International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 07 | July - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

PLANT IDENTIFICATION USING MACHINE LEARNING

Harisha K S[1],Pavitra M H[2], Spandana S R[3], Shruthi B S[4],Archana B S[5]
[1] Assistant Professor, Department of Electronics and communication engineering[1]
[2,3,4,5] B.E. Students, Department of Electronics and Communication Engineering
Government Engineering College Haveri,
Visvesvaraya technological university, Belagavi

---***---

Abstract - The preservation of therapeutic plants is crucial as it will help a broad variety of sectors, including medicine, botanic research, and plant taxonomy studies, among others. Existing technologies cannot imitate the range of therapeutic plant species present in India. The suggested technique facilitates the classification of medicinal plants by exploiting textural aspects crucial in leaf recognition and identification. The three key phases of the proposed technique are picture enhancement, feature extraction, and classification. The photographs of the leaves are shot using cellphones and then processed using digital image processing algorithms to extract the features that may be compared between them. Finally, the CNN classifier is employed to develop an automated classifier.

Key Words: Plant identification, CNN algorithm, Image processing.

1.INTRODUCTION

The living beings on the earth are dependent upon the plants. The plants are the main source of oxygen. Apart from this, the plants are used in a variety of industrial applications such as herbs, ingredients in biofuels, biomass, pharmaceuticals etc. People have been using plant as a traditional medicine. It's considered non-expensive and does not cause any side effects. Different kinds of plant species are available on the earth; recognition of these plants is difficult. There are several methods to recognize a plant. At present, plants are identified manually by taxonomist, which are prone to human errors. To avoid this, automated plant identification system has been developed by several researchers. Medicinal plants are those plants that are used in treating and preventing specific ailments and diseases that affect human beings. Knowledge accumulated by the villagers and tribal on herbal medicine remains unknown to the scientists and urban people. Many plant species associated with the rural people are on the verge of disappearing and are on vulnerable list. The impact of deforestation, urbanization and modernization is shifting the rural people from their natural habitats and their very knowledge particularly with respect to herbal drugs is slowly disappearing. Our immediate goal is to preserve this knowledge. Whatever knowledge exists today is mostly confined to the older generation. In this context, some approaches are needed to preserve and develop traditional knowledge.

Plant classification is an active research area, with plants being used in agriculture, medicine and food industry, as well as in the preparation of cosmetics and a range of food products. Individuals cannot characterize plants as effectively as botanists, who do so by classifying those utilizing leaves, flowers, seeds, and roots. Agent-based systems classify plants into species that can be used in medicine and as food. Keeping in mind the end goal, which is to provide data on therapeutic plants, it is critical to have an intelligent system framework that recognizes natural species with the assistance of their digitized databases. An intelligent system is a key strategy utilized in plant based recognition systems to create real models from plants, incorporating pattern classification and object recognition. Researchers have created a plant acknowledgment framework utilizing plant leaves, flowers, fruits and seeds and by taking into consideration the visual content of their images such as color, texture, and shape.

Plant leaf identification using machine learning refers to the process of training a computer system to recognize different types of plant leaves using algorithms and statistical models. This process involves feeding the computer system with large amounts of data containing images of plant leaves, and then using machine learning algorithms to identify patterns and features that distinguish one type of leaf from another. The goal of plant leaf identification using machine learning is to develop a system that can accurately identify a plant's species based on its leaves. This can be useful in a variety of contexts, such as agriculture, horticulture, and environmental conservation. For example, farmers can use the technology to identify and manage weeds, while conservationists can use it to monitor endangered species and their habitats.

Ayurveda is an ancient medicinal system that is practiced in India and has its origins in the Vedic times, approximately 5000 years ago. Ayurveda is considered to be the oldest healing science. This widespread and extensive use in the field of medicine makes Ayurveda the 'mother of healing'. These Ayurvedic plants are used for preparing medicines on a commercial basis. This has resulted in the production and marketing of Ayurvedic medicines to become a thriving industry with its annual turnover exceeding Rs 4000 crores. Because of this, the number of licensed Ayurvedic medicine manufacturers in India now exceeds 8500. Because of this increased commercialization of the Ayurvedic sector, several issues regarding the raw material quality used for their preparation need to be focused.

These plants are usually collected by tribal masses that are not professionally trained in the work of identifying the correct plants. Even the manufacturing units, at times, receive improper or substituted medicinal plants. Most of these manufacturing units do not have proper quality control mechanisms that can screen these plants. In addition to this, confusion due to the different local names of these plants makes the matters worse. Because of the usage of improper raw materials, Ayurvedic medicine becomes inefficient. It may result in certain unpredictable side effects too. As a result, an intelligent system needs to be developed that can properly identify the ayurvedic plant based on the leaf samples. This will, in turn, improve the quality of the medicine and will also maintain its credibility.

2. PROBLEM STATEMENT

The rapid pace of urbanization and industrialization has led to the loss of biodiversity and the endangerment of many plant species. It is essential to develop a system that can accurately

and quickly identify plant species to help with conservation efforts and restoration of degraded ecosystems. The use of machine learning in plant identification can significantly improve the accuracy and speed of the identification process. The challenge, however, lies in developing a robust and reliable algorithm that can accurately identify plant species even when dealing with variations in environmental conditions, growth stages, and the presence of pests and diseases. To identify the plant by their leaf, a type of artificial neutral network called convolutional neural network(CNN) is used.

3.OBJECTIVES

➢ Plant identification using a user-uploaded photograph would be one major goal of the proposed system.

➢ The objective of this project is to reduce manual work and increase the efficiency by the automatic identification of plants.

➢ Detect the Ayurvedic Medicinal Plants using CNN.

➢ To give formers a cost-effective and time-saving option.

➢ To deliver the most accurate results possible

4. LITERATURE REVIEW

Skanda H N[1] has proposed a leaf identification technique in which Leaves can be identified using digital fingerprint. This method works the same way a media recognition app works. By scanning the leaf by lasers, different depth points can be marked and connected to form an image which can be plotted against a graph. The area enclosed by graph form the unique digital fingerprint of the leaf which can be used to recognize the plant.PM KUMAR[2] describes image processing technique for identifying ayurvedic medicinal plants by using leaf samples. Forests and wastelands sources for over 80% of ayurvedic plants. There exists no predefined database of Ayurvedic plant leaves. A set of leaf images of medicinal plants were collected from the botanical garden. To improve the efficiency of plant identification system, machine learning techniques can be used over human visual perception as it is more effective. Weka is a collection of machine learning algorithms for data mining. It contains feature selection, regression, classification and pre-processing tools. Graphic user interface is used for accessing the functions. This proposed scheme uses some of the classifiers such as Support Vector Machine (SVM) and Multilayer perceptron (MLP). For reverting and classifying of data SVM is used. MLP is an artificial neural network which helps in routing the input data of one set to appropriate output pertaining to another set. The highest identification rate in SVM is 98.8% and 99% obtained in MLP.

H. A. Chathura Priyankara and D. K. Withanage [3] discusses the Computer-assisted android system for plant identification based on leaf image using features of SIFT along with Bag of Word (BOW) and SVM as classifiers. This identification method for android involves 8 stages. It employs client-server model of architecture. Server involves 2 main activities. The first activity is to train the SVM classifier to generate feature vector required for classification and then save it. The second activity is generation of feature vector with the help of photographs uploaded. These are uploaded by android client. The generated vector is used for identification by the SVM classifier. The process of training SVM involves SIFT descriptors along with Bag of Feature model that helps in generation of classifier. The generation of classifier involves 4 steps. In the first step, using the reduction method of data space SIFT descriptors are extracted from each leaf image belonging to the training data set. The second step is to cluster all the extracted features into feature bags using BOW methods. In the next step bow histograms are generated by taking all the images in the training dataset into consideration. In the final step all the histograms are passed to the SVM as the classification feature vector. SVM creates and saves the classifier in the server storage. The RGB image is converted into a greyscale image before extracting SIFT feature points as a pre-processing step. Following which involves extraction of key point and generating of descriptors by using SIFT algorithm that involves CBIR (content-based image retrieval) algorithm. Using k-means clustering method all the collected SIFT features from training dataset are clustered into several clusters. A histogram represents each image in the training dataset. Histograms are classified using multi-class linear support vector machine. Android implementation involves client application that consumes algorithm of leaf recognition. Dynamic Link Library (DLL) application is used to invoke

The Rafael Rojas [4] discusses the general steps for plant identification using pre-processing, feature extraction and their classifications. The availability of classic classification algorithms are not accessible, therefore it gave way for new methodologies applying data mining methods in specific domain. Considering the extraction process, initially we come across pre-processing where extraction of the available data is done to form images. These leaf images are transformed into quality binary images using normalization and segmentation processes. Most of the leaf datasets is available online and here we scale it in order to constrain the size. We also consider image normalization where brightness and contrast features are considered. Binary images of the leaves are obtained using leaf segmentation that is necessary in order to eliminate noise using morphological features. By using contour extraction, the geometric features of leaves are obtained. The Feature extraction process is used for plant recognition which considers various parameters such as area convexity, perimeter convexity and so on describing the leaf characteristics. Classification process is a supervised learning technique where we use ANN, SVM and KNN classifiers which improves classification accuracy.

The Rafael Rojas-Hernández and Asdrúbal López-Chau [4] describes the methods of shape feature extraction that is Scale Invariant Feature Transform (SIFT) and colour feature extraction Grid Based Colour Moment (GBCM) to identify plants which comprises of phases such as image acquisition, image processing, feature extraction, identification and performance measurement. The Image acquisition process mainly deals with acquiring datasets of different tree species. Image processing mainly aims to enhance image data required for further processing by discarding the undesired distortions. This process includes the phases of rotation, scaling and variations of leaf samples for further testing. Shape features and colour features are extracted using scale invariant feature transform and grid-based colour moment respectively. In SIFT both domains of spatial and frequency are considered. Geometric transforms makes it robust to illumination and noise. It also considers varying views of the object taking into consideration that helps in detection of the scale space extrema and an elaborate analysis is performed with respect to various features allowing the rejection of points corresponding to low contrast regions. The gradient magnitude and orientation is measured for each image sample. The orientation ranges from 360 degree and the Gaussian weighted circular window is used to measure the magnitude. The Grid-based colour moment is extracted using colour moment technique. Three parameters are used to calculate skewness, mean and standard deviation of an image. After acquiring these data, we go for an identification process based on Euclidian distance that determines the root square differences between values of a pair of objects considered. This methodology achieved an accuracy of 87.5% .

K. Li, Y. Ma and J. C. Príncipe [6] divides the identification of plant into three stages, they are: synthetic plant collection, spatiotemporal evolution model and automata extraction. In the first step, finite set of elements characterizes the plant development and growth in synthetic collection of plants. This finite set takes the indeterminate and complex shape. The mathematical formulation of underlying rules is named as L-system. An l-system is defined as the 3-tuple G = (V, w, P) system. The artificial regularity, also it introduces randomness to its production. In a synthetic plant collection, image processing and feature extraction method is also used. The L-systems are also visualized using truth table using turtle interpretation and saved as JPEG images to simulate the real plants. To detect the main axis and root of the plant, Hough transform is used. In the second step, that is, the spatiotemporal evolution model, KAARMA network models a dynamic system as defined by the general continuous non-linear state transitions functions and an observation function. To train a STEM, kernel adaptive KAARMA is used. In the third step, that is, the automata extraction, the discrete finite automation (DFA)is used where all the state transitions are uniquely determined by input symbols, from an initial state. The DFA is used to model the discrete time dynamical system in the discrete state space. A DFA can be represented in two ways, state transitions or lookup table. The analytical descriptor of a languages known as an Automata. The

DFA also validates the corresponding regular grammar produced by the language.

5.METHODOLOGY

1.DATA SET: The dataset for training is obtained from Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI). LIDC and IDRI consist of 1000 CT scans of both large and small tumors saved in Digital Imaging and Communications in Medicine (DICOM) format.

2. IMAGE SEGMENTATION: The segmentation of photographs is the phase where the visual image is partitioned into several parts. This normally helps to identify artifacts and boundaries. The aim of segmentation is to simplify the transition in the interpretation of a picture into the concrete picture that can be clearly interpreted and quickly analyzed.

3.PRE-PROCESSING: In preprocessing stage, the median filter is used to restore the image under test by minimizing the effects of the degradations during acquisition. Various preprocessing and segmentation techniques of lung nodules are discussed in. The median filter simply replaces each pixel value with the median value of its neighbors including itself. Hence, the pixel values which are very different from their neighbors will be eliminated.

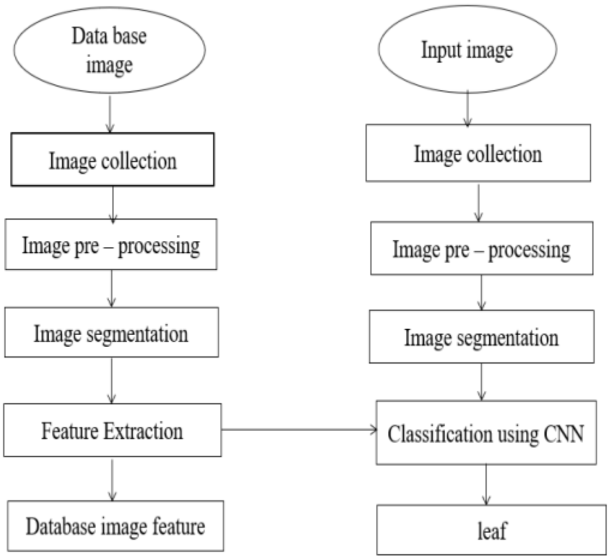


Figure1: Data flow Diagram of Training and Testing Phase

Image processing stages involves

1.Image Collection

2. Image Preprocessing

3. Image Segmentation

4. Feature Extraction

5. Training

6. Classification

4. CONVOLUTIONAL NEURAL NETWORKS: A CNN is type of a DNN consists of multiple hidden layers such as convolutional layer, RELU layer, Pooling layer and fully connected a normalized layer. CNN shares weights in the convolutional layer reducing the memory footprint and increasing the performance of the network. The important features of CNN lie with the 3D volumes of neurons, local connectivity, and shared weights. A feature map is produced by the convolution layer through the convolution of different sub-regions of the input image with a learned kernel. Then, anon- linear activation function is applied through the ReLu layer to improve the convergence properties when the error is low. In the pooling layer, a region of the image/feature map is chosen as the pixel with the maximum value among them or average values is chosen as the representative pixel so that a 2x2 or 3x3 grid will be reduced to a single scalar value. This results a large reduction in the sample size. Sometimes, traditional Fully-Connected (FC) layer will be used in conjunction with the convolutional layers towards the output stage.

Front-End Development Using Python Tkinter: Modern computer applications are user-friendly. User interaction is not restricted to console- based I/O. They have a more ergonomic graphical user interface (GUI) thanks to highspeed processors and powerful graphics hardware. These applications can receive inputs through mouse clicks and can enable the user to choose from alternatives with the help of radio buttons, dropdown lists, and other GUI elements

The data training in our CNN model has to satisfy following constraints: 1) No missing values in dataset. 2) The dataset must distinctly be divided into training and testing sets, either the training or the testing set shouldn't contain any irrelevant data out of our model domain in case of an image dataset all the images must be of the same size, one uneven distribution of image size in our dataset can decrease the efficiency of our neural network. 3) The images should be converted into black and white format before feeding it into the convolution layer because reading images in RGB would involve a 3-D num Py matrix which will reduce the execution time of our model by a considerable amount. Any kind of corrupted or blurred images should also be trimmed from the database before feeding it into the neural network. This is all about data pre-processing rules, let us dive right into the working of the convolution neural network

A CNN IS COMPOSED OF SEVERAL KINDS OF LAYERS: Convolutional layer: creates a feature map to predict the class probabilities for each feature by applying a filter that scans the whole image, few pixels at a time. Pooling layer (down-sampling): scales down the amount of information the convolutional layer generated for each feature and maintains the most essential information (the process of the convolutional and pooling layers usually repeats several times). Fully connected input layer: flattens the outputs generated by previous layers to turn them into a single vector that can be used as an input for the next layer. Fully connected layer: Applies weights over the input generated by the feature analysis to predict an accurate label.

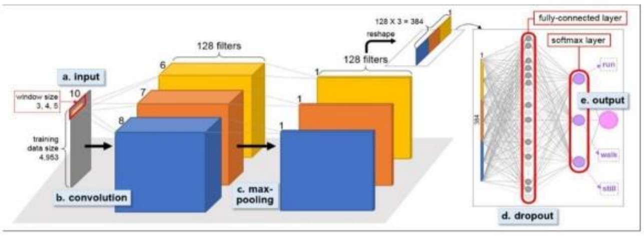


Figure2: Convolutional Neural Network General Architecture

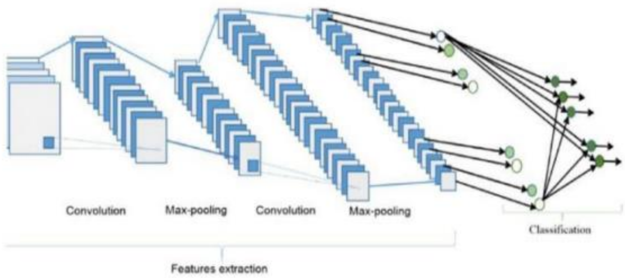
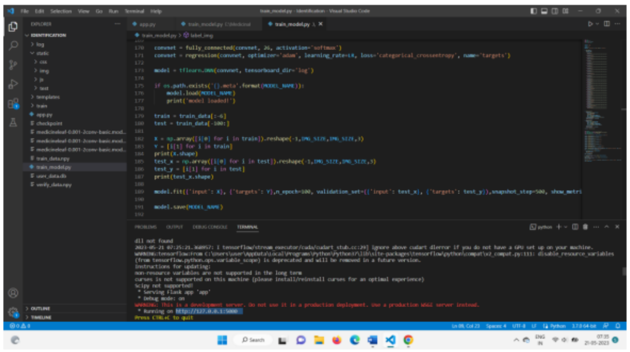
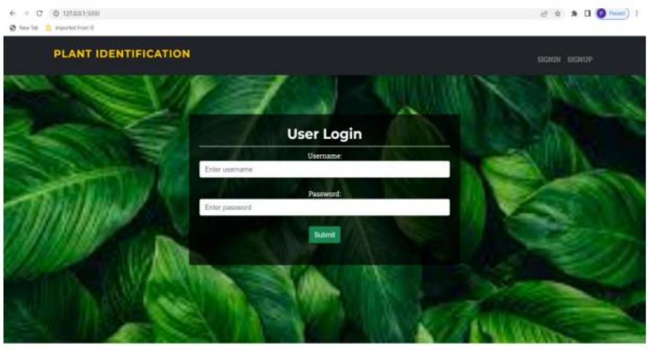


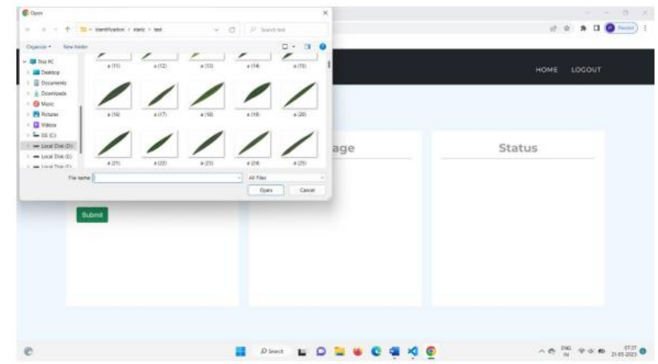
Figure2: Convolutional Neural Network Layers

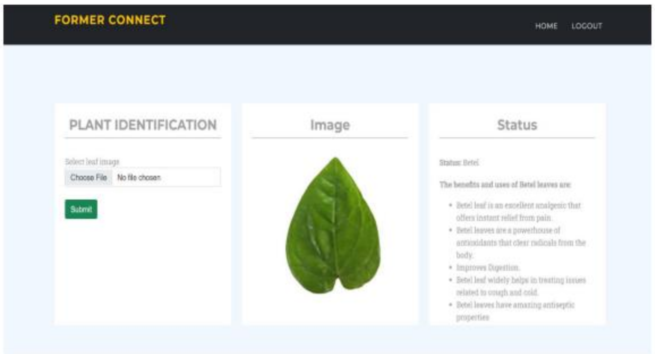
6.RESULTS

Results of this project are as show in following figures.









- Here user has to Login with user ID and Password.
- Then select the leaf which is to be identified.
- The CNN algorithm used identifies leaf and it displays name of plant and its medicinal features.

7.APPLICATIONS

Applications of this project are listed as bellow

1.Agriculture: In agriculture we can use this project to identify plant diseases

2. Conservation: We can use this project to conserve some traditional ayurvedic plants.

3. Forestry: We can use this project to identify medicinal plants and preserve them.

4.Horticulture: Here we can use this project to Identify flowers in gardening.

5.Education: It can be used in Botony labs to demonstrate identification and features of plants.

CONCLUSIONS

Plants are necessary for human survival. Herbs, particularly, are employed by indigenous populations as folk medicines from old period .Recent improvements in analytical technology have made it much easier to identify herbs depending on scientific evidence. As a result, a simple and reliable method for identifying herbs is required. Herbal identification anticipated to benefit from the combination of computation and statistical examination. This nondestructive technique will be the preferred approach for quickly identifying herbs, especially for individuals who cannot able to use expensive analytical equipment

ACKNOWLEDGEMENT

We take this opportunity to Acknowledge Our principal head of the department and All authors take part in the discussion of the work described in this article. All The authors would like to thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

REFERENCES

1. Skanda H N , Smitha S Karanth , Suvijith S , Swathi K S, 2019, Plant Identification Methodologies using Machine Learning Algorithms, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 08, Issue 03 (March – 2019),
2. P. M. Kumar, C. M. Surya and V. P. Gopi, "Identification of ayurvedic medicinal plants by image processing of leaf samples," 2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata, 2017, pp. 231-238. DOI: 10.1109/ICRCICN.2017.8234512
3. H. A. Chathura Priyankara and D. K. Withanage, "Computer assisted plant identification system for Android," 2015 Moratuwa Engineering Research Conference (MERCon), Moratuwa, 2015, pp. 148-153. DOI: 10.1109/MERCon.2015.7112336
4. Rafael Rojas-Hernández and Asdrúbal López-Chau, "Plant identification using new geometric features with standard data mining methods", Networking Sensing and Control (ICNSC) 2016 IEEE 13th International Conference on, pp. 1-4, 2016.
5. Che Hussin, N. A., Jamil, N., Nordin, S., & Awang, K. (2013). Plant species identification by using scale invariant feature transform (SIFT) and grid based colour moment (GBCM). In 2013 IEEE Conference on Open Systems, ICOS 2013 (pp. 226-230). [6735079]
6. K. Li, Y. Ma and J. C. Príncipe, "Automatic plant identification using stem automata," 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP), Tokyo, 2017, pp. 1-6. DOI: 10.1109/MLSP.2017.8168147. J. X. Du, D. S. Huang, X. F. Wang, and X. Gu, "Computer-aided plant species identification (CAPSI) based on leaf shape matching technique," *Transactions of the Institute of Measurement and Control*, vol. 28, no. 3, pp. 275–285, 2006.
7. S. Prasad, P. Kumar, and R. C. Tripathi, "Plant leaf species identification using curvelet transform," in *Proceedings of the 2011 2nd IEEE International Conference on Computer and Communication Technology*, pp. 646–652, Harbin, China, August 2011.
8. S. G. Wu, F. S. Bao, E. Y. Xu, Y. X. Wang, Y. F. Chang, and Q. L. Xiang, "A leaf recognition algorithm for plant classification using probabilistic neural network," in *Proceedings of the 2007 IEEE International Symposium on Signal Processing and Information Technology*, pp. 11–16, Giza, Egypt, December 2007.
9. H. Fu, Z. Chi, D. Feng, and J. Song, "Machine learning techniques for ontology-based leaf classification," in *Proceedings of the 8th IEEE International Conference on Control, Automation, Robotics and Vision Conference*, vol. 1, pp. 681–686, Niigata, Japan, April 2004.
10. F. Han, X. Qiao, Y. Ma et al., "Grass leaf identification using dbN wavelet and CILBP," *Advances in Multimedia*, vol. 2020, Article ID 1909875, 8 pages, 2020.
11. X. Gu, J. X. Du, and X. F. Wang, "Leaf recognition based on the combination of wavelet transform and Gaussian interpolation," in *Proceedings of the ICIC 2005: Advances in Intelligent Computing*, pp. 253–262, Anhui, China, August 2005.
12. S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523–3542, 2022.
13. Hossain and M. A. Amin, "Leaf shape identification based plant biometrics," in *Proceedings of the 2010 13th IEEE International Conference on Computer and Information Technology*, pp. 458–463, Bradford, UK, December 2010.
14. A. H. Kulkarni, H. M. Rai, K. A. Jahagirdar, and P. S. Upparamani, "A leaf recognition technique for plant classification using RBPNN and Zernike moments," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 2, pp. 984–988, 2013.