IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

Plant Identification based on Machine Learning Techniques

Ruturaj A. Wakchaure[1st], Minakshi M.Sonawane [2nd] , Payal T. Budhe[3rd]

1, 2, 3, Department of Dr. D. Y patil college of engineering and innovation, Varale , Talegov , India,

Abstract: Life on Earth depends heavily on plants. There are many different plant species accessible, and the number keeps growing. For diverse employment domains, different social groups like forestry, farming, environmentalists, and educators require species knowledge. Because of this, identifying species is of interdisciplinary interest. However, since this calls for specialized knowledge, it becomes tiresome and difficult work for non-experts who know little to nothing about common botanical phrases. However, this process may become considerably simpler because of developments in the disciplines of computer vision and machine learning. Image capture, pre-processing, feature extraction, and classification are the typical four phases in plant identification. Photos from the Swedish leaf dataset, which contains 1,125 photos, were used in this investigation.

Keywords **- Identification of Plant Species, Segmentation, SVM.**

I.Introduction

Plants are an essential component of life on our planet. They provide us with food, medication, air to breathe, and a host of other necessities that make life worthwhile. They serve as the foundation for all life. However, accurate plant identification is beyond the capabilities of the average individual because it necessitates specialized knowledge, which can only be attained by professionals with a background in botany. Moreover, because there are an infinite number of plant species, not even botanists are familiar with all the plants that exist on our planet. As a result, only a very small group of people are capable of carrying out the duty of plant identification. Botanists employ a set of clearly defined traits of a plant as a key for identification in manual identification, which is useful for identifying plant species.

The properties of an unfamiliar plant's "shape," "texture," "color," and "venation" are among the identifying keys. When carefully considered, these traits finally result in the ideal species. Furthermore, a person of ordinary intelligence is incapable of identifying a plant species from a natural environment due to the high taxonomical knowledge that is required. Therefore, traditional methods for identifying plant species appear unworkable for regular people and are difficult for expert taxonomists as well. Even for experienced botanists, identifying species is frequently a difficult undertaking.

Smartphones and digital cameras are widely available and widely used in this era of the digital world. Digital images have developed into an essential component of several industries as a result of this technical breakthrough, including computer vision, face identification, plant recognition, and health informatics [6]. Additionally, automation of the species identification procedure has been a goal of technological advancement in the area of image processing and the toolkits available to accomplish it. The process of automating species identification has been attempted on occasion, with some degree of success. The authors of [7] made one of the most effective attempts and created Leaf snap, the most comprehensive system of its kind for identifying plant species.

II. LITERATURE REVIEW

Table 1 provides a comparison of methods employed in published literature.

Table: Comparison of existing literature

Sr. No.	Title of Name	Dataset	Features/Techniques	Accuracy
1.	A leaf recognition algorithm for plant classification using a probabilistic neural network.	Flavia- (32) images.	Shape + vein Morphological descriptors (MD), PCA, SVM.PNN	90.31
2.	A shape-based approach for leaf classification using multiscale triangular representation.	Swedish	TESLA	90.40 95.20 95.73 96.53
3.	Plant recognition based on Intersecting optical model.	ICL	Shape+ texture, EnS, CDS, segmentation, KNN, SC(shape context)	95.87 with the given of (ICL1) 94.21 with the given of (ICL2)
4.	Feature decision-making ant colony optimization system for an automated recognition of plant species.	Flavia	Shape color + texture + vein, Edge Detection, Morphological feature. GLCM, MD	96.25
5.	An Automatic Leaf Based Plant Identification System.	Flavia	Shape + vein, GLCM, MD	87.40
6.	Automatic recognition of medicinal plants using the machine.	2017	Own	90.10
7.	Plant Classification Using Image Processing and Neural Network.	2018	ICL, CNN, KNN	99% (training) 96% (testing)

Many studies have been conducted over the last ten years in an effort to create reliable and effective plant identification systems. In their plan, they have produced a dataset they call Flavia, which is utilized as the standard dataset for the work of many other academics. It has 1907 pictures of leaves from 32 distinct plant species. Based on shape and vein structure, they were able to extract from the leaf photos 12 digital morphological elements and 5 fundamental geometric aspects for their investigation. Additionally, the dimensions of the input vector that would be supplied to the probabilistic neural network (PNN) for classification were reduced using principal component analysis (PCA) derived from a set of distinctive traits dubbed "Leaf Width Factor (LWF)" along with 9 other morphological features. The PNN was then used to classify the features of the leaf form using these features as inputs. The network was trained using 1200 leaf images in total, and when it was validated using 10-fold cross-validation, it had a maximum accuracy of 94% at the eighth fold. The achieved accuracy was 91.41% on average. Proposed a reliable technique for classifying leaf images that makes use of both global and local information. They used SIFT (Scale Invariant Feature Transform) and shape context (SC) as global and local features, respectively. On the ICL dataset, classification using K-nearest neighbor (k-NN) yielded an overall accuracy of 91.30%. created a technique that uses shape cues to automatically extract classification information. Forward-back propagation neural networks have been used for classification. Then, using the ICL dataset, this technique was evaluated. It achieved 99% accuracy for training photos and 96% accuracy for testing images.

Since form feature descriptors are thought to be the most discriminating feature in plant identification, the majority of research in the available literature has used them for feature extraction. One of the most important processes in image processing is feature extraction; therefore, features characterizing different parts of a plant leaf must be taken into consideration before the plant image is finally classified. Additionally, in instances where the leaves are altered or not fully developed, texture and color characteristics can better characterize a leaf image. The ideal combinations of texture and color features have been extracted for categorization in our study.

III. METHOD AND TECHNIEQUES

A collection of medicinal plant leaf photos was gathered from the botanical garden. Machine learning approaches can replace human visual perception in the plant identification system since they are more efficient.
Database: When thinking about the extraction process, the first thing we come across is pre-processing, where the available data is extracted to create images. Through the use of normalization and segmentation techniques, these leaf images are converted into high-quality binary images.

- Online Dataset

3.1 **Online Dataset:** we collect data from online sources also take image normalization and brightness and contrast features into account. Leaf segmentation, which is required to remove noise using morphological features, is used to create binary images of the leaves.

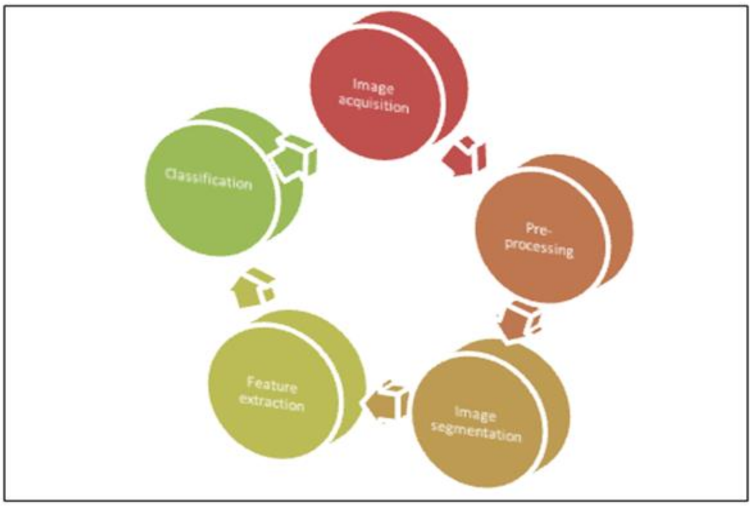

Figure 1: Flow of Operation of the System

3.2 Image Acquisition

Getting a picture of the plant is the first step in the identification process. The entire plant—a leaf, a bloom, a stem, or even the fruits can be photographed. According to the authors, there are three sorts of images depending on how they were captured: scans, pseudo-scans, and photos. In the scan and pseudo-scan categories, the leaf images are taken using the scanning method and photography, respectively; that is, they are taken indoors against a plain background. Images of plants taken in their native habitats are included in the third category. Researchers frequently employ scans and pseudo-scans because they are simple to study

3.3 Image Pre-processing

Pre-processing a picture is a crucial step since it improves the quality of the image for subsequent processing. This stage is essential since noise is present in images by nature, which could lead to less accurate classification. It is done to treat the corrupted data and remove the noise that interferes with identification.

1.3.1 Image Segmentation

In order to extract the region of interest (ROI), picture segmentation is a crucial and vital step in the image analysis process. The end result of the segmentation step is an image that has been separated into distinct segments after each pixel in the image has been individually analysed and grouped with other pixels in the image that share the same qualities. To put it another way, it's a procedure in which each pixel is given a label based on a set of criteria, and the pixels that have those same characteristics are then grouped together. In general, the images are divided into many pieces (or segments) based on the texture, colour, grey level, pixel intensity value, etc.

4. Feature Extraction

Feature extraction is carried out after pre-processing and segmenting the image into the appropriate region of interest. It is thought to be one of the most significant processes in pattern analysis and image processing. Another term for feature extraction is dimensionality reduction. An image by its very nature contains a lot of data, much of it is redundant and cannot be handled without a significant amount of processing power and memory [25]. Therefore, feature extraction is carried out to lower the number of variables for additional image processing. Thus, selecting the proper set of attributes to best characterize the image becomes crucial. In our investigation, the image is represented by a mix of texture and color properties. The K-Nearest Neighbour algorithm, which is based on the Supervised Learning technique, is one of the most basic machine learning algorithms. When a new data point is added, the K-NN algorithm classifies it based on how similar the existing data is to it and stores it all. This means that as fresh data is generated; it can be quickly categorized using the K-NN method [17]. Identification of Skin Disease Using Machine Learning 107 Fig. 2. Support Vector Machine using Linear Regression [18].

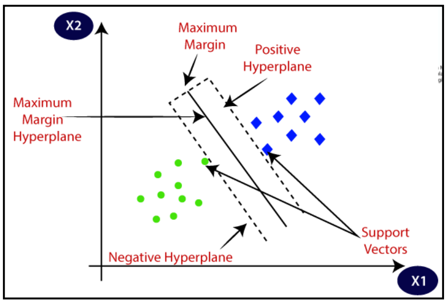

Figure 2: Support Vector Machine

5. CLASSIFICATION:

In our study, classification generally refers to ascribing a particular plant species to the image based on the feature set extracted. To put it another way, classification is the process of determining the class label of a new input image based on the training dataset's previously learned information. To classify the new data input for our investigation, we employed a supervised classification technique in which the labels of the classes (in this case, plant species) were already known.

5.1 CONVOLUTION NEURAL NETWORK:

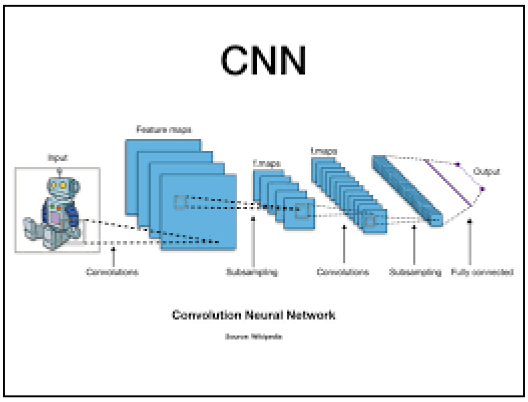

Figure 3: Convolution Neural Network

A typical Deep Learning neural network architecture in computer vision is the convolution neural network (CNN). A computer can comprehend and analyse visual data or images thanks to the field of artificial intelligence known as computer vision.

6. RESULTS AND DISCUSSION:

The Swedish dataset, which includes 1,125 photos of 15 different species was used to evaluate the proposed methodology. Typically, the collection includes pictures of a single leaf of a plant. The majority of the leaves are mature and in good condition. Only a very small percentage of the leaves are slightly bent or twisted. The dataset exhibits substantial intra-class similarity and, in a very small number of instances, inter-class similarity. The dataset's photos have been downsized to [300, 400] for this investigation. First, filtering and contrast enhancement are used to remove any unwanted information or noise from the input image. Since the photos in the dataset are colored and an RGB image contains redundant information, this step is necessary.

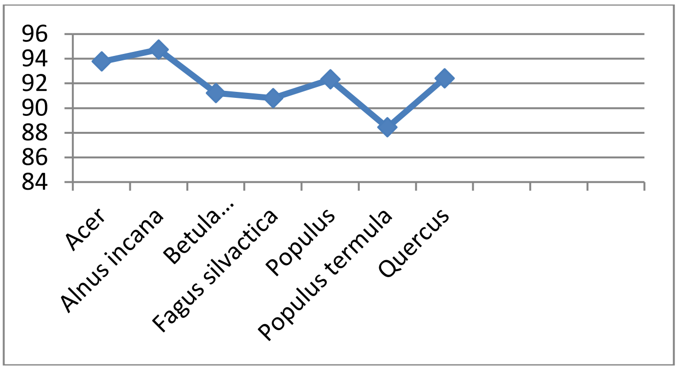

Figure 4: Result with Different using Parameter

7. CONCLUSION:

In order to categorize photographs of plant leaves, a method for automatic plant species identification has been developed in this research. This method makes use of computer vision and machine learning techniques. Phases of the study included image pre-processing, image segmentation, feature extraction, and finally image classification. In order to classify the data, a mixture of texture and color features 5 and 4, respectively was retrieved. The system achieved an average accuracy of 93.26% while being evaluated on a Swedish dataset. 15 different plant species could be automatically categorized by the model.

REFERENCES

[1] S. G. Wu, F. S. Bao, E. Y. Xu, Y.-X. Wang, Y.-F. Chang, Q.-L. Xiang, "A leaf recognition algorithm for plant classification using probabilistic neural network," In Proceedings of 2007 IEEE International Symposium on Signal Processing and Information Technology, pp. 11-16, 2007.

[2] S. Mouine, I. Yahiaoui, A. Verroust-Blondet, "A shape-based approach for leaf classification using multiscale triangular representation," In Proceedings of the 3rd ACM conference on International conference on multimedia retrieval, pp. 127-134, 2013.

[3] Z. Wang, X. Sun, Y. Ma, H. Zhang, Y. Ma, W. Xie, Z. Wang, Y. Zhang, "Plant recognition based on the intersecting cortical model," In Proceedings of 2014 International Joint Conference on Neural Networks (IJCNN), pp. 975-980, 2014.

[4] M. A. J. Ghasab, S. Khamis, F. Mohammad, H. J. Fariman, "Feature decision-making ant colony optimization system f an automated recognition of plant species," Expert System with Applications, vol. 42, no.5, pp. 2361-2370, 2015.

[5] N. Ahmed, U. G. Khan, S. Asif, "An Automatic Leaf Based Plant Identification System," Science International, vol. 28, no.1, pp. 427-434, 2016.

[6] A. Begue, V. Kowlessur, F. Mahomoodally, U. Singh, S. Pudaruth, "Automatic recognition of medicinal plants using machine learning techniques," Int. J. Adv. Comput. Sci. Appl, vol. 8, no. 4, pp. 166-175, 2017.

[7] M. M. Amlekar, A. T. Gaikwad, "Plant Classification Using Image Processing and Neural Network," Data Management, Analytics and Innovation, vol. 839, pp. 375-384, 2018.