© NOV 2024 | IRE Journals | Volume 8 Issue 5 | ISSN: 2456-8880

CNN-Driven Plant Species Recognition System

SUCHETHA N V[1], ARJUN K[2], PANCHAMI B S[3]

[1]*Department of CSE, Sri Dharmasthala Manjunatheshwara Institute of Technology, Ujire and Affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India*

[2]*Department of AIML, Canara Engineering college, Bantwal and Affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India*

[3]*Student, Department of CSE, Sri Dharmasthala Manjunatheshwara Institute of Technology, Ujire and affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India.*

Abstract- The need for innovative solutions to monitor and sustain plant species biodiversity is rising because global biodiversity declines rapidly. The traditional methods of identifying plants are frequently time-consuming and require botanists with expertise in these areas. The objective is to create a dependable, efficient, and scalable system for recognizing plant species using machine learning technology. The intent here is to construct a user-friendly tool leveraging complex machine learning techniques such as Convolutional Neural Networks (CNN), which allow scientists and the public to identify plant species correctly. The suggested approach is based on an extensive dataset of photos that depict many plant species at various phases of growth in addition to varying environmental circumstances. This assists in classification and feature extraction by CNNs that enables the model to learn specific features from these pictures and increase its extent of generalization across various plant species. This method could significantly increase plant identification's availability, speed, and accuracy, supporting conservation efforts and monitoring the world's biodiversity.

Indexed Terms- Classification, Convolutional Neural Networks, Deep Learning, Feature Extraction, Generalization, Machine Learning.

I. INTRODUCTION

Botany, ecology, conservation, and agriculture have long depended on the identification of native plant species, a labor- and time-intensive technique historically reliant on botanists and field biologists. Recent developments in AI and ML have revolutionized this field, making plant identification automated. By utilizing machine learning models, projects can improve our capacity to identify and categorize different plant species in localized locations. Deep learning techniques, computer vision, and vast datasets of botanical photos significantly ease the difficulty of plant species identification.

The aim of the project is to create an intuitive platform by using cutting-edge technology such as machine learning and computer vision. Users can upload pictures of plants they encounter onto this website, where advanced image recognition technology will identify each species and provide detailed information. This includes data on growth habits, ecological importance, taxonomic classification, and potential uses. The project aims to engage the general public in understanding and preserving the rich biodiversity of their local area while supporting researchers and botanists in their work. Local plant species identification initiatives are essential for encouraging biodiversity preservation and strengthening relationships between local populations and their natural environments.

A local plant species identification project aims to enhance our understanding of the flora in a specific geographical area. These efforts aim to catalog and identify native plant species by combining scientific knowledge, community involvement, and technology tools. By involving local communities, employing mobile apps, and integrating data from field studies, the method contributes to conservation efforts, ecological research, and the promotion of local biodiversity awareness. Delving into the technical aspects of local plant species identification techniques, these initiatives employ a multifaceted approach that integrates geospatial technologies, machine learning algorithms, and bioinformatics tools. The application serves as an educational platform, delivering knowledge about regional ecosystems and the importance of preserving native plant species.

The focus is on the challenges of accurately identifying plant species, especially in diverse ecosystems, and debate the transition from traditional methods to deep learning and computer vision. The paper [1] highlights the importance of large annotated databases from citizen-science initiatives for training these models. The complexities of identifying plant species "in the wild" are explored, and two main approaches are discussed: standard fine-grained recognition and a retrieval-based approach. The latter offers advantages such as explainability and support for open-set recognition but introduces additional complexity in efficient nearest neighbors search.

A proposed method for plant species identification makes use of deep learning, machine learning, and image classifiers. Many researchers concentrate on plant leaf-based identification since it is more accessible than other portions of the plant. This research [2] conducts a survey of the methods and classifications used to locate plants in recent years. Furthermore, this survey contains a comparison of various strategies based on the accuracy of the classifiers used. Leaf identification is critical for recognizing plant species. Plants leaf shapes differ, although each plant has rather consistent properties. These traits serve as the foundation for plant recognition. Experts have tried to identify plants based on their leaves.

The crucial challenge of automating plant species recognition through computational means is tackled. Recognizing the significance of plant species identification across various domains such as botany, agriculture, and environmental conservation, the researchers [3] introduce an innovative framework. With the use of the intrinsic gradients and edge orientations in photos, they apply the histogram of oriented gradients (HOG) technique to extract complex characteristics from leaf patterns. By integrating these extracted features into CNNs renowned for their prowess in image-related tasks, they achieve a refined classification mechanism. The synergy between HOG and CNNs not only streamlines the identification process but also elevates accuracy levels, enabling researchers and professionals to discern plant species with heightened precision.

The grape plant variety identification software for vineyards is proposed using data augmentation and CNN techniques. This research [4] addresses the challenge of identifying grape varieties in the Douro Region vineyards using in-field images, crucial for precision viticulture. The study proposes an automatic algorithm based on transfer learning and fine-tuning techniques using the AlexNet architecture. Despite challenges such as natural environment variations, low image volume, and high similarity among grape varieties, the proposed approach achieved a test accuracy of 77.30%. The use of a four-corners-in-one image warping algorithm and leaf segmentation contributed to successful classification. The datasets, collected in different harvest seasons, demonstrated promising results, offering potential assistance to Douro wine growers in automating grape variety identification for tailored vineyard management.

The challenge of manual plant species identification is addressed, highlighting the significance of automated methods for various purposes such as ecological balance, medicinal applications, and agricultural industries. The authors [5] suggest a method for classifying plant species based on leaf photos by utilizing computer vision and machine learning. The study involves image acquisition, pre-processing, feature extraction, and classification, utilizing a Multiclass-support vector machine (SVM). The evaluation with the Swedish leaf dataset demonstrates an accuracy of nearly 93.26%, with a goal to enhance further. The paper emphasizes the efficiency of automated identification over manual methods, discussing the advancements in technology, including the use of smartphones and digital cameras, in facilitating image-based recognition. The challenges associated with manually classifying plant species are discussed, as well as the potential for automation through the use of computer vision methods. The authors [6]analyse 120 peer-reviewed articles published between 2005 and 2015 as part of a comprehensive literature review. The focus is on computer vision techniques for the identification of plant species, classifying methods according to the shape, texture, color, edge, and vein structure of the examined plant organs. The essay provides a thorough summary that might help researchers and novices in these domains, emphasizing the relevance of these results for both computer vision and ecological research.

The significance of automated systems in recognizing plant species from digital images is emphasized because of the shortage of expert taxonomists. The paper [7] provides an extensive review of computational and morphometric methods, focusing on techniques such as analysing leaf outlines, flower shapes, and vein structures. Despite the potential of digital technologies, challenges persist, including specimen deformations, ambiguous class boundaries, and the need for precise feature selection.

A smartphone app [8] that uses automatic image recognition to detect different plant types is suggested. With its novel approach to plant species identification, Leafsnap tackles the drawbacks of labour-intensive manual procedures and offers a flexible, easily navigable alternative. The system's focus on using computer vision to simplify the identification challenge highlights its usefulness, allowing professionals as well as a broad spectrum of users, including scientists, ecologists, foresters, and even schoolchildren, to quickly identify different species of trees. The system's accuracy is improved while handling a variety of photos taken under real-world situations by combining a leaf/non-leaf classifier with a strong color-based segmentation algorithm. The extraction of curvature-based shape features effectively represents complex leaf shapes, which adds to the system's effectiveness.

Recent research highlights Convolutional Neural Networks (CNNs) as effective tools for detecting potato leaf diseases like late blight and early blight [9]. These CNNs utilize curated datasets of potato leaf images, employing techniques such as data augmentation and transfer learning for enhanced accuracy. Results show customized CNN models achieving up to 99.22% accuracy, surpassing traditional methods. With the help of visualization tools like confusion matrices, evaluation metrics like precision, recall, and F1-score verify their efficacy. Future research aims to expand datasets and explore CNN generalizability across diverse potato varieties and growing conditions.

Plant identification research use techniques such as discretization, feature selection with information gain, and picture segmentation (using PSO)[10]. Evaluation often focuses on datasets like Flavia, showing better accuracy in identifying plants despite image variations. Deep learning for improved feature extraction and classification in a variety of plant datasets may be the subject of future research.

Plant species identification is experiencing a paradigm change, moving away from expert-dependent approaches toward state-of-the-art deep learning and computer vision techniques. Furthermore, as technology progresses, a newfound synergy emerges between AI-driven approaches and large databases from citizen science projects, enhancing the resilience and scalability of these systems. Even while these developments offer previously unheard-of accuracy and efficiency, difficulties still exist. Problems like specimen variability, imprecise classifications, and the fine-grained subtleties of plant species require a balanced combination of domain-specific knowledge and computer power. Therefore, even while technology spurs advancement, it is still essential to combine cutting-edge techniques with botanical expertise in a synergistic way. This cohesive strategy not only improves identification procedures but also encourages interdisciplinary teamwork, propelling breakthroughs in ecological research, agriculture, biodiversity protection, and other fields. The table 1 specified below provides a comparative analysis of current systems in light of the proposed approaches.

Table 1: Comparative Analysis

Sl. No	Author(s)	Algorithms/ Techniques	Performance Measures
1.	Picek Lukáš, Šulc Milan, Patel Yash, Matas Jiři	CNN, Retrieval-based approach, ViT	91.15%
2.	Gargi Chandrababu, Ojus Thomas Lee, Rekha K S	Machine Learning, Deep Learning, Image Classifiers	Not Specified

3.	Truong Quoc Bao	HOG, CNN	93%
4.	Carlos S. Pereira, Raul Morais, Manuel J. C. S. Reis	CNN, Transfer Learning, Image Processing, Data Augmentation	89.75%
5.	Kaur Surleen, Kaur Prabhpreet	Hidden Markov Model	93.26%
6.	Wäldchen, Jana &Mäder, Patrick	Pattern recognition, image processing, and computer vision	91.2%
7.	James S. Cope, David Corney, Jonathan Y. Clark, Paolo Remagnino, Paul Wilkin,	Digital morphometrics, Image Processing	Not Specified
8.	Neeraj Kumar, Joao V.B. Soares, Ida C. Lopez, David W. Jacobs, Peter N. Belhumeur, Arijit Biswas, and W. John Kress	Computer Vision, Leaf Segmentation, Feature Extraction Logistic Regression	96.8%
9	Abdullah Walid, Md. Mehedi Hasan, Tonmoy Roy, Md. Selim Hossain, Nasrin Sultana	Convolutional Neural Networks, data augmentation and transfer learning	99.23%
10	Heba F. Eid	Image segmentation (using PSO), feature selection with information gain, and discretization	Not Specifed

This all-encompassing perspective ensures that new technologies complement the valuable knowledge and deep understanding ecologists and botanists have gained over decades. By proposing a system for plant category identification, The study attempts to address the challenges associated with manual plant species identification, which can be slow and prone to mistakes. This research involves processing a large set of photos of local plant species using CNN and image classifiers along with computer vision and machine learning techniques.

The goals of this project go beyond just identification. They include collecting many images, implementing CNN in detail, and creating an easy-to-use model for accurate web-based plant species identification. The study also aims to check the model's accuracy and reliability through thorough testing and validation. This will provide a dependable tool to help people understand and take care of their local plants. By doing this, the project not only makes plant identification easier but also helps people connect more with their local environment, supporting efforts to conserve biodiversity and take care of nature.

II. METHODOLOGY USED

The development of a machine learning system for plant species classification can be accomplished by utilizing the subsequent methodology:

Dataset Collection

Collecting a comprehensive dataset is crucial for building an effective machine learning model. The dataset should encompass a wide variety of local plant species to ensure the model's ability to generalize well. Images should be gathered from diverse sources to capture the natural variability of plants, including different angles, lighting conditions, and backgrounds.It's essential to ensure the dataset's quality by verifying image resolution, clarity, and relevance to the target task. Annotation of images with corresponding plant species labels is necessary for supervised learning, enabling the model to learn associations between image features and plant categories effectively. The dataset used is "Indian Medicinal Plant Dataset" from Kaggle.

Preprocessing and Feature Extraction

Prior to feeding photos into the model, preprocessing procedures are carried out to standardize and improve the quality of images. Resizing and cropping images focus on the relevant parts containing the plant to eliminate unnecessary background noise. Normalizingcolor helps to mitigate variations in lighting conditions and camera settings, making the model more robust. Applyingalgorithms like Gaussian and Morphological operations can aid in extracting meaningful features from images, such as edge detection and shape analysis. Noise removal techniques, such as filters or denoising algorithms, contribute to improving the signal-to-noise ratio in images, thereby enhancing the model's accuracy.

Model Training

Splitting the dataset into training, validation, and testing sets is essential for evaluating the model's performance accurately.Using gradient descent and other iterative learning algorithms, the model's parameters are optimized using the training set. The validation set is utilized to tune hyperparameters and prevent overfitting by assessing the model's generalization ability on unseen data. The testing set evaluates the final model's performance on completely unseen data to provide an unbiased estimate of its accuracy's classifiers are a popular option because of their efficacious handling of nonlinear decision boundaries and high-dimensional data.

Performance Analysis

Various metrics are employed to assess the model's performance and guide further improvements. Training loss and validation loss curves provide insights into the model's convergence and overfitting tendencies during training. Metrics such as accuracy, precision, recall, and F1-score offer a comprehensive evaluation of the model's classification performance across different classes. Monitoring the loss ratio for each epoch helps in identifying the trade-off between bias and variance, guiding regularization strategies. Plant species identification systems that are reliable and accurate can be developed by constant model iteration and improvement based on performance analysis.

III. SYSTEM DESIGN

Gathering, cleaning, and pre-processing the data collection is the initial phase. The ML classification algorithms Morphological image processing and Gaussian image processing are then used to train the data set. Following pre-processing, extract the features using a binary image or any other picture format. Use the leaf contour image. Finally, the prediction is completed. The interface displays the anticipated name of the plant. To improve user involvement, we'll add more features. The web application will be implemented using the most accurate algorithm after the accuracy has been calculated. Local plant species identification project seamlessly integrates a user-friendly interface with a robust machine learning pipeline. The architecture comprises key components to ensure efficient processing, accurate identification, and user engagement. Figure 1 shows the architecture of the proposed system.

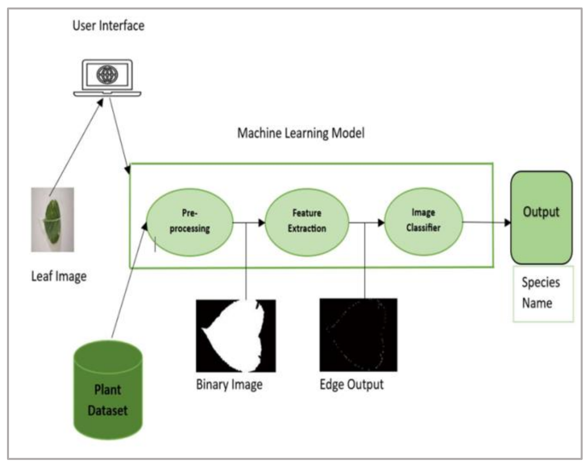


Figure 1: Architecture of the proposed system

User Interface: The user interface serves as the entry point for the identification process, allowing users to upload plant images effortlessly.

Image Upload: Users submit plant images through the interface, providing the system with diverse visual data for analysis.

Preprocessing: Upon image submission, the system engages in preprocessing the dataset. This phase involves operations like resizing, normalization, and noise reduction to optimize image quality. The pre-processed image is a binary image or contour image.

Feature Extraction: Extracting crucial features from pre-processed images. This process captures distinctive characteristics essential for accurate plant species classification. The input to this is the binary image.

Image Classifier: The retrieved features are used by a trained machine learning model to classify the plant species. To produce accurate predictions, the classifier makes use of patterns it has learnt from the training data. The edge detected image is given as input for classification.

Output Display: The identified plant species name is displayed through the interface, providing users with real-time results. The output includes the predicted species name and a corresponding image and some information about the image.

A system flowchart is a way of depicting how data flows in a system and how decisions are made to control events. It is necessary to import, clean, and preprocess the raw dataset. Using our user interface, first upload the image to the model. After incorporating image processing algorithms, feature extraction is done. The image from n sets of classes should be predicted by the trained model. Should the accuracy fall short of our expectations, then will retrain the model. The end product is the plant species name. The goal of the local plant species identification project is to correctly identify different plant species from photos.

Figure 2 depicts the system flowchart.

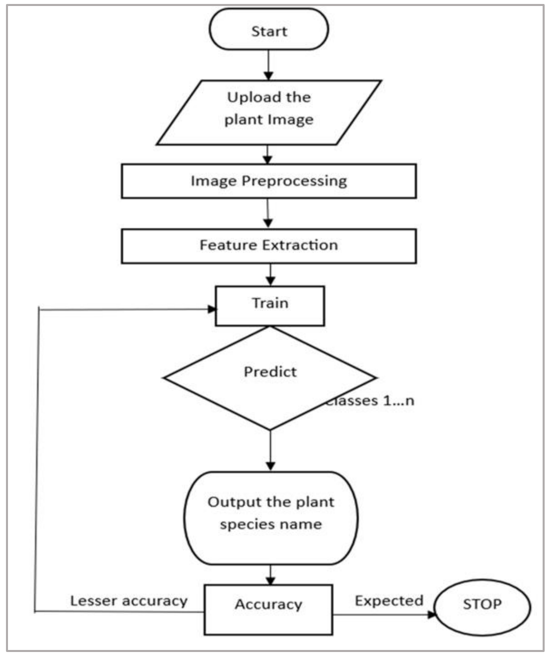


Figure 2: Flowchart of the Proposed System

IV. RESULT AND DISCUSSION

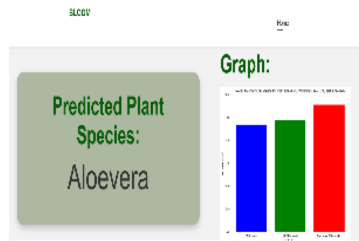
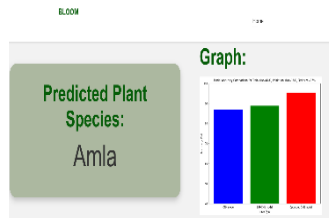
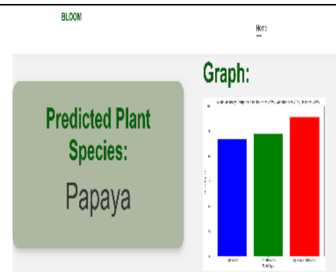
The subset of images used for training the system is displayed in Figure 3.

Figure 3: Subset of the dataset used

The system can identify up to 50 species of plants. Table 2 presents the results of three sample test cases.

Table 2: Unit test cases

Test case No.	Input	Expected behavior	Observed behavior	Status P=Pass F=Fail

1		Aloevera		P
2		Amla		P
3		Amla		F

The primary objective of the project was to predict plant species using machine learning algorithms. Table 3 displays the analysis conducted on three algorithms with varying training, testing, and validation sizes. It was determined that Optimized CNN yielded the highest accuracy across all cases.

Table 3: Analysis of the three algorithms

Training Size	Testing Size	Validation Size	Accuracy (%)		
			CNN	LBP-CNN	Optimized CNN
60%	20%	20 %	52.02	60.22	82.24
66%	18%	16 %	86.70	88.70	95.41
80%	10%	10 %	89.36	90.96	96.32

Figure 4 is the bar graph for the accuracy of three models where train size was 60%, validation size was 20% and test size was 20%.

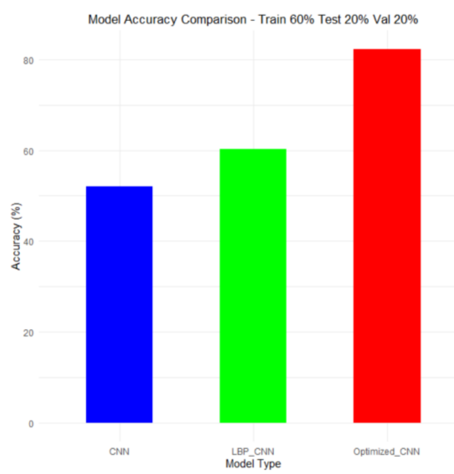


Figure 4: Analysis 2

Figure 5 is the bar graph for the accuracy of three models where train size was 66%, validation size was 18% and test size was 16%.

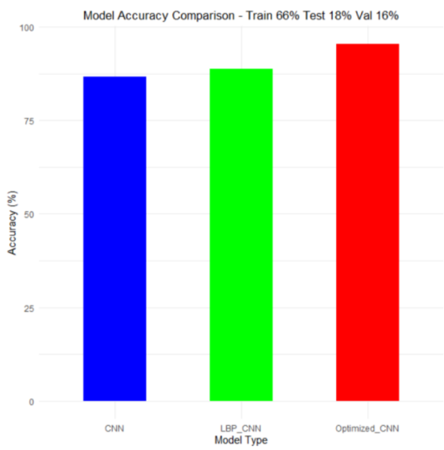


Figure 5: Analysis 3

Figure 6 is the bar graph for the accuracy of three models where train size was 80%, validation size was 10% and test size was 10%

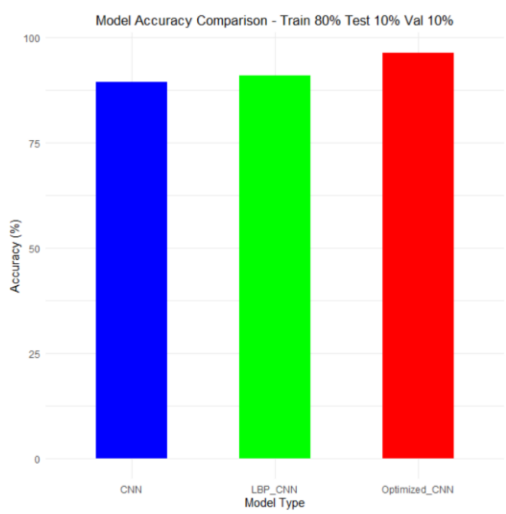


Figure 6: Analysis 4

CONCLUSION

In conclusion, the completion of dataset training, development of a user-friendly interface, and planned implementation of preprocessing techniques represent significant milestones in the local plant species identification approach. These initiatives demonstrate our dedication to using technology to improve the environment. The combination of trained datasets and intuitive interfaces empowers users to contribute seamlessly to the identification process. Looking ahead, integrating advanced preprocessing techniques promises to enhance the accuracy and efficiency of our identification algorithms. In addition to advancing our knowledge of the regional flora, this study helps communities recognize and value the region's tremendous biodiversity.

The successful completion of dataset training and the development of a streamlined user interface mark pivotal achievements in our ongoing venture, "Local Plant Species Identification". The meticulous training of datasets has strengthened our algorithmic models, establishing a robust foundation for precise plant identification. The user-friendly interface serves as a gateway, making the application accessible to a broader audience, including citizen scientists and nature enthusiasts.

For the future of the website, several integrations and enhancements can further expand its impact:

- Enhance User Interface: Improving usability will enhance accessibility and user engagement.
- Mobile Responsiveness Integration: Ensuring the website is user-friendly across all devices.
- Localization: Making the website available in multiple languages to reach a wider audience.
- Update and Expand CNN Model: Continuously updating the CNN model with more data and advanced algorithms to improve accuracy.
- Implement Advanced Features: Adding features like real-time image recognition or augmented reality for field identification.
- Collaborate with Botanical Experts: Partnering with botanists or organizations to verify identifications and provide expert advice, enhancing credibility.
- Educational Partnerships: Collaborating with academic institutions and conservation groups to integrate the website into educational curricula and conservation programs.

These enhancements aim to make the website more effective, engaging, and impactful in fostering biodiversity conservation and ecological awareness.

REFERENCES

[1] L. Picek, M. Šulc, Y. Patel, and J. Matas, "Plant recognition by AI: Deep neural nets, transformers, and kNN in deep embeddings”.

[2] G. Chandrababu, O. T. Lee, and K. S. Rekha, "Identification of Plant Species using Deep Learning".

[3] Q. B. Truong, "Plant species identification from leaf patterns using histogram of oriented gradients feature space and convolution neural networks.

[4] C. S. Pereira, R. Morais, and M. J. C. S. Reis, "Deep Learning Techniques for Grape Plant Species Identification in Natural Images".

[5] S. Kaur and P. Kaur, "Plant Species Identification based on Plant Leaf Using Computer Vision and Machine Learning Techniques".

[6] J. Wäldchen and P. Mäder, "Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review," Archives of Computational Methods in Engineering.

[7] J. S. Cope, D. Corney, J. Y. Clark, P. Remagnino, and P. Wilkin, "Plant species identification using digital morphometrics: A review.

[8] N. Kumar, P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. C. Lopez, and J. V. B. Soares, "Leafsnap: A Computer Vision System for Automatic Plant Species Identification,".

[9] Abdullah Walid, Md. Mehedi Hasan, Tonmoy Roy, Md. Selim Hossain, Nasrin Sultana, "Deep Learning-Based Potato Leaf Disease Detection Using CNN in the Agricultural System", International Journal of Engineering and Manufacturing (IJEM), Vol.13, No.6, pp. 9-22, 2023. DOI:10.5815/ijem.2023.06.02.

[10] Heba F. Eid, "Performance Improvement of Plant Identification Model based on PSO Segmentation", International Journal of Intelligent Systems and Applications (IJISA), Vol.8, No.2, pp.53-58, 2016. DOI:10.5815/ijisa.2016.02.07