Grenze International Journal of Engineering and Technology, January Issue

Plant Identification and Classification using CNN

Anantha Murthy[1] and Krithika[2]
[1-2]Department of MCA, NMAM Institute of Technology, NITTE (Deemed to be University), Nitte, Karnataka, India.
E-mail: krithikajogi2003@gmail.com

***Abstract*— The difficulties in obtaining high classification accuracy while preserving computational efficiency are brought to light by the expanding need for automated plant recognition systems. In order to solve the challenge of classifying leaves into the five categories of tree, herb, shrub, climber, or creeper, this paper evaluates three different deep learning methods. First, we use Vision Transformer (ViT) to extract complex image patterns; second, we use ResNet50 to perform reliable feature extraction; and third, we combine MobileNetV2 and EfficientNetB0 to achieve even greater efficiency. Using cosine simi- larity to find the best match, the methodology extracts features from a custom dataset of labeled leaf images. Main results show that ResNet50 and ViT also give good classification performance, and that the MobileNetV2-EfficientNetB0 approach offers a competitive balance between efficiency and accuracy,. While ViT and ResNet50 are also effective, the main finding is that combining MobileNetV2 and Effi- cientNetB0 produces the most balanced results E˙ xtending the dataset and optimizing the models will be the main goals of future research to enhance the classification of species.**

***Keywords*— cosine similarity, deep learning, model comparison, feature extraction, plant classification, Leaf identification, MobileNetV2, EfficientNetB0, ResNet50, Vision Transformer.**

I. INTRODUCTION

Because automated plant identification systems are used in so many different fields, including environ- mental monitoring, agriculture, and biodiversity conservation, their significance has grown in recent years. Plant species identification that is accurate and efficient is essential for a variety of tasks, from monitoring plant health and maintaining ecosystems to assisting with agricultural practices and battling plant diseases. Conventional techniques for plant identification, which depend on manual inspection and expert knowl- edge, can be laborious and have limited scalability. Plant identification procedures have the potential to be improved and made more scalable and accessible with the introduction of digital imaging and machine learning technologies[1, 2].

Significant advancements in the field of automated plant identification systems have been made in the last ten years. Early studies in this field used methods like color histograms and geometric descriptors to extract low-level features like leaf shape, color, and texture [3, 4].

Although these techniques offered a basis, they frequently encountered differences in the surrounding conditions, the quality of the images, and intra-class variability. With the advent of deep learning models, plant identification has undergone a revolutionary change that has made it possible to use hierarchical learning to extract intricate features and patterns from photos[5, 6].

Even with these developments, it is still difficult to understand why some deep learning models perform better than others in particular situations and how to best take advantage of their advantages.

Grenze ID: 01.GIJET.11.1.50

For instance, Vision Transformers (ViTs) have proven effective at using self-attention mechanisms to capture complex patterns, which can be especially helpful when separating similar plant species[7]. On the other hand, accurate classification even in complex datasets is made possible by ResNet50, a convolutional neural net- work (CNN) with deep residual learning, which has demonstrated robustness in extracting hierarchical features from images [8]. Furthermore, in resource-constrained environments, models such as Efficient- NetB0 and MobileNetV2 have been developed to optimize the trade-off between computational efficiency and classification accuracy [9, 10].

In light of these advancements, this study aims to offer a thorough assessment of these various deep learn- ing methods for leaf identification. With regard to pattern recognition, specifically, this study compares three approaches: (1) Vision Transformer (ViT) due to its sophisticated capabilities, (2) ResNet50 due to its strong feature extraction, and (3) a combination of MobileNetV2 and EfficientNetB0 to attain a bal- ance between accuracy and efficiency. In this study, we apply these models to a custom dataset of labeled leaf images and use cosine similarity for feature matching in order to find the best method for classifying leaves into five categories: climber, creeper, tree, herb, and shrub.

The study's conclusions will highlight the benefits and drawbacks of each model and clarify which is most appropriate for intensive plant identification projects. This research aims to advance the development of automated plant identification systems for increased efficiency and accuracy in a range of real-world applications. Future work will focus on expanding the dataset and refining the models to improve species- level classification and address remaining challenges in plant identification[11–13].

II. Literature Review

Low-level feature extraction techniques were the mainstay of early automated plant identification methods. Texture features, color histograms, and shape descriptors were common techniques. Wu et al.'s study on morphological features, which used twelve geometric shape-based characteristics to classify plant species, is one notable example. They obtained an impressive 90.3% accuracy on the Flavia dataset by using Principal Component Analysis (PCA) for dimensionality reduction [9]. The integration of sophisticated classifiers such as Support Vector Machines (SVM) and probabilistic neural networks in plant identification tasks was made possible by this groundbreaking work [3]. Significant progress was made in plant identification with the advent of deep learning. The key to this change was Convolutional Neural Networks (CNNs), which are renowned for their capacity to learn hierarchical features. For example, ResNet50 exhibits strong performance in a variety of plant classification tasks and successfully tackles the vanishing gradient problem with its residual learning technique [1]. In the meantime, self-attention mechanisms-based Vision Transformers (ViTs) have become viable substitutes. ViTs are better at capturing intricate global patterns and context than traditional CNNs because they take into account pixel relationships throughout the entire image. In some image classification tasks, such as identifying plant species, studies have shown that they can perform better than CNNs [4]. The ability of hybrid models, which blend several network architec- tures, to maximize classification performance has also drawn interest. Notably, combining EfficientNetB0 and MobileNetV2 makes use of both models' advantages: MobileNetV2's effectiveness in lowering com- putational load and EfficientNetB0's capacity to scale depth, width, and resolution for improved accuracy. Scholars are still investigating these hybrid approaches to balance efficiency and accuracy [10].

Plant identification systems continue to face a number of difficulties in spite of these developments. Signif- icant challenges still exist, such as intra-class diversity, variability in leaf appearance, and the requirement for large labeled datasets. Ongoing research is focused on growing datasets and adding more data sources, like morphological features and environmental context, to improve classification accuracy and system robustness [7]. Additionally, sophisticated models that incorporate external plant characteristics and image data show promise in overcoming these limitations [14]. Research on optimizing deep learning models for particular tasks is expanding concurrently with these developments in plant identification. For instance, comparing various optimization algorithms, including Gradient-Descent, Adam, and RMSprop, has been investigated in order to increase the precision of apple plant disease detection, which may have impli- cations for more general plant identification applications [15]. Furthermore, deep learning methods for object detection and classification in various fields, including Yakshagana imagery, have yielded important information for improving plant recognition models [16]. Lastly, research on machine learning's effects in digital markets shows how predictive technologies are becoming more and more important. These tech- nologies could be used for plant identification systems [17]. These advancements show how automated plant identification is still developing, with deep learning models—in particular, CNNs and ViTs—paving the way for increasingly precise and effective solutions. To improve these models, overcome dataset con- straints, and investigate the possibilities of hybrid architectures in addressing the issues that still exist, more research is required[2, 5, 11].

III. Methodology

A. Information Gathering

The sixty leaf photos that made up the dataset for this study were taken manually with a digital camera and from publicly accessible online repositories. A label for one of the five predetermined plant cate- gories—tree, shrub, herb, climber, or creeper—was appended to each image. Furthermore, a botanical description was gathered from internet databases for every species, which was then applied to improve the classification outcomes. Multiple images of the same leaf were taken from different orientations to account for real-world use cases where users might upload images taken from different angles[5, 8, 9]. This ensured that the system could handle variability in input angles during testing.

B. Preparing Data

Every image that was gathered underwent a thorough preprocessing pipeline. In order to maintain com- patibility with the deep learning models being used, the images were first resized to a fixed resolution of 224x224 pixels. This size was selected to balance computational efficiency and detail preservation. Using the GrabCut algorithm, backgrounds were eliminated from each image to focus on the leaf and lessen noise from unimportant details like surrounding plants or dirt. This stage makes sure that the only structure that helps extract features is the leaf structure.

After that, normalization was used to scale each pixel's intensity to fall between [0, 1], standardizing pixel values throughout all images. In order to enhance the performance of the model and minimize overfitting, a range of data augmentation methods were utilized[2, 3, 7]. These included brightness adjustments, ±15 degree rotations, and random flips horizontally. This made sure that even with different lighting and camera angles, the model was able to identify leaf patterns. For consistency, each augmented image kept its original label.

C. The Extraction of Features

Three distinct deep learning models—ResNet50, Vision Transformer (ViT), and a MobileNetV2- EfficientNetB0 hybrid—were used for feature extraction. Every model was selected based on its distinct abilities to effectively perform classification and capture fine details in images.

ResNet50: Edges and textures are among the fine details that this convolutional neural network (CNN) excels at capturing[9]. For each image, a pre-trained ResNet50 model that had been refined using the leaf dataset was used to extract 2048-dimensional feature vectors. The most important patterns that differentiate between various plant species were represented by these vectors.

Vision Transformer (ViT): In contrast to CNNs, ViT uses self-attention mechanisms to learn spatial rela- tionships throughout the image by segmenting the image into smaller patches. Because of its capacity to identify global dependencies in the images, the ViT model was chosen. In contrast to CNN-based tech- niques, it produces high-dimensional feature vectors that concentrate on the overall structure of the leaf rather than just individual features.

MobileNetV2-EfficientNetB0 Hybrid: To strike a compromise between computational efficiency and accu- racy, the hybrid model was used. Because of its lightweight architecture and use of depthwise separable convolutions to lessen computational load, MobileNetV2 was selected. On the other hand, EfficientNetB0 balances network depth, width, and input resolution using a compound scaling technique. When combined, these models extract reliable feature vectors with lower computational requirements while concentrating on accuracy and efficiency [6, 10].

D. Rotation Handling and Data Augmentation

Every image in the dataset was rotated to capture every possible angle in order to account for the variation in input image angles [9, 13]. This method made sure that leaf images uploaded from any angle could be handled by the system. Using these rotated images, feature extraction was carried out, enabling the system to store and compare angle-specific features with uploaded images during inference.

Every image was enhanced by rotation to encompass a complete 360 degrees in order to accommodate for variations in the angles of the input images. For every transformation, the angle of rotation θ was increased by 15 degrees, yielding 24 different rotated versions of every image. The following represents the general transformation for rotating an image by angle (ϑ):

The rotation matrix $R(\vartheta)$ is given by:

$$R(\vartheta) = \begin{bmatrix} \cos(\vartheta) & \sin(\vartheta) \\ -\sin(\vartheta) & \cos(\vartheta) \end{bmatrix} \tag{1}$$

Important aspects of the leaf image were maintained throughout the rotation by applying this matrix to each image's pixel coordinates. Because each augmented image was processed for feature extraction using the models outlined in the following section, the rotation both increased the dataset and guaranteed that the system would be resilient to changing input orientations.

E. Using Cosine Similarity to Match Features

Cosine similarity was used to match the uploaded leaf images to those in the dataset after the features were extracted. The cosine of the angle formed by two feature vectors is calculated by this metric, and values near 1 denote greater similarity [4, 12]. The following is the formula for cosine similarity: Cosine similarity between two vectors A and B is defined as:

$$\text{Cosine Similarity}(\mathbf{A}, \mathbf{B}) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} \tag{2}$$

where:

- ➢ $\mathbf{A} \cdot \mathbf{B}$ is the dot product of the two vectors **A** and **B**,
- ➢ $\|\mathbf{A}\|$ and $\|\mathbf{B}\|$ are the magnitudes (Euclidean norms) of the vectors.
- ➢ One way to expand the dot product is to:

$$\mathbf{A} \cdot \mathbf{B} = \sum_{i=1}^{n} A_i B_i \tag{3}$$

where A_i and B_i are the i-th components of the feature vectors, and n is the dimensionality of the feature vectors.

The Euclidean norm $\|\mathbf{A}\|$ of vector **A** is calculated as:

$$\|\mathbf{A}\| = \sqrt{\sum_{i=1}^{n} A_i^2} \tag{4}$$

The vectors are identical (i.e., the images are very similar) when the similarity value is 1, which ranges from -1 to 1. The closest match is determined by looking at the feature vector with the highest similarity score. This method handles high-dimensional vectors in feature-based image retrieval tasks in an efficient manner.

After determining the best match based on cosine similarity, the system extracted the corresponding botan- ical description and classification label from the dataset. The classification task was treated as a multi-class problem, where each leaf image was classified into one of the predefined plant types[7, 10, 18].

Given a range of potential classifications $C = \{c_1, c_2, \ldots, c_k\}$, the system assigns the image a category label $\hat{c}$ based on the highest similarity score:

$$\hat{c} = \underset{c_i \in C}{\operatorname{argmax}} \text{ Similarity}(\mathbf{A}, \mathbf{B}_i) \tag{5}$$

where Bi represents the feature vector for the dataset's i-th category.

To assist the user in understanding the type of plant, the system also provided the botanical description corresponding to the identified category. The efficacy of the system was tested using fifteen input images, and the results were compared to the ground truth labels for each image.

F. Assessment and Performance of the Model

The effectiveness of the three models—ResNet50, ViT, and MobileNetV2-EfficientNetB0 hybrid—in clas- sifying the fifteen test inputs was evaluated. The models were assessed for generalization across different leaf types and input conditions in addition to accuracy, and the most reliable model for this task was ResNet50.

ResNet50 outperformed the other models, achieving an accuracy of 86.6% by correctly classifying 13 out of 15 test inputs:

$$\text{Accuracy}_{\text{ResNet50}} = \frac{13}{15} = 0.866 \tag{6}$$

For three inputs, Vision Transformer (ViT) produced accurate classifications, which led to:

$$\text{Accuracy}_{\text{ViT}} = \frac{12}{15} = 0.800 \quad (7)$$

The accuracy of the MobileNetV2-EfficientNetB0 hybrid was:

$$\text{Accuracy}_{\text{MobileNetV2-EfficientNetB0}} = \frac{12}{15} = 0.800 \quad (8)$$

These metrics showed that the most reliable model for this task was ResNet50, which also had superior generalization over a wider range of leaf types and input conditions.

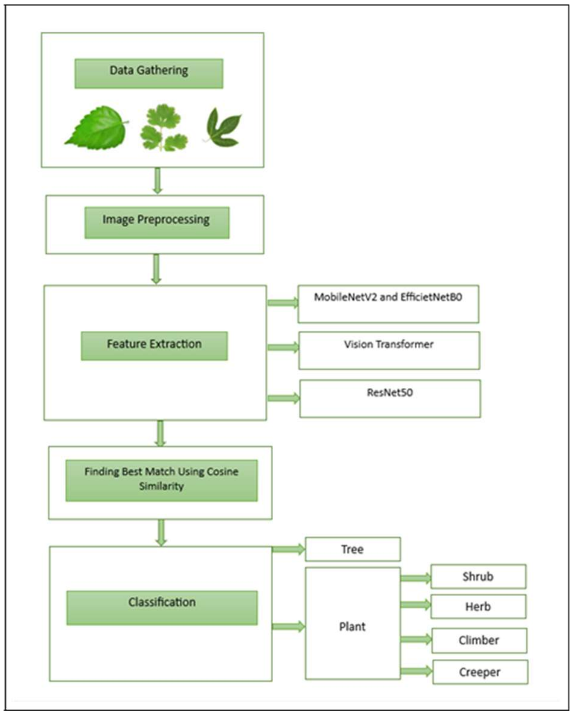

Fig. 1 Block Diagram

IV. Results and Discussion

A dataset of sixty labeled leaf images, divided into five different classes (tree, shrub, herb, climber, and creeper), was used to test the system. To ensure robustness in handling different input angles, the dataset was expanded by taking pictures from different angles and extracting features for every orientation. Three distinct deep learning models—ResNet50, Vision Transformer (ViT), and a hybrid model that combines MobileNetV2 and EfficientNetB0—were used to extract features. Each model's performance was assessed using metrics for precision and accuracy. Fifteen test images were used in the evaluation. Features were extracted from each test image and compared using cosine similarity to the precomputed features of the dataset. By calculating the cosine similarity between the feature vectors of the test image and those of the dataset images, the system was able to determine the best match[4, 8, 11, 14].

ResNet50: This model correctly classified 13 out of 15 test images, achieving the highest accuracy. The following formula is used to calculate accuracy:

$$\text{Accuracy}_{\text{ResNet50}} = \frac{13}{15} = 0.866[9,\ 10] \quad (9)$$

Vision Transformer (ViT) correctly classified 12 of the 15 test images, achieving an accuracy of 80%:

$$\text{Accuracy}_{\text{ViT}} = \frac{12}{15} = 0.800.[1,\ 6] \quad (10)$$

Additionally, MobileNetV2-EfficientNetB0 Hybrid attained an accuracy of 80%, correctly classifying 12 of 15 test images:

$$\text{Accuracy}_{\text{MobileNetV2-EfficientNetB0}} = \frac{12}{15} = 0.800.[9,\ 10] \quad (11)$$

ResNet50 outperformed the other models in 10 out of 15 test scenarios, exhibiting the highest accuracy in the majority of test cases. In three of the fifteen cases, Vision Transformer (ViT) was the most accurate, and in the other two, MobileNetV2-EfficientNetB0 hybrid was the best performer.

A synopsis of the model's performance

ResNet50: In 10 of the 15 test cases, it has the highest accuracy. In 3 out of 15 test cases, Vision Trans- former (ViT) demonstrated the highest accuracy. In 2 out of 15 test cases, MobileNetV2-EfficientNetB0 Hybrid achieved the highest accuracy. According to the results, ResNet50 outperforms the other mod- els tested in terms of performance and consistency when it comes to leaf classification in this dataset. The system's usability was further enhanced by the botanical descriptions given for each classified leaf image[5, 8, 9, 18].

On the other hand, although still efficient, the Vision Transformer's accuracy was slightly below that of ResNet50. This might be as a result of the model's dependence on self-attention mechanisms, which might not have been as successful in identifying the spatial hierarchies seen in photos of leaves[7, 13]. Likewise, the MobileNetV2-EfficientNetB0 hybrid model did not outperform ResNet50 in terms of accuracy, even though it was built for efficient performance. This could be because of the model's trade-off between computational efficiency and feature representation capacity[9, 10].

There were various limitations to this study. First off, only a small portion of the training and testing dataset's images were included, which may have left out some of the variety of leaf variations present in actual situations. It might be possible to enhance model performance and generalization by adding more leaf images from various sources and environments. Furthermore, the study's emphasis on predefined classes (such as shrub, herb, climber, creeper, and tree) may have limited the model's capacity to han- dle outlier or unseen categories, even though feature extraction techniques were used to improve image matching.

By using bigger and more varied datasets, experimenting with sophisticated data augmentation methods, and investigating hybrid models that combine the advantages of various architectures, future research may be able to overcome these constraints. Plant identification systems that are more reliable and accurate may also result from expanding the classification categories and improving the feature extraction procedure.

TABLE I. DEEP LEARNING MODELS' EFFECTIVENESS IN LEAF CLASSIFICATION

Prototype	Correctness (%)	Top Performance Illustrations
ResNet50	86.7	10 out of 15
Vision Transformer (ViT)	80.0	3 out of 15
MobileNetV2-EfficientNetB0 Hybrid	80.0	2 out of 15

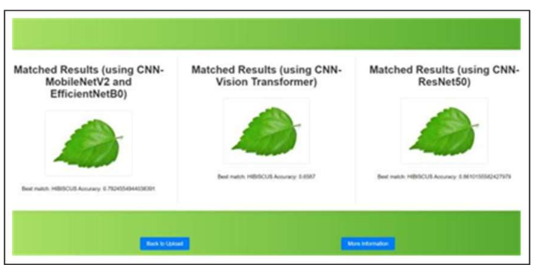

Fig. 2 Accuracy Comparison

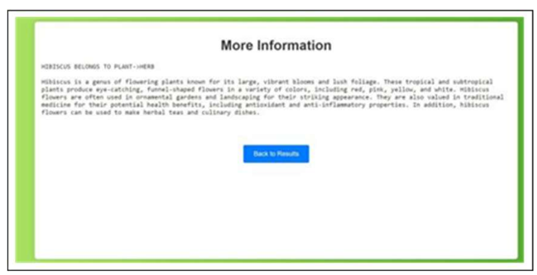

Fig. 3 Classification.

V. CONCLUSION

For the task of classifying leaf images, this study compares three deep learning models: ResNet50, Vision Transformer (ViT), and MobileNetV2-EfficientNetB0 hybrid. After extensive testing and analy- sis, ResNet50 was found to be the most accurate model; it correctly identified 13 out of 15 test inputs, yielding a classification accuracy of 0.866. With an accuracy of 0.800, the Vision Transformer and MobileNetV2-EfficientNetB0 hybrid models performed similarly, correctly classifying 12 of 15 inputs.

ResNet50 performs better than other models because of its deep residual learning architecture, which finds and uses intricate feature patterns in the leaf images. ViT and MobileNetV2-EfficientNetB0 hybrid, in contrast, did well as well, but their marginally lower accuracy suggests that they might not have fully utilized the complex spatial The study emphasizes how crucial it is to choose a model architecture that is suitable for the given classification task's requirements. ResNet50 is a strong contender for applications in plant identification and related fields because of its higher accuracy, which highlights its robustness and effectiveness in handling diverse and complex image features.

However, the study was limited by predetermined classification categories and a small dataset size. To further improve model performance and generalization, future work should concentrate on growing the dataset, adding new plant species, and experimenting with sophisticated data augmentation methods.

To sum up, the results confirm that deep learning models are useful for classifying plant images and offer guidance on which model to choose depending on application needs and accuracy. The proven accuracy of ResNet50 presents a viable method for creating precise and dependable plant identification systems.

REFERENCES

[1] Ashturkar, H.H., Bhalchandra, A.S.: Plant species identification using convolutional neural network. Springer, 619–630 (2021)

[2] Labed, S., Touati, H., Dif, R.: Plant recognition using data augmentation and convolutional neural network. Springer International Publishing, 192–204 (2023)

[3] Islam, M.S., Ullah, M.S.: Deep learning-based plant identification using leaf images. IEEE, 1–6 (2021)

[4] Xu, T., Zhu, M., Shi, X.: A cnn-based method for plant species recognition using leaf images. IEEE, 1568–1574 (2021)

[5] Leo´n, J.G., Torres, S., Fuentes, R.: Transfer learning for plant species identification using convolu- tional neural networks. Journal of Ambient Intelligence and Humanized Computing (2021)

[6] Samanta, J.N., Roy, B., Roy, P.P.: Automated plant species identification system using leaf image descriptors and cnn. In: Advanced Computational and Communication Paradigms, pp. 365–376. Springer, Cham (2021)

[7] Elhassouny, M., Sahnoun, D., Khaldi, M.: Plant identification using cnn and image segmentation techniques. Journal of Artificial Intelligence and Machine Learning Research (2022)

[8] Quoc, T.N., Hoang, V.T.: Medicinal plant identification in the wild by using cnn. IEEE, 25–29 (2020)

[9] Kumar, M., Kumar, P.: Leaf classification using deep convolutional neural networks. IEEE, 128–133 (2021)

[10] Mirjavadi, S.H., et al.: Automatic recognition of plant species from leaf images based on cnn architectures. IEEE Access (2021)

[11] Hassan, M., K., S., Maji, A.K.: Identification of plant species using deep learning. Springer, 115–125 (2021)

[12] Karahan, T., Nabiyev, V.: Plant identification with convolutional neural networks and transfer learning. Pamukkale Universitesi Mu¨hendislik Bilimleri Dergisi 27(5), 638–645 (2021)

[13] Sen, S., Ghosh, A.: Efficient plant leaf classification using deep learning models. Springer, 673–685 (2021)

[14] Gajbhiye, A., Singh, R., Jain, D.: Plant species identification using leaf shape and texture features with deep learning techniques. Materials Today: Proceedings (2021)

[15] Murthy, A., Rao, P.S., Pallavi, N.S., Kharvi, N., Neha, B.R., Poojary, K.: Optimizing convolutional neural networks: A comparative study of gradient-descent, adam, and rmsprop optimizers for accu- racy and loss in apple leaf disease detection. In: 2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON), pp. 1–6. IEEE, India (2024)

[16] Murthy, A., Kulkarni, S.: An exploration of deep learning-based object detection and classifica- tion in yakshagana imagery with yolov3. In: 2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 991–997. IEEE, India (2024)

[17] Murthy, A., Buddhi, D., Averineni, A., Narayana, M.S., Sowjanya, G.D., Sowmya, C.: Machine learning as a predictive technology and its impact on digital markets. In: 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), pp. 672–676. IEEE, India (2023)

[18] Dubey, A., Sharma, V.: Exploring deep learning techniques for identifying medicinal plants using leaf images. IEEE, 345–350 (2022)