International Journal for Innovative Engineering and Management Research

PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL

www.ijiemr.org

COPY RIGHT

IJIEMR Transactions, online available on 18th Dec 2024. Link

https://ijiemr.org/downloads.php?vol=Volume-13&issue= Issue12

DOI:10.48047/IJIEMR/V13/ISSUE12/27

Title: "PLANT IDENTIFICATION IN A COMBINED-IMBALANCED LEAF DATASET"

Volume 13, ISSUE 12, Pages: 223- 229

Paper Authors

Mr. L. Suneel, Mr. G. Chenna Rao, Mr. I. B. Koushik, Mr. J. Vivek

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

PLANT IDENTIFICATION IN A COMBINED-IMBALANCED LEAF DATASET

Mr. L. Suneel[1], Mr. G. Chenna Rao[2], Mr. I. B. Koushik[3], Mr. J. Vivek[4]

[1]Assistant Professor, Department of ECE, CMR Institute of Technology, Medchal, Hyderabad.

[2,3,4]Bachelor's Student, Department of ECE, CMR Institute of Technology, Medchal, Hyderabad.

ABSTRACT- Plant identification is essential for biodiversity conservation, agriculture, and environmental monitoring. However, identifying species becomes challenging when dealing with imbalanced datasets, where some species are underrepresented. This project focuses on a combined imbalanced leaf dataset to develop a robust system for plant identification and damage assessment. Advanced machine learning and image processing techniques are employed to analyze leaf morphology and texture for accurate classification. The system also integrates innovative strategies to identify underrepresented species with equal precision. Moreover, it incorporates a damage detection module to assess leaf health and recommend remedies for diseased or damaged leaves, such as nutrient management, pest control, and environmental adjustments. Convolutional neural networks are used for feature learning, while data augmentation helps address dataset imbalances. Comprehensive testing shows significant improvements in precision and recall metrics. This approach has applications in agriculture, biodiversity research, and sustainable farming, enabling efficient species identification and health monitoring.

KEYWORDS: Convolutional Network Networks, Random Forest, Imbalanced dataset

1.INTRODUCTION

Plant identification plays a critical role in biodiversity preservation, agricultural productivity, and environmental health monitoring. The rapid loss of plant species due to climate change, deforestation, and urbanization has heightened the need for efficient plant identification systems. Traditional methods rely heavily on expert knowledge and manual classification, which are time-consuming and prone to errors, especially when dealing with vast datasets. This project addresses these challenges by focusing on the identification of plant species using a combined imbalanced leaf dataset, which represents diverse plant species with varying sample distributions. Leveraging advanced machine learning algorithms and image processing techniques, the system performs species identification and health assessment with high accuracy. A unique feature of this project is the integration of a damage detection module to assess leaf health, identifying issues like pest infestation or nutrient deficiencies and providing remedies. Techniques such as convolutional neural networks for feature extraction and data augmentation for imbalance correction are utilized. This innovative approach has applications in agriculture, research, and environmental conservation, making plant identification and leaf health management efficient, accurate, and scalable.

Plant identification is a fundamental aspect of biodiversity conservation, sustainable agriculture, and ecosystem management. Understanding plant species and their health is critical for addressing environmental challenges such as climate change, deforestation, and declining

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org
crop productivity. This project focuses on developing a comprehensive plant identification system using a combined imbalanced leaf dataset, representing diverse plant species with varying sample distributions.

Fig 1: Displays a laptop identifying a banana leaf.

The proposed system employs advanced machine learning algorithms and image processing techniques to overcome the challenges posed by imbalanced datasets. The identification process involves three critical stages: image preprocessing, feature extraction, and classification. Initially, the system preprocesses leaf images to enhance quality and remove noise. In the next stage, features such as leaf shape, texture, and color are extracted using sophisticated algorithms. Finally, these features are fed into a machine learning model for accurate species classification.

In addition to species identification, the system integrates a damage detection module to assess leaf health. This module identifies issues like pest infestations, fungal infections, or nutrient deficiencies, and suggests remedies for improving plant health.Data augmentation techniques are applied to mitigate dataset imbalance and ensure robust performance.Finally, the extracted features from leaf images will be converted into digital data for classification. The plant species are identified using a template-matching approach. This system introduces a novel method for plant identification, contributing to environmental conservation and agriculture by automating species recognition.

2. LITERATURE REVIEW

Plant identification has been a longstanding challenge in the field of environmental science and agriculture. Traditional methods rely on expert knowledge and manual inspection, which are time-consuming and prone to errors, especially with large datasets. In recent years, machine learning techniques have revolutionized this process, improving the accuracy and efficiency of plant species identification. Convolutional Neural Networks (CNNs) have been widely adopted for feature extraction from leaf images due to their ability to handle large datasets and learn hierarchical representations of features. Several studies have focused on tackling the issue of dataset imbalance, where some plant species are underrepresented. Techniques like data augmentation, oversampling, and under sampling have been applied to address this problem and improve model performance.

Recent advancements in transfer learning have further enhanced classification accuracy by utilizing pre-trained models, reducing the need for vast training datasets. Despite these improvements, challenges such as environmental variability, image noise, and the complexity of leaf shapes remain significant hurdles in automated plant identification systems.into numericalreadable characters. A plant identification system generally works in four main stagesnamely image acquisition, leaf detection, feature extraction.

Fig 2: Camera System Identifying a Plant Leaf.

3.SYSTEM MODEL

Software Model:

The **Software Model** is the key to the system’s functionality, employing image processing techniques implemented using Python and OpenCV. It involves three main steps:

- Capturing leaf image (RGB).
- Preprocessing image for clarity.
- Extracting features and classifying

The first step involves capturing an image of a plant leaf using a high-resolution camera or USB camera connected to the system. The image is acquired in RGB format, which provides detailed color and texture information. This format is essential for effective feature extraction and preprocessing. The images must maintain high quality and minimal noise to ensure accurate analysis. The camera is strategically positioned to capture the leaf without any obstruction or shadows, enabling the system to acquire data with maximum precision. Proper lighting is also crucial during this phase, as it ensures the captured image has consistent brightness and contrast, making it ideal for further processing.

The second step is the preprocessing of the captured image. In this phase, the system converts the RGB image into a grayscale format to simplify the data and focus on the leaf's structural features. Filters, such as Gaussian or median filters, are applied to reduce noise and enhance the image's clarity. Edge detection techniques like Sobel or Canny algorithms are used to highlight key features, including the leaf's shape and vein patterns. These edges play a vital role in distinguishing different plant species. Thresholding techniques may also be used to segment the leaf from its background, ensuring a clean, isolated image for subsequent steps. The preprocessing phase is vital for improving the accuracy and reliability of the feature extraction process.

The third step involves the extraction of unique features from the preprocessed leaf image. Advanced image processing techniques, such as histogram analysis, shape descriptors, and texture analysis, are employed to identify the distinguishing characteristics of the leaf. For example, features like the vein structure, leaf margin, and texture patterns are critical for species classification. The extracted features are then compared with a pre-trained dataset that contains detailed information about a wide range of plant species. Classification algorithms such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), or Convolutional Neural Networks (CNN) are used to analyze the features and determine the plant's identity.

These machine learning models are trained on thousands of labeled leaf images to ensure robust performance across diverse datasets. After feature extraction and classification, the system outputs the identified plant species along with a confidence score.

This score represents the model's certainty regarding its prediction. For instance, a confidence score of 95% indicates that the model is highly confident in its classification. The system can handle a combined dataset containing hundreds of plant species, making it a versatile tool for research, agriculture, and environmental studies. It ensures accurate identification, even in challenging conditions, such as varying lighting or partial occlusion of the leaf.

The complete detail of the software model is shown in figure 3. This research adopted an Object-Oriented System Development methodology. An object-oriented software design focuses on "objects," such as leaves, instead of just functions or processes. This methodology ensures the system evolves through analysis, design, coding, and testing phases. Enhancements are introduced iteratively, allowing for continuous improvement. Incremental releases of software modules, such as leaf preprocessing, feature extraction, and classification, are delivered to ensure accuracy and adaptability across various plant species in the dataset.

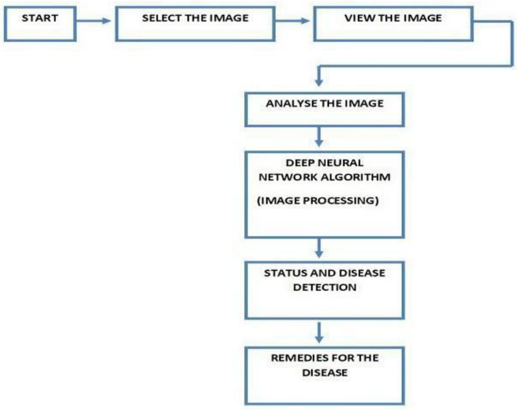

Figure 3: Steps of detection of damaged plant leaf recognition software model

4.WORKING

Requirements analysis: This phase is essential to ensuring the success of the project. Expectations need to be meticulously detailed and documented. The process is iterative, involving significant communication between stakeholders, end users, and the project team. Key stakeholders include botanists, environmental researchers, and data scientists, while end users may include agricultural experts and students.

Design: During this phase, the technical design requirements are defined. User requirements guide how the plant identification system will be structured, specifying technical needs such as the combined leaf dataset format, features like leaf matching algorithms, database integration, security measures, and hardware requirements.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

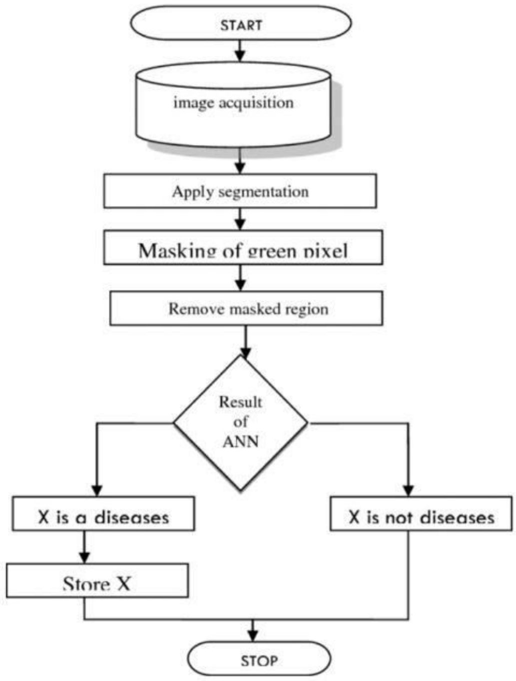

Fig 4: Represents a design flow of plant identification in a combined

Code: At this stage, the design is translated into executable code. Tools such as compilers, interpreters, and debuggers are employed. Programming languages like Python and AIML are utilized to develop key components, including the leaf identification algorithms.

Test: This stage occurs after the plant identification system has been developed. Various types of testing are performed, including performance and integration testing. User acceptance testing is the final phase, carried out by end users to ensure the system meets their expectations. At this point, defects may be found, and further work may be required in the analysis, design, or coding stages to refine the system and ensure accuracy in plant identification.

Maintenance: This phase ensures that the plant identification system has passed the user acceptance stage and is fully operational. If needed, users are trained or provided with documentation to help operate and maintain the software. Timely maintenance is performed, updating the system to adapt to any changes in the user environment or technology. This phase may encounter challenges such as hidden bugs and unforeseen issues in real-world scenarios.

5.RESULTS

This section presents the simulation results of the developed plant identification system.

Firstly, the system automatically detects the leaf images stored on the PC. Various images of leaves from different plants with varying shapes and colors are used for processing. The effects of daylight and environmental lighting conditions are considered during the analysis. The images are in RGB format, and the resolution is 800 x 600 pixels for accurate identification.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Fig 5: Analysing a image from given input

After processing the leaf image, the next step was the color segmentation algorithm. The images, after executing the color segmentation, highlight regions that match the desired leaf color. It can be observed that the algorithm successfully detects the region of interest (ROI) that contains the leaf. The next algorithm is used to enhance and extract the leaf shape. Once the leaf is isolated, it is converted into a binary format for further analysis and identification based on its unique features.

Fig 6: Displays the status of the leaf

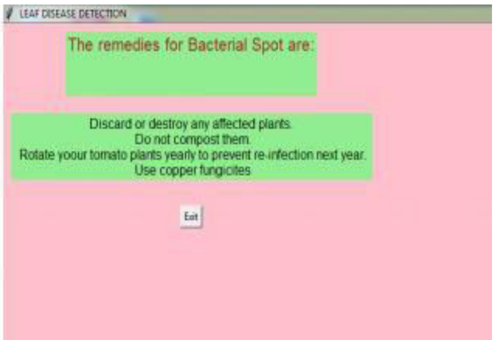

Fig 7: Displaying the remedy for the given leaf

The process begins with leaf segmentation using row and column techniques to isolate individual leaves from the dataset. Feature extraction identifies key characteristics for plant identification. Then, damage detection assesses any issues like discoloration or holes and suggests remedies for the leaf's recovery.

6.CONCLUSION

In this paper, an automated plant identification system using leaf segmentation and feature extraction has been developed. The system employs advanced image processing techniques to identify plant species from a combined leaf dataset. The system's performance has been evaluated on real leaf images, and results show that it can robustly identify various plant species even under different environmental conditions. Moreover, the system effectively detects leaf damage, such as discoloration, holes, or spots, and provides appropriate remedies for plant care. While the implementation works well, there are areas for improvement. The current system can be sensitive to noise and varying light conditions, which could affect accuracy. By using a higher-resolution camera or employing advanced preprocessing techniques, the system's robustness and accuracy could be significantly enhanced. Additionally, the segmentation algorithm can be optimized to handle leaves with complex shapes or overlapping regions. Further improvements in the damage detection algorithm, such as incorporating more advanced image recognition techniques, could lead to better detection of subtle leaf damage. Lastly, integrating a more diverse dataset would increase the system's ability to generalize across different plant species and environmental factors.

REFERENCES

1. K. Radhakrishna, D. Satyaraj, H. Kantari, V. Srividhya, R. Tharun and S. Srinivasan, "Neural Touch for Enhanced Wearable Haptics with Recurrent Neural Network and IoT-Enabled Tactile Experiences," 2024 3rd International Conference for Innovation in Technology (INOCON), Bangalore, India, 2024, pp. 1-6,
2. Karne, R. K., & Sreeja, T. K. (2023, November). Cluster based vanet communication for reliable data transmission. In AIP Conference Proceedings (Vol. 2587, No. 1). AIP Publishing.
3. Karne, R., & Sreeja, T. K. (2023). Clustering algorithms and comparisons in vehicular ad hoc networks. Mesopotamian Journal of Computer Science, 2023, 115-123.
4. Karne, R. K., & Sreeja, T. K. (2023). PMLC-Predictions of Mobility and Transmission in a Lane-Based Cluster VANET Validated on Machine Learning. International Journal on Recent and Innovation Trends in Computing and Communication, 11, 477-483.
5. Mohandas, R., Sivapriya, N., Rao, A. S., Radhakrishna, K., & Sahaai, M. B. (2023, February). Development of machine learning framework for the protection of IoT devices. In 2023 7th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1394-1398). IEEE.
6. Kumar, A. A., & Karne, R. K. (2022). IIoT-IDS network using inception CNN model. Journal of Trends in Computer Science and Smart Technology, 4(3), 126-138.
7. Karne, R., & Sreeja, T. K. (2022). Routing protocols in vehicular adhoc networks (VANETs). International Journal of Early Childhood, 14(03), 2022.
8. Karne, R. K., & Sreeja, T. K. (2022). A Novel Approach for Dynamic Stable Clustering in VANET Using Deep Learning (LSTM) Model. IJEER, 10(4), 1092-1098.
9. RadhaKrishna Karne, D. T. (2021). COINV-Chances and Obstacles Interpretation to Carry new approaches in the VANET Communications. Design Engineering, 10346-10361.
10. RadhaKrishna Karne, D. T. (2021). Review on vanet architecture and applications. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(4), 1745-1749.