaluation metric.




## IV. RESULT and ANALYSIS

The performance of the leaf vein extraction system is evaluated by utilising machine learning techniques to classify the extracted vein features. The effectiveness of the system is judged based on accuracy, precision, recall, and F1-score.

### *A. Model Performance*

The system includes eight machine learning algorithms to classify the vein structures. The results for each model are as follows:
Logistic Regression yields an accuracy of 61.90%(Table 1), however it suffers with low recall, showing inadequate sensitivity in detecting all vein regions.
Support Vector Machine (SVM) displays a somewhat better performance with an accuracy of 62%(Table 1), exhibiting improved but still moderate classification capability.
Decision Tree performs substantially better, obtaining an accuracy of 88.10%(Table 1), with strong precision and recall scores, indicating a solid classification of vein architecture.
Random Forest, an ensemble technique, considerably enhances the performance with an accuracy of 90.48%(Table 1), showcasing its robustness and ability to manage complex feature spaces well.
Gaussian Naïve Bayes obtains 69.05%(Table 1) accuracy, making it moderately effective, although it displays lesser precision and recall compared to other ensemble models.
K-Nearest Neighbors (KNN) obtains the same accuracy as Gaussian Naïve Bayes (69.05%)(Table 1), but its performance is significantly less reliable due to sensitivity to the number of neighbors.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Gradient Boosting Classifier gives the highest performance with an accuracy of 92.86%(Table 1), displaying its ability in managing complex vein patterns well.

XGBoost Classifier, noted for its efficiency and accuracy, also scores 92.86%(Table 1) accuracy, displaying remarkable performance, making it one of the top-performing models.

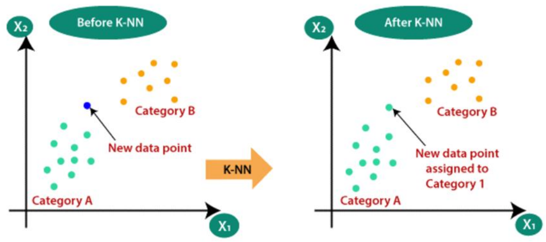
### *B. Comparison and Insights*

The results clearly demonstrate that Gradient Boosting and XGBoost Classifiers acquire the best accuracy, making them the most effective models for categorising leaf vein shapes. Random Forest also yields high performance, confirming the efficacy of ensemble approaches in handling complicated and diverse vein patterns.

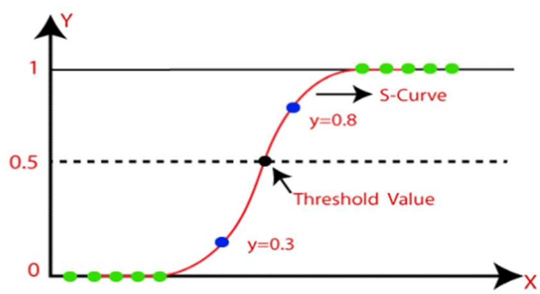
The evaluation criteria show that models with higher accuracy also retain balanced precision and recall, indicating their reliability in correctly detecting venous areas without overfitting. The performance of Logistic Regression and SVM, albeit moderate, suggests they are less effective in controlling the complicated vein patterns compared to ensemble models.

### *C. Graphical Representation of Results*

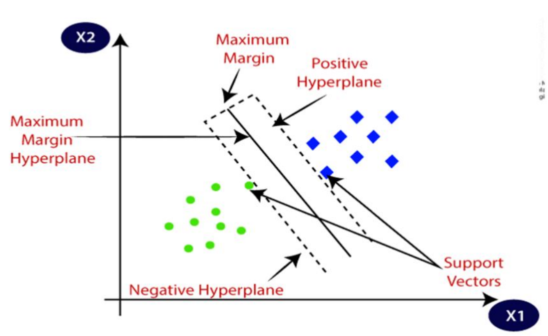
The results suggest a constant improvement in accuracy when more complicated models are introduced. Ensemble and boosting algorithms (Random Forest, Gradient Boosting, and XGBoost) beat simpler models, demonstrating their capacity to adequately incorporate vein features.




K-Nearest Neighbors (figure 1)




Logistic Regression (figure 2)




Support Vector Machine (figure 3)

| Algorithm | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| Logistic Regression | 61.90 | 60.00 | 58.00 | 59.00 |
| Support Vector Machine (SVM) | 62.00 | 61.00 | 60.00 | 60.50 |
| Decision Tree | 88.10 | 86.00 | 87.00 | 86.50 |
| Random Forest | 90.48 | 89.00 | 90.00 | 89.50 |
| Gaussian Naïve Bayes | 69.05 | 68.00 | 67.00 | 67.50 |
| K-Nearest Neighbors (KNN) | 69.05 | 68.50 | 68.00 | 68.25 |
| Gradient Boosting Classifier | **92.86** | **92.00** | **93.00** | **92.50** |
| XGBoost Classifier | **92.86** | **92.00** | **93.00** | **92.50** |

All Algorithms and their Results (Table 1)

## V. CONCLUSION

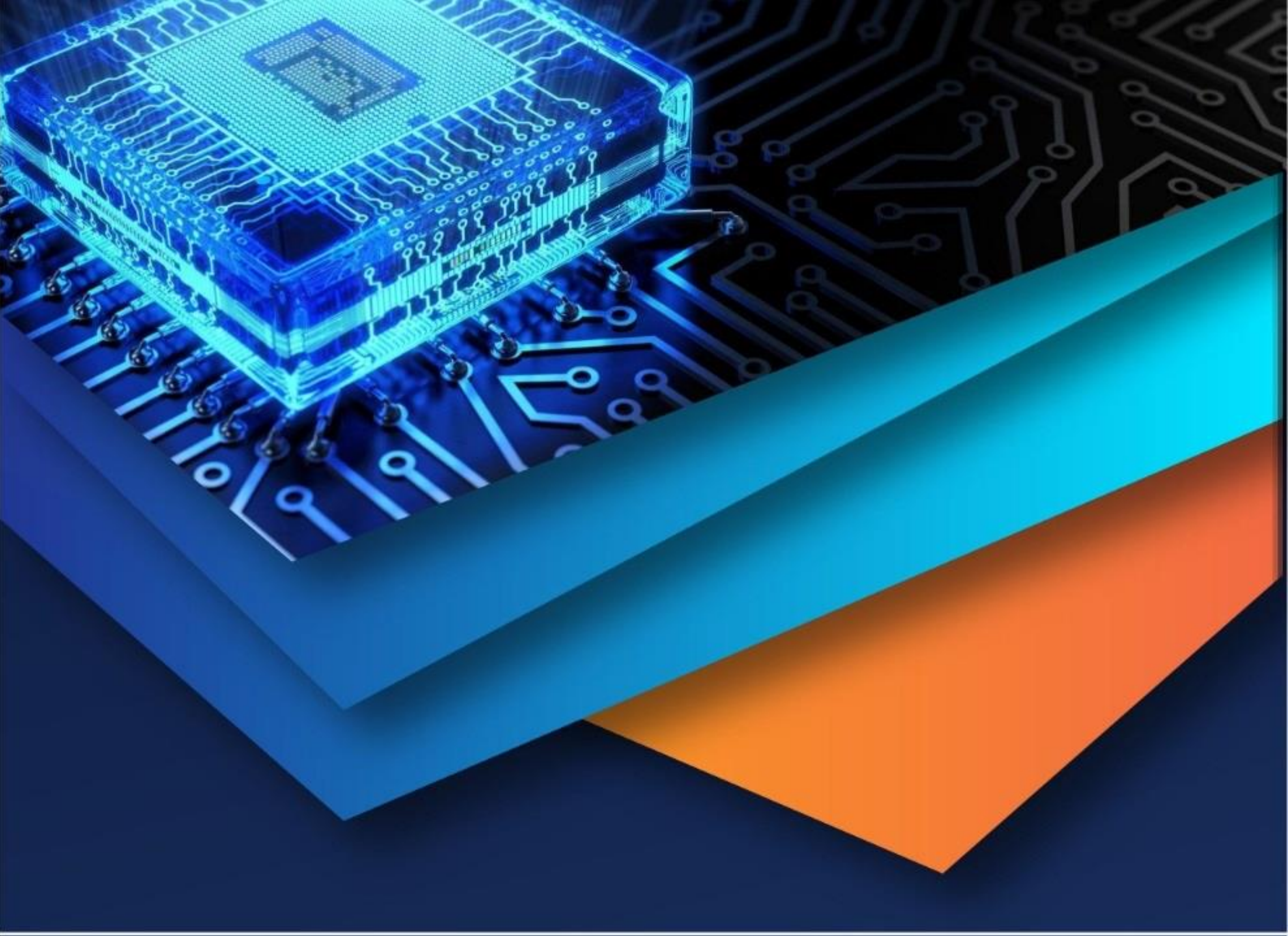
The leaf vein extraction technique proposed in this research provides an automated and accurate approach for recognising and classifying vein patterns from leaf photographs. By merging image processing techniques, morphological operations, feature extraction using Local Binary Pattern (LBP), and machine learning algorithms, the system effectively isolates and analyzes vein patterns. The results reveal that ensemble and boosting algorithms outperform simpler models in vein classification. Both Gradient Boosting and XGBoost Classifiers attain the highest accuracy of 92.86%, making them the most dependable models for this task. Random Forest, with an accuracy of 90.48%, likewise displays strong performance, illustrating the usefulness of ensemble approaches. The system's scalability and efficiency make it highly suitable for botanical research, plant taxonomy, and ecological investigations. Its capacity to process big datasets with minimum operator involvement decreases the time and effort necessary for vein extraction and classification. Furthermore, the automated approach provides consistent and reliable results, making it a significant tool for plant species identification and disease diagnosis.

In summary, the suggested methodology provides a stable and scalable solution for leaf vein extraction, presenting substantial promise for future applications in agriculture, botany, and environmental monitoring.

## REFERENCES

[1] Fu H., and Chi Z., "A two-stage approach for leaf vein extraction," In: Proceedings of International Conference on Neural Networks and Signal Processing, vol. I, Nanjing, Jiangsu, China, December 12–15, 2003, pp. 208–211
[2] Y. Li, Z. Chi, D.D. Feng Leaf vein extraction using independent component analy-sis IEEE International Conference on Systems Man and Cybernetics (2006), pp. 3890-3894
[3] Kirchgessner N., Scharr H., and Schurr U., "Robust vein extraction on plant leaf images," In: 2nd IASTED International Conference Visualization, Imaging and Image Processing, Malaga, Spain, 9–12 September, 2002.
[4] R. Radha, S. Jeyalakshmi An effective algorithm for edges and veins detection in leaf images IEEE World Congress on Computing and Communication Technologies (2014), pp. 128-131
[5] Park J, Hwang E, and Nam Y, "Utilising venation features for efficient leaf image retrieval," The Journal of Systems and Software, 2008, 81: 71–82.
[6] Jin Yingen. Botany. Beijing: Science Press, 2006.
[7] Li Y.F., Zhu Q.S., Cao Y.K., Wang C.L, "A Leaf Vein Extraction Method Based On Snakes Technique," In: International Conference on Neural Networks and Brain 2005, vol. 2, 13–15 Oct. 2005, pp. 885–888.
[8] Clarke, J. ,Barman ,S. ,Remagnino ,P. ,Bailey ,K. ,Kirkup ,D. ,Mayo ,S. ,Wilkin , P.: Venation pattern analysis of leaf images. Lecture Notes In Computer Science 4292, 427–436(2006)
[9] H. Fu, Z. Chi A two-stage approach for leaf vein extraction Proceedings of the IEEE International Conference on Neural Networks and Signal Processing, 1 (2003),pp. 208-211
[10] T. Chan, L. Vese An Active Contour Model Without Edges Scale-Space Theo-ries in Computer Vision (1999), pp. 141-151
[11] I. Philipp, T. Rath Improving plant discrimination in image processing by use of different colour space transformations Comput. Electron. Agric., 35 (1) (2002), pp. 1-15
[12] X. Zheng, X. Wang Leaf vein extraction based on gray-scale morphology International Journal of Image Graphics Signal Process., 2 (2) (2010), p. 25
[13] D.S. Bright, E.B. Steel Two-dimensional top hat filter for extracting spots and spheres from digital images J. Microsc., 146 (2) (1987), pp. 191-200
[14] H. Fu, Z. Chi Combined thresholding and neural network approach for vein pattern extraction from leaf images IEE Proc. Vis. Image Signal Process., 153 (6) (2006), pp. 881-892
[15] Nam Y., Hwang E., and Kim D., "A similarity-based leaf image retrieval scheme: Joining shape and venation features," Computer Vision and Image Understanding,2008, 110 (2): 245-259.
[16] A. Salima, Y. Herdiyeni, S. Douady Leaf vein segmentation of medicinal plant using Hessian matrix IEEE International Conference on Advanced Computer Science and Information Systems
[17] K.B. Lee, K.S. Hong An implementation of leaf recognition system using leaf vein and shape Int. J. Bio-Sci. Bio-Technol., 5 (2) (2013), pp. 57-66
[18] Soille P., "Morphological image analysis applied to crop field mapping," Image and Vision Computing, 2000, 18: 1025 1032

[19] M.G. Larese, R. Nam´ ıas, R.M. Craviotto, M.R. Arango, C. Gallo, P.M. Granitto Automatic classification of legumes using leaf vein image features Pattern Recognit., 47 (1) (2014), pp. 158-168
[20] T. Ojala, M. Pietik¨ ainen, and D. Harwood (1994), "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions", Proceedings of the 12th IAPR International Conference on Pattern Recognition (ICPR 1994), vol. 1, pp. 582–585.
[21] T. Chokey and S. Jain, "Quality Assessment of Crops using Machine Learning Techniques," 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 2019, pp. 259-263. doi: 10.1109/AICAI.2019.8701294
[22] N. Shah and S. Jain, "Detection of Disease in Cotton Leaf using Artificial Neural Network," 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 2019, pp. 473-476. doi: 10.1109/AICAI.2019.8701311
[23] S. Singh and S. Jain, "Detection and Classification of Plant Disease Using Artificial Intelligence," 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2024, pp. 1-5, doi: 10.1109/ICRITO61523.2024.10522449.
[24] Jain, Sarika & Ali, Kunwar. (2022). Brinjal Disease Classification Using Deep Learning. 1-5. 10.1109/ICRITO56286.2022.9965112.




10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

TOGETHER WE REACH THE GOAL


# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY