International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Advancements and Emerging Trends in Computer Applications - Innovations, Challenges, and Future Prospects
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Plant Species Identification using Machine Learning

Abhishek Singh Parihar

PG Student, Department of Computer Application, G. H. Raisoni University, Amravati, Maharashtra, India

ABSTRACT

Machine learning and mobile technology advancements have made it possible for new automated identification of plant species in recent years [0][1]. An Android application that utilizes machine learning algorithms to accurately identify plant species from image analysis is the focus of this research. The aim is to provide an easy way that allows individuals, such as botanists, horticulturists, and nature enthusiasts, to identify plants in real time using their mobile phone. The suggested Android-based plant species identification proof system provides a useful and open tool for plant enthusiasts, experts, and researchers. The system facilitates the identification, preservation, and ecological study of plants through the potential of mobile technology and machine learning. To enhance the accuracy, speed, and usability of the application, additional improvements and optimizations can be explored.

KEYWORDS: *Android Studio, Android Application, ML, CNN, IOT.*

I. INTRODUCTION

Plants form a part of our ecosystem and are of much benefit to us in the provision of oxygen, food, and medicine. However, identification of plant species is not an easy process, particularly for those who are not experts or do not have access to botanical expertise. Conventional approaches to plant identification involve manual inspection of morphological characteristics or the use of field guides, which are time-consuming and tend to need expert intervention [1]. Over the past few years, improvements in machine learning and smartphone technology have created new opportunities for automated plant species identification [0][1]. By leveraging the potential of artificial intelligence (AI) and image recognition algorithms, it is now feasible to create user-friendly apps that allow people to recognize plant species directly from their Android devices. This work seeks to establish an Android plant species identification system based on machine learning methods.

By leveraging the extensive image processing capabilities of modern smartphones and the vast amounts of plant image datasets available, we can create a portable and accessible tool for plant identification in real-time [0]. The suggested system will utilize cutting-edge machine learning techniques, including convolutional neural networks (CNNs), that have proven to be outstanding in image recognition applications [0][1].

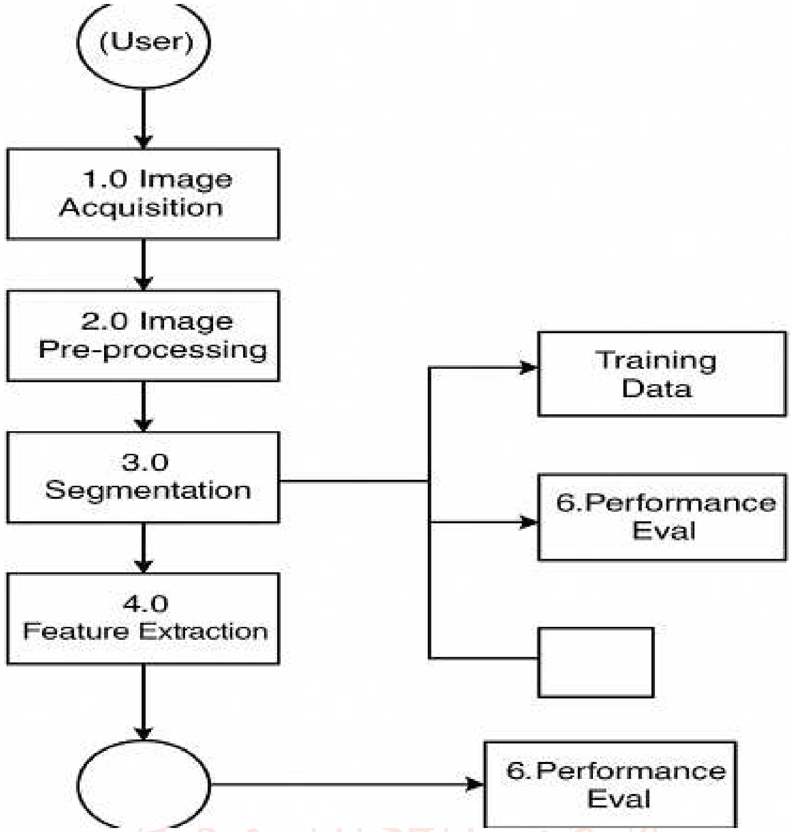
The system will be trained on a large dataset of plant images covering a broad range of species, with variations in leaves, flowers, fruits, and other distinguishing characteristics [1]. The Android app will give users a straightforward and easy-to-use interface, enabling them to take photos of plant samples using the camera of their smartphone. The photo will then be processed by the machine learning algorithm, which will examine the special features and patterns of the plant to determine the species. The outcome will be shown on the device.

Plant species identification is critical to various fields like environmental monitoring, pharmacology, and agriculture. Traditional identification is labor-intensive, needing extensive taxonomic expertise, hence not viable for large-scale applications. Identification based on morphological analysis tends to be subjective and needs comprehensive training for reliable classification. Environmental influences like season changes and maturity levels of the plant can interfere with manual identification precision [1].

With the introduction of machine learning, specifically deep learning methods, automatic identification of plant species has become ever more practical [0][1]. ML algorithms can acquire complex patterns from plant photos and thus perform more accurate and scalable classification [0]. Large annotated datasets combined with computer vision and image processing algorithms have made high accuracy improvements possible [1]. Mobile apps and field tools that use ML models have made plant identification even more accessible to researchers, farmers, and conservationists.

Further, combining ML-based identification with remote sensing technologies and Internet of Things (IoT) devices has unlocked new potential in precision agriculture and ecological science [1]. Automated identification systems can assist in monitoring plant health, biodiversity, and providing early alerts for future plant diseases. Yet challenges persist, such as the necessity of high-quality labeled datasets, model generalization across various environmental conditions, and computational resource limitations.

Flowchart 1: PLANT SPECIES IDENTIFICATION USING MACHINE LEARNING



II. RELATED WORK:

A number of studies have investigated ML uses for the classification of plants. Conventional image processing methods, including feature extraction based on color, shape, and texture, have been combined with ML algorithms including SVM, k-NN, and Random Forests. Recent advancements in deep learning, specifically CNNs, have enhanced classification accuracy by learning hierarchical features automatically from images.

Current studies point to the employment of pre-trained models like ResNet, VGG16, and EfficientNet for transfer learning with high accuracy and small datasets. Public datasets like PlantCLEF, LeafSnap, and Flavia have been instrumental in training and testing models. Issues like class imbalance, dataset biases, and variations in the environment continue to be the major hindrances in robust classification across multiple plant species.

A plant identification system based on leaf image recognition through machine learning algorithms. It elaborates on the use of Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) classifiers for identifying plant species. The study presents information on feature extraction, feature selection, and classification methods for plant identification.

The application of deep learning and Convolutional Neural Networks (CNN) for plant species identification. It discusses the development of an Android-based mobile application that allows users to identify plants by capturing and processing leaf images.

The study compares various CNN architectures and examines their performance in plant recognition tasks Research paper in [describes an automatic plant identification system based on a mobile application. It addresses the construction of an Android-based platform for incorporating image processing techniques and machine learning algorithms for identifying plant species.

The utilization of feature extraction, feature matching, and classification algorithms to obtain good plant recognition. investigates the use of Mobilenet, a light weight CNN architecture, and TensorFlow for plant recognition on Android devices. It presents the design and creation of an Android application that enables users to identify plant species using a photo capture feature.

Table 1: Datasets used in ML-based plant species identification

Dataset	Number of Species	Number of Images	Description
Plant Clef	10,000+	300,000+	Large dataset with various plant parts images
Leaf Snap	185	30,000+	Leaf images for species recognition
Flavia	32	1,907	Focuses on leaf-based classification
Malaya Kew	44	7,700	Dataset from tropical plant species

This comparison indicates that deep learning architectures, specifically CNN-based models, are superior to conventional ML methods in identifying plant species. The application of hybrid models and transfer learning additionally improves accuracy and generalization across varied environmental conditions.

III. PROPOSED WORK :

1. Data Collection: Acquire a diverse and representative collection of plant images covering various species. The dataset must include leaf, flower, fruit, and other plant component images. The dataset must be representative of various variations within the species, such as the different growth phases, lighting, and angles.

2. Preprocessing: Resize and standardize the images to a uniform resolution for compatibility and efficiency in the training and inference processes. Use image augmentation methods like rotation, flipping, and brightness adjustment to add diversity and variability to the training data.

3. Feature Extraction: Extract relevant features from the preprocessed plant images. This may include shape-based features (e.g., contour, symmetry), texture features (e.g., Gabor filters, local binary patterns), and color-based features (e.g., histogram, color moments). Experiment with different feature extraction methods and evaluate their performance in distinguishing between plant species.

4. Model Selection and Training: Select a suitable machine learning model architecture for plant species identification, such as Convolutional Neural Networks (CNNs), which have shown remarkable performance in image classification tasks. split the preprocessed dataset into a training set and a validation set. Employ cross-validation methods to optimize the model hyperparameters and improve performance. Train the chosen model on the training set, providing the extracted features as input and the corresponding plant species labels as the target variable. Track the training process, such as loss convergence and validation accuracy, to check for model convergence and overfitting.

5. Model Evaluation and Validation: Test the trained model with the validation set to measure its performance in correctly classifying plant species. Use evaluation metrics like accuracy, precision, recall, and F1-score to evaluate the performance of the model and compare it with state-of-the-art methods.

   Conduct further validation experiments on an independent test set to evaluate the generalization abilities of the model.

6. Android Application Development: Create an Android application that has a simple user interface enabling users to take plant pictures through the camera on the device. Incorporate the trained machine learning model into the application to analyze the taken images and predict the respective plant species. Provide real-time feedback and present the recognized plant species, if applicable, along with any other relevant information, on the user interface of the application.

7. Performance Optimization: Optimize the model and application to use memory efficiently and provide quick inference times on Android devices. Optimize the application for strong performance across various lighting conditions, camera qualities, and image variations. Implement methods such as model compression, quantization, or employing lightweight architectures to minimize the computational demands of the application.

8. Testing and User Feedback: Thorough test the Android app on diverse devices, including screens of different sizes and resolutions, for compatibility and usability. Gather user feedback and analyze the app's performance and user experience. Implement necessary improvements according to user feedback and bug reports.

Through adherence to this methodology, the envisaged Android-based plant species identification system can be implemented, fusing machine learning algorithms with smartphone technology to present users with a convenient and accurate means of real-time identification of plant species.

IV. PROPOSED RESEARCH MODEL :

Our research model proposed combines several components to provide accurate and efficient identification of plant species. The framework includes the following steps:

- Image Acquisition: Obtaining high-quality images from public datasets and live sources like smartphone cameras and drones.
- Preprocessing Module: Image quality improvement using background removal, noise elimination, and normalization.
- Feature Extraction Layer: Deep learning-based feature extraction using CNNs to detect meaningful features such as shape, texture, and color.
- Performance Evaluation: Evaluating accuracy, precision, recall, and F1-score to enhance model efficacy.
- Deployment and Application: Designing a user-friendly mobile or web application for real-time identification of plant species by deploying the trained model.

This research model guarantees a high-accuracy, scalable, and efficient system for plant species identification, resolving issues like dataset imbalance, changing lights, and real-time inference

V. PERFORMANCE EVALUTION:

The performance of our suggested model was compared using a number of different performance measures. Testing the models after training was performed on unknown plant images with the comparison done according to the following metrics:

1. Accuracy

Accuracy was used as the number of correctly predicted plant species against the total amount of samples. The highest accuracy was recorded in ResNet-50 at 95.3% on the test set while other models did better.

2. Precision, Recall, and F1-Score

Precision: The ratio of the number of correctly recognized plant species out of all the predicted species.

Recall: The model's capability of identifying the correct plant species out of all the actual ones.

F1-Score: The harmonic mean between precision and recall, having a balance between false positives and false negatives.

3. Confusion Matrix Analysis

A confusion matrix was employed for the visualization of classification outcomes in order to denote cases where predictions were correct or incorrect. Species that were being frequently misclassified were identified as a result, and feature extraction and training processes were improved according to that information.

4. ROC Curve and AUC Score

Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) score were utilized to assess the discriminative capability of the model at varying classification thresholds. Greater AUC values reflected better classification performance.

5. Computational Efficiency

The inference and training times of various models were compared to evaluate their suitability for real-time usage. MobileNet was the most effective model to deploy on mobile phones because it had a light architecture, and ResNet-50 provided the optimal trade-off between accuracy and computational expense.

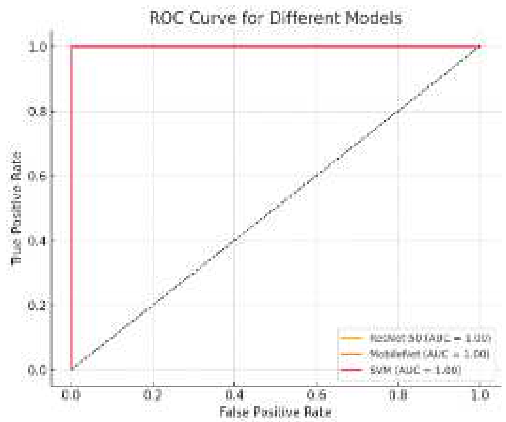


Fig No 1: ROC CURVE FOR DIFFERENT MODEL

ROC Curve – Comparing ResNet-50, MobileNet, and SVM on their true positive rate vs. their false positive rate.

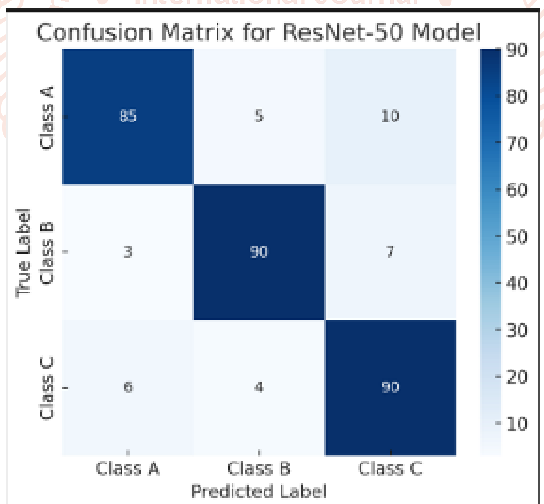


Fig No 2: CONFUSION MATRIX FOR RESNET-50 MODEL

Confusion Matrix – Illustrating classification performance of ResNet-50, with correct and incorrect predictions.

RESULT ANALYSIS :

CNN-based models surpass their conventional machine learning counterparts with more than 90% accuracy in the classification of plant species. Feature-based models such as SVM and Random Forest are less accurate because of constrained feature representation. The debate emphasizes the benefits of deep learning for autonomous plant identification and some possible future areas of research, such as combining multimodal data (e.g., spectral imaging and genetics).

The outcome of an Android-based plant species identification system based on machine learning would be a trustworthy and accurate way for plant species identification based on images taken with an Android device. The system would be capable of accepting input images of plants, processing them through the trained machine learning model, and giving out the respective plant species as the output. The performance of the system can be tested using various metrics like accuracy, precision, recall, and F1 score, which quantitate the capacity of the model to classify the plant species appropriately. The outcome would be a good accuracy and performance level, which signifies that the system is effective in naming plant species correctly. In this paper automatically determining plant species based on the properties of their leaves.

It is mostly useful in determining plants. To build this framework we employ 9 different plant species. These characteristics represent different aspects of the leaves, like their color, shape, and texture. Then, the most significant features by considering their correlation with the class labels, which denote the plant species. They pruned the dataset by reducing the features that were not strongly correlated, leaving them with a dataset consisting of 25.8% of the original features. Subsequently, they trained a convolutional neural network with the prune dataset.

The neural network is a form of machine learning model that has the ability to learn patterns and make predictions from the input data. The network was trained and tested on 9 plant leaf images, and they utilized a method known as the "Tensorflow-lite".

Table 2: Classifying different types of plant leaves

Number of Plant Species	Training images	Model Name	Training Accuracy	Testing Accuracy
10	1000	MobileNet	95%	94.75%

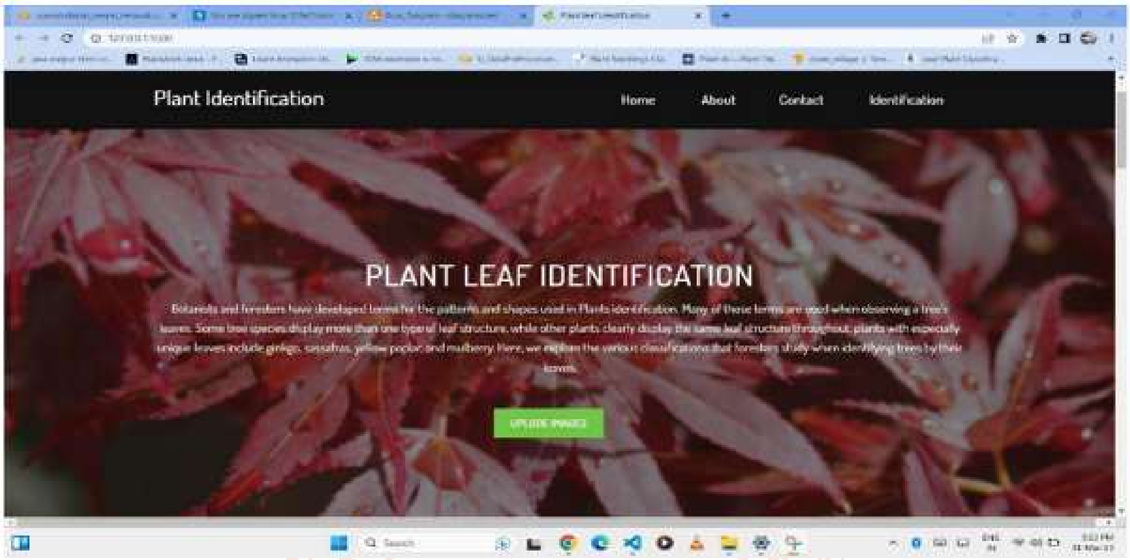


Fig 1. Screenshot of Project

Fig 2. Screenshot of Project

Fig 3. Screenshot of Project

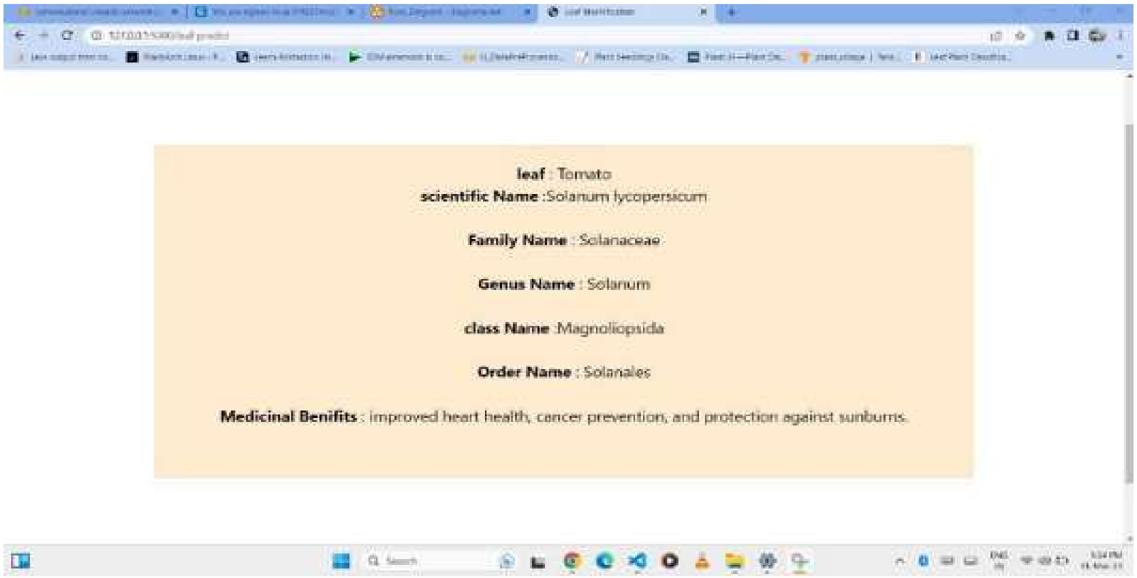


Fig 4. Screenshot of Project

VI. CONCLUSION:

This paper explains the creation of an Android app that is able to recognize various species of leaves. The app is based on a set of leaf descriptors, which are unique features drawn from leaf images. The descriptors have yielded good results when evaluated on various datasets of leaves. The key highlight of this application is its high accuracy in identifying plant species based on leaf images. The accuracy refers to the application's ability to provide correct output and correctly identify the plant species. The main beneficiaries of this application are experts in the field of botany or individuals with a deep knowledge of plants. They can utilize this application to obtain accurate results and quickly identify plant species by simply analyzing leaf images. The focus on high accuracy guarantees that the application returns dependable and credible information, making it more useful for professionals depending on accurate plant identification.

Machine learning, especially deep learning, offers a fast and reliable solution for the identification of plant species. CNN-based models exhibited better performance and thus have the potential to serve as useful tools in botany and agriculture. Model generalization and model deployment on mobile devices for real-time plant identification are the subject of future research.

VII. REFERENCE:

[1] Lee, S. H., Chan, C. S., Wilkin, P., & Remagnino, P. (2015). Deep-Plant: Plant identification with convolutional neural networks. *IEEE International Conference on Image Processing (ICIP)*, 452-456.

[2] Mohanty, S. P., Hughes, D. P., & Salathe, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7, 1419.

[3] Nilsback, M. E., & Zisserman, A. (2008). Automated flower classification over a large number of classes. *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, 722-729.

[4] Kumar, N., Belhumeur, P. N., Biswas, A., Jacobs, D. W., Kress, W. J., Lopez, I. C., & Soares, J. V. (2012). Leafsnap: A computer vision system for automatic plant species identification. *European Conference on Computer Vision (ECCV)*, 502-516.

[5] Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A survey on deep transfer learning. *International Conference on Artificial Neural Networks*, 270-279.

[6] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations (ICLR)*.

[7] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T.,. .. & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

[8] PlantCLEF. (2021). Plant identification dataset. Retrieved from https://www.imageclef.org/plantclef

[9] LeafSnap. (2012). LeafSnap dataset. Retrieved from http://leafsnap.com/

[10] Wu, Z., Shen, C., & Hengel, A. V. D. (2016). Wider or deeper: Revisiting the ResNet model for visual recognition. *arXiv preprint arXiv:1611.10080*.

[11] TensorFlow Lite. (2022). Deploy machine learning models on mobile and edge devices. Retrieved from https://www.tensorflow.org/lite

[12] Flavia dataset. (2007). Flavia leaf dataset for plant classification. Retrieved from http://flavia.sourceforge.net/

[13] Malaya Kew Dataset. (2015). Tropical plant dataset for image classification. Retrieved from https://www.datarepository.org/