

A Identification of Piper Plant Species Based on Deep Learning Networks

A.Pravin^a, C.Deepa^b

^aResearch Scholar in Computer Science, Sri Ramakrishna College of Arts and Science, Coimbatore

^bAssociate Professor in IT, Sri Ramakrishna College of Arts and Science, Coimbatore

^apravinresearchscholar@gmail.com, ^bdeepapkd@gmail.com

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

Abstract: Medicinal plants are widely used in non-industrialized societies, mainly because they are readily available and cheaper than modern medicines. These herbs that have medicinal quality provide rational means for the treatment of many internal diseases, which are otherwise considered difficult to cure. This is the reason why medicinal plant related analysis is growing in popularity across the researchers. The prime difficult in this medicinal plant treatment is the identification of those plants. Without any expert, the identification is difficult. The image processing methodologies are the dominant method for solving this kind of problem. This paper is addressing a solution for the medicinal plant identification using deep learning networks. The deep learning algorithm is a class of machine learning algorithms that uses multiple layers to progressively extract higher level features from the raw input. From this approach, different kind of plants can be easily identified and the state-of-art of this approach is the speed of operation and precision in identification. The proposed approach is implemented on both dataset as well as experimental images

Keywords: Preprocessing technique, Feature extraction and Deep Belief Network

1. Introduction

Medicinal plants are an important source of new chemical compounds with potential therapeutic effects. The plant parts include leaf, stem bark, flower, fruit, root and rhizomes are used as a single drug or ingredient of formulation of Indian traditional medicines like Ayurveda, Siddha etc., since ages. Many plants synthesize substances that are useful to the maintenance of health in humans and animals. In present scenario, the World Health Organization (WHO) encouraging the developed countries to use herbal medicines. Since, it would be the growing recognition that natural products have fewer or even no side effects; for others it would be their accessibility and affordability. Now, the use of herbs has increased greatly in Western countries, also in places such as India and China. In Europe, at least 2,000 MAP (Medicinal and Aromatic Plants) species are traded commercially of these 1,200 to 1,300 are being native to Europe (Open Course Ware). For the efficient usage of herbs, identification of species is an important part. Identification is simply the determination of the similarities or differences between two elements, i.e., two elements are the same or they are different. The comparison of an unknown plant with a named specimen and the determination that the two elements are the same also involves classification, i.e., when one correctly decides that an unknown belongs to the same group (species, genus, family., etc.) as a known specimen, the information stored in classification systems becomes available and applicable to the material at hand. Both processes identification and classification involves comparison and judgment and require a definition of criteria of similarities. Identification is, therefore, a basic process in classification with nomenclature playing an essential role in the retrieval of information and as a means of communication. According to Black welder (1967) "identification enables us to retrieve the appropriate facts from the system (classification) to be associated with some specimen at hand" and is "better described as the recovery side of taxonomy." In practice one commonly identifies a plant by direct comparison or the use of keys and arrives at a name. To classify the required leaf from other species an algorithm is developed and it functions, based on the optimization techniques and Fuzzy Relevance Vector Machine algorithm. This classifier is made to learn based on the training features and labels are assigned to the leaf images using machine learning.

2. Related Work

The author of [1] given databases that are applied on the Semantic Annotation Based Clustering (SABC) for image and Semantic Based Clustering (SBC) for webpage content. The main intention of the proposed work is to accurately retrieve both the images and web pages. In experiments, the performance of the proposed SABC technique is evaluated and analyzed in terms of computation time, precision and recall.

The work [2] introduced a novel plant species classifier based on the extraction of morphological features using a Multilayer Perceptron with Ada-boosting. This framework comprises pre-processing, feature extraction, feature selection, and classification for plant species identification and achieved more than 90% accuracy.

The authors of [3] combined the IoT technology with deep learning to build an IoT system for crop fine-grained disease identification. This system is automatically detecting crop diseases and send diagnostic results to farmers also they proposed multidimensional feature compensation residual neural network (MDFC-ResNet) model for fine-grained disease identification in the system.

In [4] presented a NP-hard problem where an approach to identify the plant species from the contour information from occluded leaf image. Classifying occluded plant leaves is even more challenging than full leaf matching because of large variations and complexity of leaf structures.

In [5], Bacterial foraging optimization based Radial Basis Function Neural Network (BRBFNN) was introduced for identification and classification of plant leaf diseases automatically. For assigning optimal weight to Radial Basis Function Neural Network (RBFNN) the Bacterialforaging optimization (BFO) was utilized that further increases the speed and accuracy of the network to identify and classify the regions infected of different diseases on the plant leaves.

The [6] describes an automatic method for segmenting 3D point cloud data of vegetation, acquired from commodity scanners, into its two main components: branches and leaves, by using geometric features computed directly on the point cloud.

The main content of [7] was to extract plant leaf features and identify plant species based on image analysis. The class label of the test set can be obtained by reconstructing the deep learning model with the smallest error set. The results show that this method has the shortest recognition time and the highest correct recognition rate

.In the article [8], the author proposed an overlapping-free individual leaf segmentation method for plant point clouds using the 3D filtering and facet region growing. In order to separate leaves with different overlapping situations, a new 3D joint filtering operator was developed, which integrates a Radius-based Outlier Filter (RBOF) and a Surface Boundary Filter (SBF) to help to separate occluded leaves.

Concentric circles based method to explore the surface of the leafs was introduced in [9]. This work also to counts the changes of color in binary images, then, the changes are analyzed to detect compound leaves. The method was produced maximum accuracy in leaf detection.

In [10], a new CNN-based method D-Leaf was proposed. The leaf images were pre-processed and the features were extracted by using three different Convolutional Neural Network (CNN) models namely pre-trained AlexNet, fine-tuned AlexNet and D-Leaf. These features were then classified by using different machine learning techniques and CNN based work provided good results.

3. Methodology

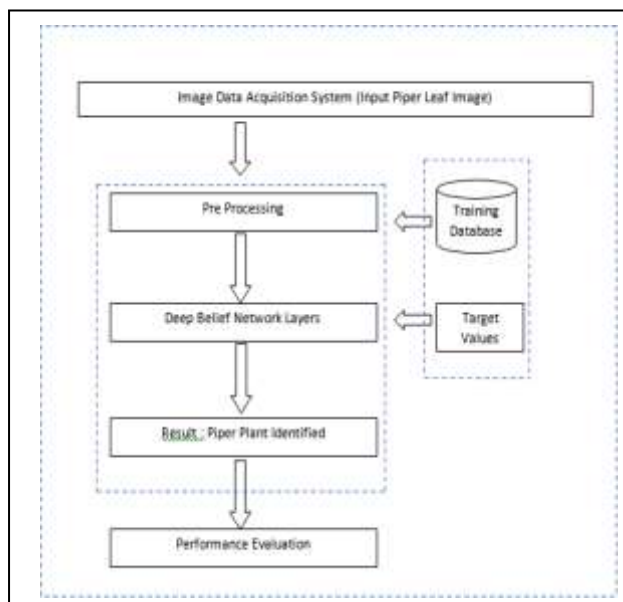


Figure.1 Proposed methodology

Plants have been used for medicinal purposes long before recorded history. It plays a major role in medicines, food, perfumes and cosmetics industries. By knowing the herbal plants and its usage it can be used for above applications. In this digital era, people don't have adequate knowledge to identify various herbal plants which are used by our ancestors for long time. Presently, the identification of herbal plants is purely based on the human perception or knowledge. If, human based manual identification and classification may occurs some errors at the time of identifying large numbers of herbal leaf species. To rectify this problem automatic recognition is proposed to bring the overall efficiency in identifying the species. This algorithm aims to predict the piper plants in a very

convenient and accurate way by using a computerized method like image and data processing techniques. The deep learning approach of piper leaves predictions is purely based on the leaf shape, texture, color and its features.

Preprocessing technique:

The raw image is available in the form of jpeg image format and the RGB mode of test image is having the noisy pixels. The preprocessing steps are eliminating the unwanted pixels. The pixels are inspected with the help of image processing tool available in MATLAB. The pixel is having three intensity values in RGB mode. The pixel representation is shown in the following figure

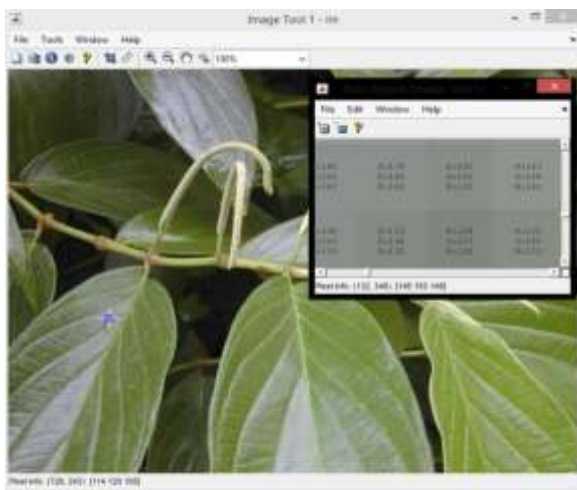


Figure.2 Pixel representation in test leaf image

The Histogram is the graphical representation of distribution of pixel in an image. This histogram plot gives the intensity distribution in R, B and G channel image. By analyzing this plot, it is possible equalize the intensity values. The following images are showing the histogram plots of R, G and B channels respectively.

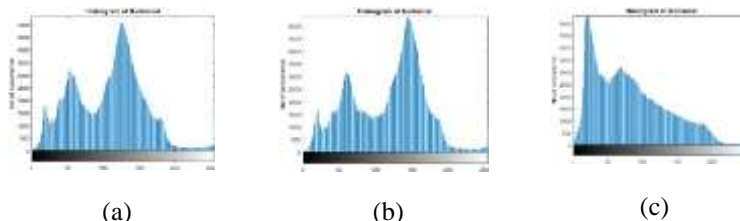


Figure.3 Histogram image –(a) Red channel (b) Green channel (c) Green channel

The histogram equalization will result into enhancing the intensity level and by the way the noisy pixels can be eliminated. The following figure shows the histogram equalized image of three color channel.

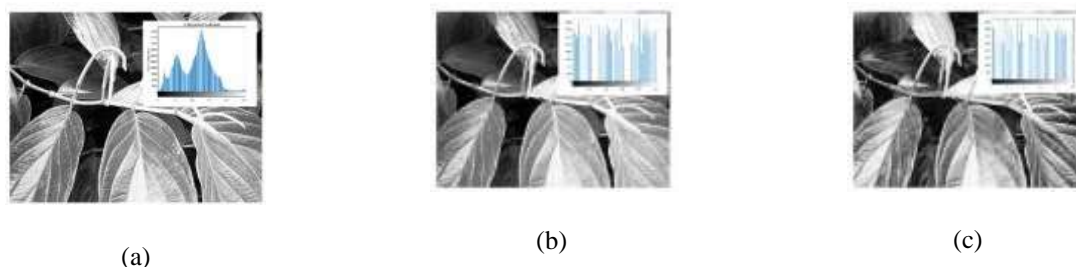


Figure.4 Histogram equalized image – (a) Red channel (b) Green channel (c) Green channel

Feature extraction:

It is obvious that the most distinguishing feature of a plant’s leaf is its shape. The shape of a leaf is invariant to plant maturity unlike the shape, texture and colour features, which might vary with maturity, climate or location. A set of features that best describe the shape of a leaf is the matter of interest in this research. The following features are used for classification of leaves: Area, perimeter, centroid, eccentricity, equivalent diameter,

extentmajor axis, minor axis, energy, correlation, entropy, inverse difference, sum of average, mean and standard of hue, mean and standard of saturation, mean and standard of value.

(i) Shape-based feature extraction

The shape features used in the proposed system are the area, centroid, eccentricity, equivalent diameter, extent, major axis length, minor axis length and perimeter. The feature (Value) are computed based on shape feature extraction by applying the shape based features for a single leaf by measuring the image properties of leaf using the Region Props Command.

(ii) Color based feature extraction

Various studies used the color histograms and color co-occurrence matrices for assessing the similarity between two leaf images. The color spaces are used for producing the color histograms. Each color is represented as a mixture of the three primary color channels (Red, Green and Blue) in the RGB color space. The shortcoming of this scheme is the sensitivity to illumination changes. The RGB color space is converted into the HSV (Hue, Saturation and Intensity) color model, which separates the intensity from the chromaticity.

Deep Belief Network

In machine learning, a deep belief network (DBN) is a generative graphical model, or alternatively a class of deep neural network, composed of multiple layers of latent variables, with connections between the layers but not between units within each layer.

When trained on a set of examples without supervision, a DBN can learn to probabilistically reconstruct its inputs. The layers then act as feature detectors. After this learning step, a DBN can be further trained with supervision to perform classification.

DBNs can be viewed as a composition of simple, unsupervised networks such as restricted Boltzmann machines (RBMs) or autoencoders, where each sub-network's hidden layer serves as the visible layer for the next. An RBM is an undirected, generative energy-based model with a "visible" input layer and a hidden layer and connections between but not within layers. This composition leads to a fast, layer-by-layer unsupervised training procedure, where contrastive divergence is applied to each sub-network in turn, starting from the "lowest" pair of layers.

The observation that DBNs can be trained greedily, one layer at a time, led to one of the first effective deep learning algorithms. Overall, there are many attractive implementations and uses of DBNs in real-life applications and scenarios.

Training of DBN

The training method for DBN proposed by Geoffrey Hinton for use with training "Product of Expert" models is called contrastive divergence (CD). CD provides an approximation to the maximum likelihood method that would ideally be applied for learning the weights. In training a single RBM, weight updates are performed with gradient descent via the following equation:

$$\omega_{ij}(t + 1) = \omega_{ij}(t) + \frac{\partial \log(p(v))}{\partial \omega_{ij}}$$

Where, $p(v)$ is the probability of a visible vector, which is given by $p(v) = \frac{1}{Z} \sum_h e^{-E(v,h)}$. Z is the partition function and $E(v, h)$ is the energy function assigned to the state of the network. A lower energy indicates the network is in a more "desirable" configuration. The gradient $\frac{\partial \log(p(v))}{\partial \omega_{ij}}$ has the simple form $\langle v_i h_j \rangle_{model} - \langle v_i h_j \rangle_{data}$ where $\langle \dots \rangle_p$ represent averages with respect to distribution p . The issue arises in sampling $\langle v_i h_j \rangle_{model}$ because this requires extended alternating Gibbs sampling. CD replaces this step by running alternating Gibbs sampling for n steps. After n steps, the data are sampled and that sample is used in place of $\langle v_i h_j \rangle_{model}$.

The procedure of work as follows:

1. Initialize the visible units to a training vector.
2. Update the hidden units in parallel given the visible units: $p(h_j = 1|V) = \sigma(b_j + \sum_i v_i \omega_{ij})$. σ is the sigmoid function and b_j is the bias of h_j .

3. Update the visible units in parallel given the hidden units: b_j is the bias of v_j . This is called the "reconstruction" step.
4. Re-update the hidden units in parallel given the reconstructed visible units using the same equation as in step 2.
5. Perform the weight update: $\Delta\omega_{ij} \propto \langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{reconstruction}$.

Once an RBM is trained, another RBM is "stacked" atop it, taking its input from the final trained layer. The new visible layer is initialized to a training vector, and values for the units in the already-trained layers are assigned using the current weights and biases. The new RBM is then trained with the procedure above. This whole process is repeated until the desired stopping criterion is met. Although the approximation of CD to maximum likelihood is crude (does not follow the gradient of any function), it is empirically effective.

4. Result and Discussion

The proposed system is implemented on the globally available dataset 'keralaplants' and where different Piper images are collected. The various types of piper leaves are taken in various classes that includes Piper argyrophyllum, Piper barberi, Piper betle, Piper chaba, Piper colubrinum, Piper galeatum, Piper hapium, Piper hymenophyllum, Piper longum, Piper mullesua, Piper nigrum var. hirtellosum, Piper pseudonigrum, Piper schmidtii, Piper silentvalleyensis, Piper trichostachyon, Piper trioicum and Piper wightii. The test images were taken with the help of digital single-lens reflex camera (DSLR) with the size of 1366 x768 Pixel. The images were stored in jpeg format. The following figure is showing the test image in RGB mode.



Figure.5 Test image in RGB mode

The image consists of undesired noises in pixel level and this can be eliminated by means of median filters. The result of this preprocessing steps is shown in the following figure.



Figure.6 Filtered image in grayscale

The histogram equalization is done on the filtered leaf image and the response of equalization process is shown in the following figure.



Figure.7 Histogram equalized image in grayscale

The region of interest for this process is green part of the leaf. This part of the leaf is separated and the color pixel values are undergoing for the feature extraction process.

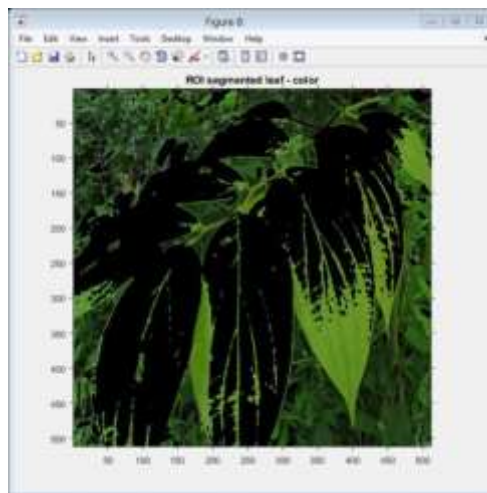


Figure.8 ROI Segmented Leaf- Color

The pixel distribution of both gray scale and binary ROI are shown in the following figure. The grayscale ROI is having pixel value from 0-255 at the leaf part and other parts are having '0' pixels. Similarly, binary ROI is having pixel value as '1' in leaf parts and other parts are having '0' pixels.

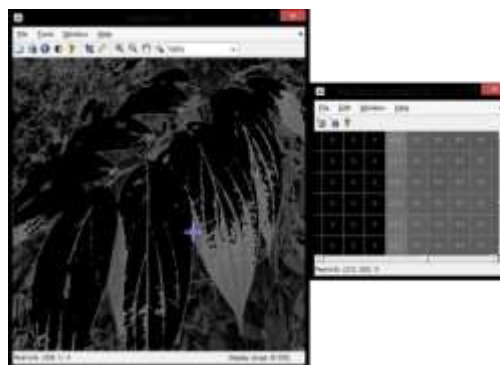


Figure.9 Gray scale ROI pixel distribution



Figure.10 Binary ROI pixel distribution

The hue-saturation-value is another type of color model where the color model is represented with the cylindrical coordinates. This work includes HSV modelled image and the HSV transformation is taken for ROI. The following figure shows the HSV image of extracted ROI.

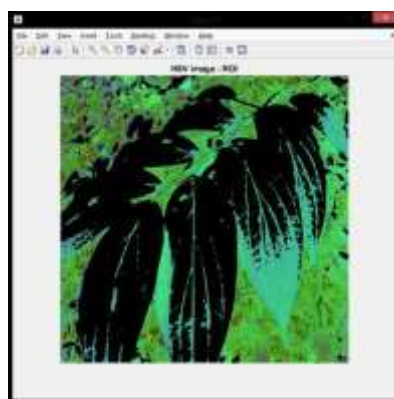


Figure.11 HSV image of extracted ROI

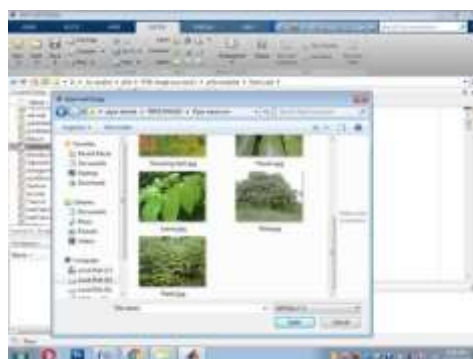


Figure.12 (Input) Species Image

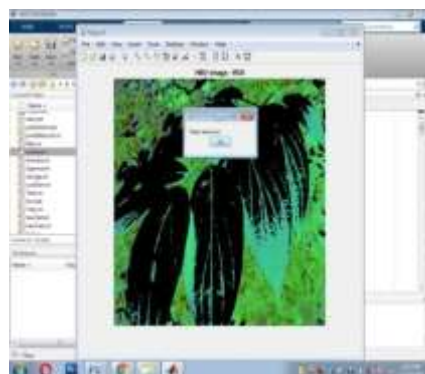


Figure.13 (Output) Species Identified

The color and shape features are extracted from the grayscale and binary ROI images and the extracted values are displayed on the MATLAB command window. The following figure is showing the feature values shown in the command window.

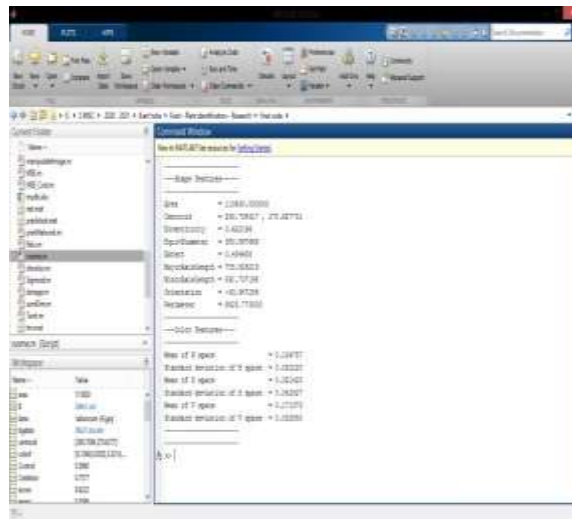


Figure.14 Feature values shown in the command window.

The six different piper leaf images are taken for the evaluation process. In each category 25 samples are taken. The proposed algorithm is providing the maximum accuracy and this can be evaluated with the help of confusion matrix. The following table is showing this confusion matrix for the proposed work.

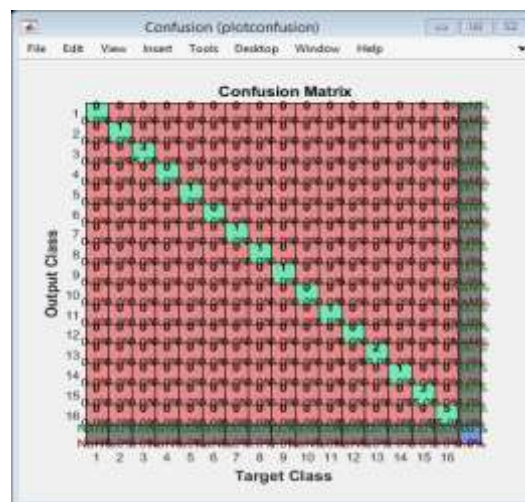


Figure.15 Confusion matrix of input image

The accuracy for each species are tabulated in the following table and the maximum accuracy is achieved as 100%.

Piper leaf image	Accuracy (%)
Aduncum	84
Amalago	88
Angustifolium	100
Arboreum	96

Auritum	96
Betle	100

Table 1: Maximum Accuracy of each piper Species

The deep learning networks are acute in predicting or classifying or clustering process. So this piper leaf identification process also producing the maximum efficiency. The performance curve of the proposed deep learning network is shown in figure.

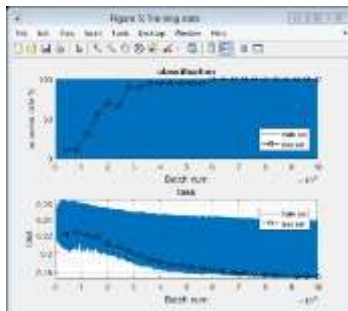


Figure.16 performance curve of the proposed deep learning network

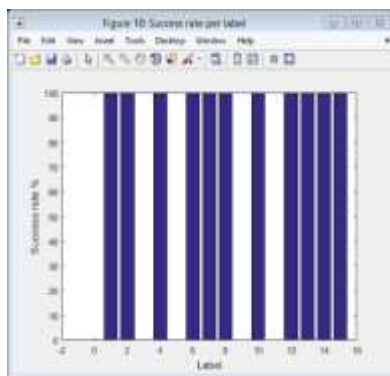


Figure.17 Success rate per label

5. Conclusion

Medicinal plants are widely used in non-industrialized societies, mainly because they are readily available and cheaper than modern medicines. Medicinal Plants contains hundreds of chemical compounds for the protection against insects, fungi, diseases, and herbivorous mammals. These herbs that have medicinal quality provide rational means for the treatment of many internal diseases, which are otherwise considered difficult to cure. This is the reason why medicinal plant related analysis is growing in popularity across the researchers. The prime difficult in this medicinal plant treatment is the identification of those plants. Without any expert, the identification is difficult. The image processing methodologies are the dominant method for solving this kind of problem. The secondary issue on this platform is regarding choosing the effective algorithms. This paper is addressing a solution for the medicinal plant identification using deep learning networks. The deep learning algorithm is a class of machine learning algorithms that uses multiple layers to progressively extract higher level features from the raw input. From this approach, different kind of plants can be easily identified and the state-of-art of this approach is the speed of operation and precision in identification. The proposed approach is implemented on both dataset as well as experimental images

References

1. C. Deepa. "SABC-SBC: a Hybrid Ontology Based Imageand Webpage Retrieval for Datasets" 108ISSN 0146-4116, Automatic Control and Computer Sciences, 2017, Vol. 51, No. 2, pp. 108–113. © Allerton Press, Inc., 2017.
2. Kumar, Munish, et al. "Plant Species Recognition Using Morphological Features and Adaptive Boosting Methodology." IEEE Access 7 (2019): 163912-163918.

3. Hu, Wei-Jian, et al. "MDFC–ResNet: An Agricultural IoT System to Accurately Recognize Crop Diseases." *IEEE Access* 8 (2020): 115287-115298.
4. Chaudhury, Ayan, and John L. Barron. "Plant Species Identification from Occluded Leaf Images." *IEEE/ACM transactions on computational biology and bioinformatics* (2018).
5. Chouhan, Siddharth Singh, et al. "Bacterial foraging optimization based radial basis function neural network (BRBFNN) for identification and classification of plant leaf diseases: An automatic approach towards plant pathology." *IEEE Access* 6 (2018): 8852-8863.
6. Digumarti, SundaraTejaswi, et al. "Automatic segmentation of tree structure from point cloud data." *IEEE Robotics and Automation Letters* 3.4 (2018): 3043-3050.
7. Huixian, Jiang. "The Analysis of Plants Image Recognition Based on Deep Learning and Artificial Neural Network." *IEEE Access* 8 (2020): 68828-68841.
8. Li, Dawei, et al. "An overlapping-free leaf segmentation method for plant point clouds." *IEEE Access* 7 (2019): 129054-129070.
9. Chau, Asdrubal Lopez, et al. "Detection of compound leaves for plant identification." *IEEE Latin America Transactions* 15.11 (2017): 2185-2190.
10. Tan, Jing Wei, et al. "Deep learning for plant species classification using leaf vein morphometric." *IEEE/ACM transactions on computational biology and bioinformatics* (2018).