Plant's Identification Methodologies using Machine Learning Algorithms

Shreyash Dhakate[1], Vrushabh Ramteke[2], Divya Shahare [3], Zakhil Landge[4]
U.G. Student, Department of Computer Science & Engineering, J.D. College of Engineering and Management, Nagpur, Maharashtra, India
Prof. Dipali P. Pethe.[5]
Asst. Prof, Department of Computer Science & Engineering, , J.D. College of Engineering and Management, Nagpur, Maharashtra, India

ABSTRACT:

Plants are the backbone of all life and there are about 40 million plant species on Earth providing us with oxygen, food and many essential products helping for the existence of human life. A good understanding of plants is essential to help in the process of identification of new or rare plant species to improve the balance in the ecosystem. The matching of specimen plant to a known Taxon is termed as plant identification which implies assigning a particular plant to a known taxonomic group by comparing certain characteristics. Plant identification which has evolved over hundreds of years ago depends on the criteria and the system used. As identification enables us to retrieve the appropriate facts associated with different species to serve a particular kind of application, plant identification is essential. This paper includes various methodologies of numerous authors who have worked on different plant identification techniques.

Keywords:-Machine Learning, Feature Extraction, Plant Species Identification, Segmentation, SVM

I. INTRODUCTION

In a world where the intricate interplay between plant diversity, ecological equilibrium, and human sustenance is paramount, the accurate identification of plant species assumes critical importance. However, the task of identifying and understanding various plant species can be formidable, particularly for non-experts, owing to the vast number of species and their nuanced differentiating characteristics. To address this challenge and harness the potential of cutting-edge technology, the "Plant Reader using ML" project emerges as an innovative initiative at the intersection of botany, technology, and conservation.

The central aim of the project is to leverage the power of machine learning to simplify and enhance the process of plant species classification.

Through the seamless integration of sophisticated image recognition techniques and modern deep learning methodologies, this project aspires to provide a user-friendly and efficient solution for identifying and categorizing plant species. By enabling botanists, horticulturists, educators, and enthusiasts alike to accurately identify plants based on images of their leaves, flowers, and distinctive attributes, the project seeks to bridge the gap between expert knowledge and broader accessibility. Enhance the accuracy of plant identification through continuous learning and updates using user-contributed data. Facilitate the democratization of botanical knowledge by making plant identification accessible to non-experts and enthusiasts. Contribute to ecological conservation efforts by enabling faster and more efficient identification of plant species in various ecosystems.

Plants are of central importance to natural resource conservation. Plant species identification provides significance information about the categorisation of plants and its characteristics. Manual interpretation is not precise since it involves individual's visual perception. Sampling and capturing digital leaf images are convenient which involves texture features that help in determining a specific pattern. The most important feature to distinguish among plant species are venation and shape of a leaf. As information technology is progressing rapidly, techniques like image processing, pattern recognition and so on are used for the identification of plants on basis of leaf shape description and venation which is the key concept in the identification process. Varying characteristics of leaves are difficult to be recorded over time. Hence it is necessary to leaves are used in most of the plant identification methodologies due to their attractive properties and availability throughout the year.

II. LITERATURE SURVEY

In the past decade a lot of research has been done in order to develop efficient and robust plant identification systems. Have proposed on of the earliest plant identification system. In their scheme, they have created their own dataset named Flavia, which has been used by various other researchers as standard dataset for their work. It consists of 1907 leaf images of 32 different plant species. In their study, they extracted 5 basic geometric and 12 digital morphological features based on shape and vein structure from the leaf images. Further, principal component analysis (PCA) was used to reduce the dimensions of input vector to be fed to the probabilistic neural network (PNN) for classification. They used a three-layered PNN which achieved an average accuracy of 90.32%. Hossain et al. extracted a set of unique featured called "Leaf Width Factor (LWF)" with 9 other morphological features using the Flavia dataset. These features were then used as inputs to PNN for classification of leaf shape features. A total of 1200 leaf images were used to train the network and then PNN was tested using 10-fold cross validation, which achieved maximum accuracy of 94% at 8th fold. The average accuracy attained was 91.41%. Wang et al. [9] proposed a robust method for leaf image classification by using both global and local features. They used shape context (SC) and SIFT (Scale Invariant Feature Transform) as global and local features respectively. K- nearest neighbour (k-NN) was used to perform classification on ICL dataset which achieved an overall accuracy of 91.30%.

Authors in developed a scheme which extracted 12 common digital morphological shape and vein features derived from 5 basic features. They implemented both k-NN and support vector machine (SVM) which attained an accuracy of 78% and 94.5% respectively when tested on Flavia dataset. Pham et al. in their computer-aided plant identification system compared the performance of two feature descriptors i.e. Histogram of oriented gradients (HOG) and Hu moments. For classification, they selected SVM due to its ability to work with high dimensional data. They obtained accuracy of 25.3% for Hu moments and 84.68% for HOG when tested with 32 species of Flavia dataset. Moline et al. , in their study introduced new multiscale shape-based approach for leaf image classification. They studied four multiscale triangular shape descriptors viz. Triangle area representation (TAR), Triangle side length representation (TSL), Triangle oriented angles (TOA) and Triangle side lengths and angle representation (TSLA). They tested their system on four image datasets: Swedish, Flavia, Image CLEF 2011 and Image CLEF 2012. With Swedish dataset they computed classification rate as 96.53%, 95.73%, 95.20% and 90.4% for TSLA, TSL, TOA and TAR respectively using 1-NN. Authors in proposed a method for plant identification using Intersecting Cortical Model (ICM) and used SVM as the classifier. This study used both shape and texture features viz. Entropy Sequence (Ends) and Centre Distance Sequence (CDS). They attained accuracy of 97.82% with Flavia dataset, 95.87% with ICL1 and 94.21% with ICL2 (where ICL1 and ICL2 are subsets of ICL dataset)

III. METHODOLOGY

The paper, describes image processing technique for identifying Ayurveda medicinal plants by using leaf samples. Forests and wastelands sources for over 80% of Ayurveda plants. There exists no predefined database of Ayurveda 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. The paper discusses the Computer-assisted android system for plant identification based on leaf image using features of SIFT along with 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 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 communication between the web service and the OpenCL implementation of image processing. This methodology obtains an average accuracy of about 96.48% on 20 different species. The paper 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 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%.

The paper discusses about the leaf features that uses shape contour which is represented mathematically. The distance travelled from the starting point is denoted by arc length, the periodic function of curve segment which is centred on the point depicts the perpendicular distance from that point to the straight line which connects it. The convexity and concavity measures of the arc are then considered, on the basis of these observed values functions operate on two different multi-scale shape information features. Capturing of the shape details is focused by smaller scale and the global properties are reflected by large scale. To achieve scale invariance consideration, maximum value is taken to normalize it and then subjected to Fourier transforms describes about the shape, in addition with standard deviation methodologies to enhance the power of discrimination of the shape descriptor. Then we consider the dissimilarity between the obtained shapes.

The extraction process is done on phone itself where bandwidth reduces drastically. Then the server returns the closest matches of the databases opened showcasing the result in a webpage. The method proposed is 30 times faster obtaining the response almost instant. This paper briefs about the idea of a graphical identification tool which uses computer aided system for automatic identification technique. Graphical tool describes three main components namely graphical interface, identification of plants and result interface. The graphical interface characterises plants based on leaf, venation etc. as graphical icons. After this, comparison of similarities between the user-defined inputs with respect to the original database containing plants are subjected for the identification process. Finally the result interface provides the result of identification and also provides sorting of plants present in the database in a decreasing order based on their similarities. Even though plant identification process is made easier with the graphical tool, the feature extraction process still remains as base for the identification process. This might sometimes lead to improper identification. So, the automatic plant identification technique is used to overcome the disadvantages of the graphical tool process. In automatic synthetic plant collection, spatiotemporal evolution model and automata extraction. In the first step, finite set of elements characterises the plant development and growth in synthetic collection of plants. This finite set takes the indeterminate and complex shape

IV. ALGORITHM USED:-

SVM-Support Vector Machines (SVM) is a supervised machine learning algorithm used for classification and regression analysis. It is a powerful and versatile algorithm that works well in both linear and non-linear data separation.

CNN-A Convolutional Neural Network (CNN) is a type of deep learning model designed specifically for processing and analysing structured grid-like data, such as images and video.

ANN-Artificial Neural Networks (ANNs) are a fundamental component of many machine learning and deep learning models used in various applications, including plant species identification.

V. PROPOSED SYSTEM:

The flow of operation of the proposed system is shown in figure 1. The details of each step are discussed in the subsequent sub-sections.

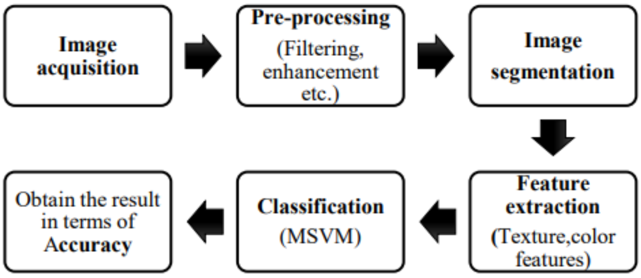

Fig. 1. Flow of operation of the system

Image pre-processing Image pre-processing is an important step as it helps to enhance the quality of image for further processing. This step is necessary as an image inherently contains noise and this may result in lower classification accuracy. It is performed to remove the noise that hampers the identification process and handle the degraded data. A series of operations are followed to improve the image of the leaf which include, converting the RGB image to grayscale, then from grayscale to binary, followed by smoothing, filtering etc. . Pre-processing mechanism used in this paper contains noise handling along with resizing operation and image enhancement.

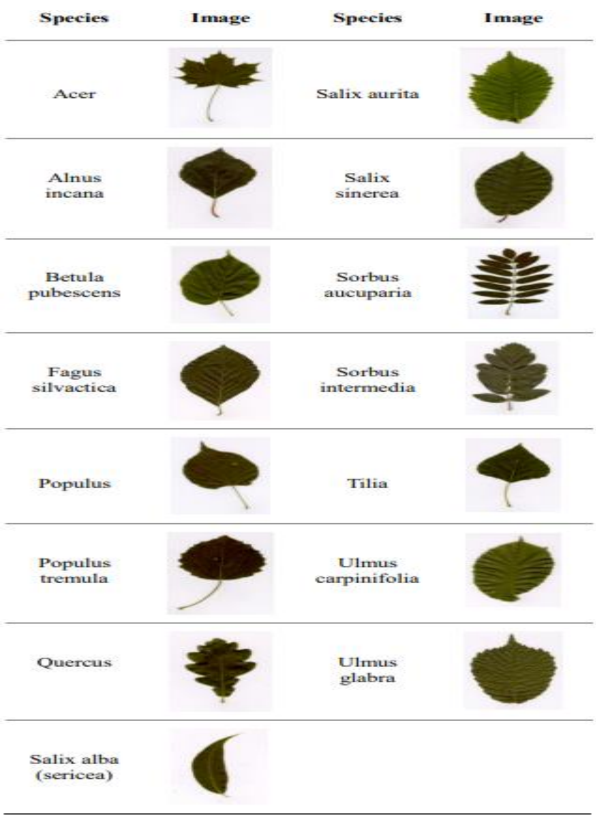

Fig. 2 Plant Species in Dataset

VI. IMAGE ENHANCEMENT

Since we are working on colour images, image enhancement becomes an essential step to perform. Moreover, the next step involves colour image segmentation for which the image contrast and texture needs to be enhanced to obtain better results. Image enhancement removes any redundant pixels present in the colour image before performing segmentation. In our study we have enhanced the contrast of the image by contrast stretching which improves the contrast in an image by expanding the dynamic range of intensity values it contains. This step is followed by contrast adjustment which saturates the top one percent and bottom one percent of all pixel values

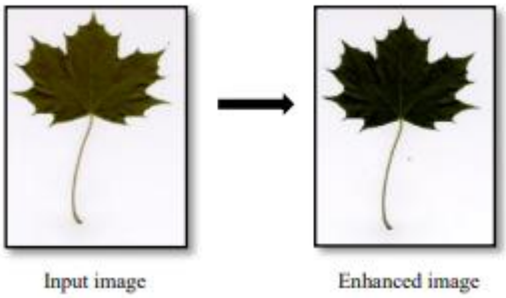

FIG. 3. IMAGE ENHANCEMENT

VII. FUTURE SCOPE

In the future, the field of plant identification using machine learning algorithms holds promise for further advancements, including the development of innovative methods such as digital fingerprinting and touchscreen-based tracing for leaf recognition. These emerging techniques, along with the continued growth of digital data storage and accessibility, are expected to enhance the accuracy and efficiency of plant species identification, making it even more accessible and valuable for ecological conservation and botanical knowledge dissemination.

VIII. CONCLUSION

Most of the methodologies mentioned above require the usage of a reference table or an inbuilt data set. This means a pre-analysis and initial collection of data has to be done in order to be used as reference for future comparison. Avoiding this preliminary step is difficult, but the content can be stored in a more efficient way with the advance of ml where digital data can be stored in the form of logical pools. New methods can be used based on the advancement of the present technology. Therefore, we propose the following new methods.

1) Leaves can be identified using digitally 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.

2) Leaf recognition can be done by tracing its outline on a digital screen such as a camera. Just like how a Swype keyboard on our phones work, the path taken by the user's finger to trace the leaf image can be linked to a present algorithm. Once the finger is lifted from the screen, the Path is mapped and the similar path is extracted from dataset and leaf is recognized.

IX. REFRENCES

[1] P. M. Kumar, C. M. Surya and V. P. Gobi, "Identification of Ayurveda 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.

[2] H. A. Chatham Priyanka 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.

[3] 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.

[4] 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] IEEE Computer Society. DOI: 10.1109/ICOS.2013.6735079.

[5] B. Wang, D. Brown, Y. Gao and J. L. Salle, "Mobile plant leaf identification using smart-phones," 2013 IEEE International Conference on Image Processing, Melbourne, VIC, 2013, pp. 4417- 4421.DOI: 10.1109/ICIP.2013.6738910.

[6] N. H. Pham, T. L. Le, P. Grard and V. N. Nguyen, "Computer aided plant identification system," 2013 International Conference on Computing, Management and Telecommunications (ComManTel), Ho Chi Minh City, Vietnam, 2013, pp. 134-139. DOI: 10.1109/ComManTel.2013.6482379.

[7] 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.

[8] M. Martinez-Ramon and J. Arenas-Garcia, "An Adaptive Combination of Adaptive Filters for Plant-Identification" Digital Signal Processing, 2002. DSP 2002. 2002 14th International Conference on, Volume: 2, DOI:10.1109/ICDSP.2002.1028307.

[9] Z. Q. Zhao, Y. Hong, P. Zheng and X. Wu, "Plant identification using triangular representation based on salient points and margin points," 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, 2015, pp. 1145-1149. DOI: 10.1109/ICIP.2015.7350979.

[10]P. M. Kumar, C. M. Surya and V. P. Gopi, "Identification of Ayurveda 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.

[12] 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.

[13] Rafael Rojas-Hernández and Asdrúbal López-Chau, "Plant identification using new geometric features with standard datamining methods", Networking Sensing and Control (ICNSC) 2016 IEEE 13th International Conference on, pp. 1-4, 2016.