A Novel Approach to Classification of Ayurvedic Medicinal Plants using Neural Networks

Sameer A Kyalkond
Student,
Dept. of CSE,
JSSATE, Bangalore

Sudhanva S Aithal
Student,
Dept. of CSE,
JSSATE, Bangalore

V Manikanta Sanjay
Student,
Dept. of CSE,
JSSATE, Bangalore

Punit S Kumar
Student,
Dept. of CSE,
JSSATE, Bangalore

Abstract—**Creating an automated categorization system for medicinal plants is a time-consuming and challenging task. India is a nation with a varied variety of plant species, each with its unique set of therapeutic qualities. Because it is hard for humans to recall the names of all plant species and their applications, previous knowledge is essential for manual identification and categorization. The preservation of these 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 in the classification of medicinal plants by exploiting textural aspects that are 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.**

Keywords—Deep Learning; Convolutional Neural Network; Classification; Medicinal plant; Leaf features; Ayurveda;

I. INTRODUCTION

Apart from providing oxygen and water to the planet's inhabitants, plants contribute significantly to the protection of life and biodiversity. Medicinal plants are a kind of plant that is used to treat a number of different disorders. It is important to maintain the knowledge of medicinal plants that has been passed down through generations. Computer vision, pattern recognition, and image processing techniques all have promise for identifying and classifying medicinal plants. Finding a medicinal plant that has all of the necessary therapeutic properties is one of the most difficult undertakings. While herbal treatment has no known adverse effects, a patient's life may be lost if the incorrect plant is employed to heal him or her. A fully automated approach is necessary at this stage to accurately identify the therapeutic plants. Medicinal plant identification and categorization permits the application of Ayurvedic treatments. Agronomists, botanists, ayurveda practitioners, forest department officials, and those engaged in the manufacture of ayurvedic pharmaceuticals all benefit from appropriate categorization of medicinal plants. Competent taxonomists, on the other hand, are in short supply in this field. People are increasingly preferring ayurvedic remedies above all others. [1] Taxonomists classify plants according on the characteristics of their leaves and flowers, as well as their stems and branches. The year-round presence of leaves is the strongest indicator of plant categorization.

This gap may be addressed by computer vision and image processing algorithms that are capable of detecting and classifying medicinal plants without the assistance of a huge number of professionals. Plants are classified according to their morphological and spatial properties, which include shape, color, and texture. On the other hand, color is not a good categorization criterion since it varies with the seasons and may appear at several stages of development on the same leaf. Taxonomists classify medicinal plants based on their leaf features. Numerous research on this subject have been undertaken. The similarity in form, color, and texture amongst classes makes this a challenging challenge to address. The remainder of the paper is summarized below. Section II summarizes and discusses some significant projects in the taxonomy of plant species. The third part outlines the planned study's instruments, methodology, and approach. Section IV summarizes the findings, and Section V closes the research. Additionally, this page has some future suggestions.

II. LITERATURE SURVEY

Jing Wei Tan and colleagues [1] proposed the use of a CNN model called DLeaf to classify leaves in plants. DLeaf extracted features using a CNN and classified them using an ANN in that order. Segmenting the venation from scaled leaf pictures is accomplished using the Sobel edge detection technique. The D-Leaf model was reported to have a classification accuracy of up to 94.88 percent.

Gopal A et al [2] suggested an automated method for identifying the leaves of medicinal plants. We distinguish distinct kinds of leaves using boundary-based features, moment characteristics, and color characteristics. After training and testing with 100 and 50 leaves, respectively, the classification effectiveness was 92 percent. R. Janani et al. [3] created an artificial neural network-based classification system for medicinal plant species based on leaf color, texture, and shape. Each leaf was subdivided into 36 training, seven validation, and twenty test leaves. The model included a total of 63 leaves. The classification of leaves was based on the identification of eight hardly detectable qualities among the leaves' twenty distinct characteristics. The eight characteristics are compactness, eccentricity, skewness,

kurtosis, energy, correlation, sum-variance, and entropy. This approach has a 94.4 percent success rate.

D.Venkatarama et al [4] suggested a method for classifying medicinal plant leaves and extracting their therapeutic properties using computer vision. The leaf class is identified using a Probabilistic Neural Network classifier. The procedures include pre-processing, feature extraction, categorization, and retrieval of therapeutic properties. Calculating and comparing feature vectors to other datasets is a critical part of categorization.

Shitala Prasad et al [5] demonstrated an effective method for collecting leaves as well as methods for transforming the acquired picture to the device-independent l color space. Principle Component Analysis (PCA) is used to reduce and optimize this feature map (PCA). The dataset was ImageNet. When the model is sent through the Fully Connected layer, a 3x4096-pixel feature vector is generated. SVM was applied to this feature set using PCA techniques, and the accuracy was found to be 97.6 percent for lVGG-16 and 98.2 percent for l-VGG-16.

Amala Sabu et al [6] identified medicinal plants in the Western Ghats using machine learning and computer vision. Classification of the SURF and HoG features was performed using the kNN algorithm. Leaf veins and twenty various places of interest on the leaves are also represented using an SURF feature descriptor. This methodology use the k-NN classification algorithm to categorize leaves with a k value of 1. Despite its high accuracy of over 96 percent, the approach for feature extraction utilized in this model is computationally intensive.

Manojkumar P et al [7] determined the optimal combination of characteristics necessary to accurately classify medicinal leaves. - Classification of objects is based on geometric qualities, color features, texture features, HU invariant moments, and Zernike moments. Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM) classifiers are utilized, with the MLP classifier obtaining a maximum accuracy of 99 percent for Geometric-Colour-Texture classification. A 38-property Zernike combination.

A. Salima et al [8] pioneered the use of a Hessian matrix to segment leaf veins. Hessian matrix segmentation is followed by thinning and visual assessment in this procedure. The testing and assessment technique utilizes a total of 80 leaves. The segmentation result was evaluated using segmentation score approaches in this process. 53.75 percent of leaf images were awarded a segmentation value of 2, whereas 42.5 percent were assigned a score of 1.

Guillermo et al [9] suggested classifying plants based on their leaf venation using Convolutional Neural Networks (CNNs). For vein segmentation, the unconstrained variant of the Hit or Miss Transform is utilized. The attributes of interest were extracted from a 100x100 pixel centered region inside the segmented vein pattern. This technique makes use of a range of classification algorithms and methodologies, including SVM, PDA, and RF. SVM, PDA, and RF are three examples.

III. OBJECTIVES

1.Plant identification using a user-uploaded photograph would be one major goal of the proposed system.
2.To create a graphical user interface that is easy to use.
To ensure that the user interface is simple to grasp and intuitive to use.
3.To provide instructions on how to care for the plant.
4.To deliver the most accurate results possible.
5.To give farmers with a cost-effective and time-saving option.
6.For the sake of providing a dependable and efficient system.

IV. METHODOLOGY

A. Deep Learning:

Plant identification and classification have been carried out utilizing standard image processing and classification approaches for many decades. These approaches classified objects based on their shape, texture, and color.

These characteristics include aspect ratio, eccentricity, kurtosis, skewness, energy, correlation, cumulative variance, entropy, and compactness. The significant disadvantage of these traditional approaches is the lengthy calculation time necessary for manual feature extraction. All conventional approaches have been supplanted by machine learning techniques in the modern era.

Deep learning is a subset of machine learning in which a computer model learns to classify pictures directly from them. Increased accuracy, the capacity to process large volumes of picture data, the built-in ability to utilize GPUs for parallel processing, and the availability of pre-trained Convolutional Neural Networks all contribute to deep learning's appeal. Given that the system must classify a large amount of visual data, the deep learning technique is the optimal solution.

B. Dataset:

The proposed dataset consists of 1050 medicinal and 550 non- medicinal leaves commonly found across Karnataka. More than 20 leaves from 100 different plant species were collected and performed sampling on it. The leaves with severe deformities are removed and selected 20 leaves with a significant difference in the shape, color, and size. Apparently, 20 leaves of each class are selected for the further scanning process. Both the top and bottom sides of these selected 20 leaves are scanned to create 40 leaf images per species.

Only the leaf area is selected and cropped using GIMP image editor and saved each image in jpg format. A common naming convention is used to label each image, plant species name followed by a unique sequence number. Sampling ensures the diverse nature of the dataset in plant species level and helps the model to provide more accurate classification results. The images are converted to a resolution of 256x256 and are trained with epoch.

C. Proposed Architecture:

The suggested system's CNN architecture is based on the AlexNet architecture, which was invented and built by Google.

The first layer is the input layer, which contains the dimension information for the input pictures. The second layer, called convolution, consisted of 90 (7x7) filters with a stride size of two. Following this layer is a ReLU layer that thresholds the output, followed by a max pooling layer with a 2x2 filter size. This layer precisely halves the size of the output. Following this max pooling layer is a second convolution layer that acts on 512 kernels of size 5x5 with stride of 2. Following that, a ReLU layer is added, followed by a max pooling layer with a filter size of 3x3 and a stride of 1.

The next two layers are two convolutional layers stacked one on top of the other in the following combinations. Both use 3x3 kernels with a stride of 2, with the first and second containing 480 and 512 kernels, respectively. Following these layers, a ReLU layer and a max pooling layer with a filter size of 2x2 and a stride of 2 are added. The output of the previous maximum pooling layer is sent to the first fully connected layer, which has 6144 neurons.

6144 neurons are also included in the second completely linked layer. And the third completely linked layer has 40 neurons, which is the number of classification classes for the medicinal plant. Finally, the third fully connected layer's output is sent to the softmax and ReLu classification layers, which generate classification probabilities for each species.

Four stages are taken by the suggested CNN model.

1) Capture of Images: Leaf samples of therapeutic plants were gathered from ayurvedic practitioners and Karnataka's tropical woods. At least twenty leaves were gathered from 100 distinct medicinal plant species. Conduct a simple hand sample and remove any leaves that are significantly damaged. For each medicinal plant species, 30 leaves are selected for scanning and 60 picture samples are obtained.

2) Image Preprocessing: The preprocessing phase's purpose is to convert the scanned images to a dimension of 256x256x3, one of the accepted input image formats for CNN. We utilized RGB photos and converted them to the requisite 256x256x3 format if they were not already in it. Due to the fact that the dimensions of the photos in our dataset vary, we first padded them into a NxN dimension. Finally, the padded picture was enlarged to a resolution of 256x256x3.

3) Feature Extraction: A variety of CNN models were created using a variety of different layer counts, filter counts, and filter sizes, as well as a variety of different training choices. The dataset is used to train and test all of the developed models, and the results are compared to those of the Alexnet CNN model. The performance of these models is obviously dependent on the quality of the training data, the number of convolutional, max-pooling, and ReLU layers used, the training choices used, and the number of neurons in the first two fully connected layers.

The model's accuracy is heavily dependent on the number of training photos and the number of iterations used to train the CNN. Following the training phase, the model is calibrated using photos from the test set. Finally, the model outputs the training and validation accuracy using an accuracy-loss graph

during training, as well as classification accuracy using a confusion matrix during classification.

V. RESULTS AND INTERPRETATIONS

For our case study, we have compared different ratios in which the dataset can be divided into test and train datasets. The results obtained are given below.

A. Result Obtained For Ratio of 60:40:

• Epoch - 90
• Accuracy - 89.74%
• Loss - 44.85%
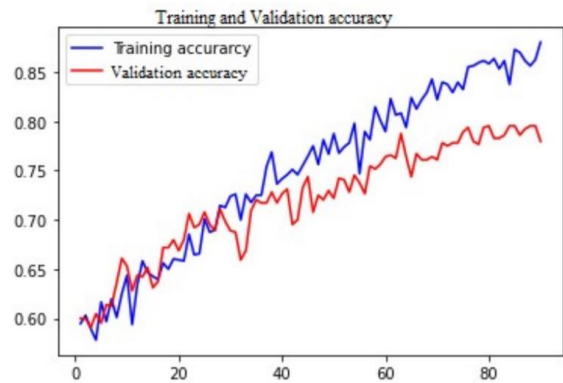• Validation Accuracy - 77.97%
• Training and testing accuracy graph -

Fig. 1. Accuracy graph for 60:40
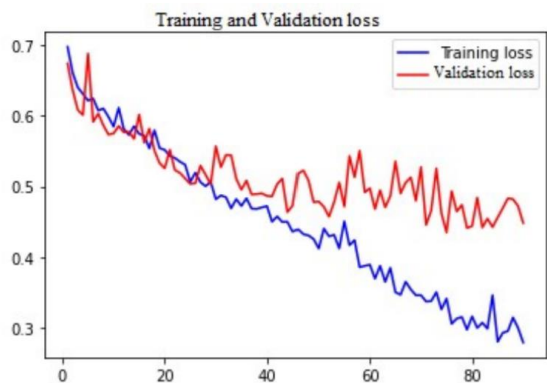
• Training and testing loss graph -

Fig. 2. Loss graph for 60:40

B. Result Obtained For Ratio of 70:30:

• Epoch – 90

• Accuracy - 83.05%

• Loss - 50.84 %

• Validation Accuracy - 77.08%

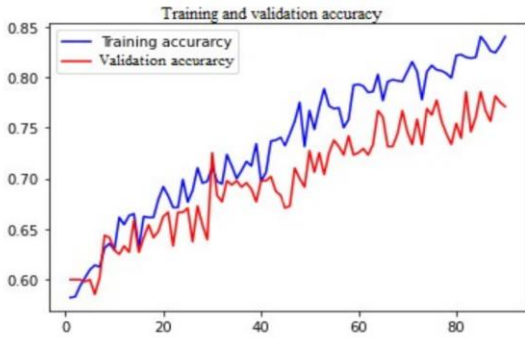• Training and testing accuracy graph –

Fig. 3. Accuracy graph for 70:30
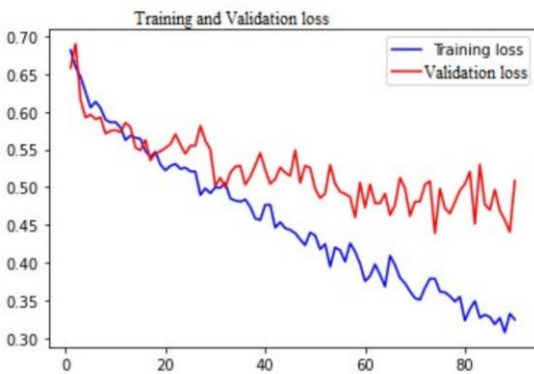
• Training and testing loss graph –

Fig. 4. Loss graph for 70:30

C. Result Obtained For Ratio of 80:20:

• Epochs - 90
• Accuracy - 90.01%
• Loss - 42.04%
• Validation Accuracy - 81.88%
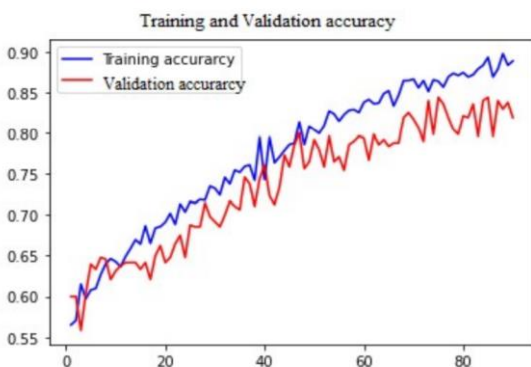• Training and testing accuracy graph –

Fig. 5. Accuracy graph for 80:20
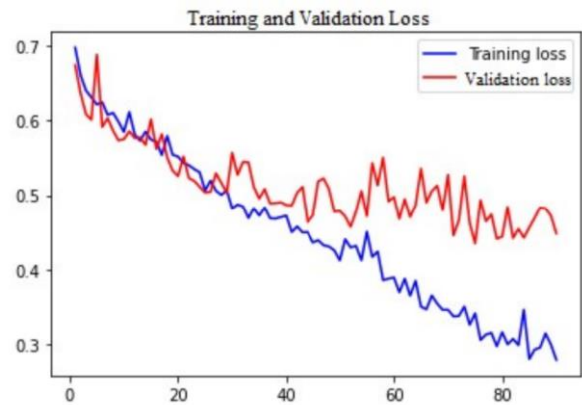
• Training and testing loss graph –

Fig. 6. Loss graph for 80:20

D. Application:

When the code is executed on VS Code, prediction is obtained as follows. The given example is of a leaf which is non-medicinal.
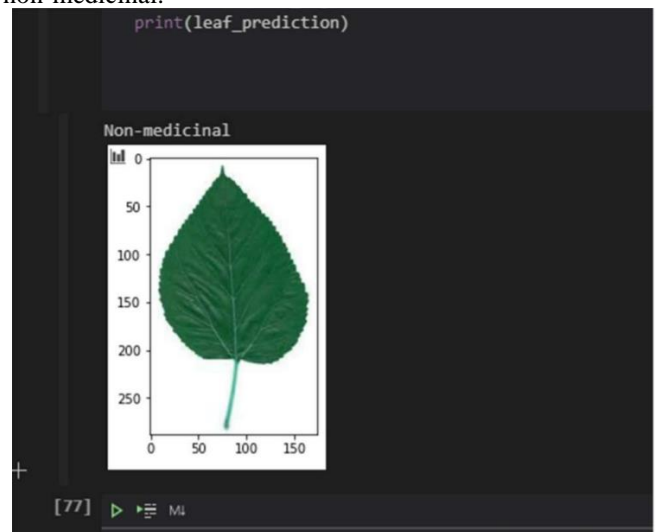

Fig. 7. Backend Prediction

When the file output.py is executed. This is the window is displayed.
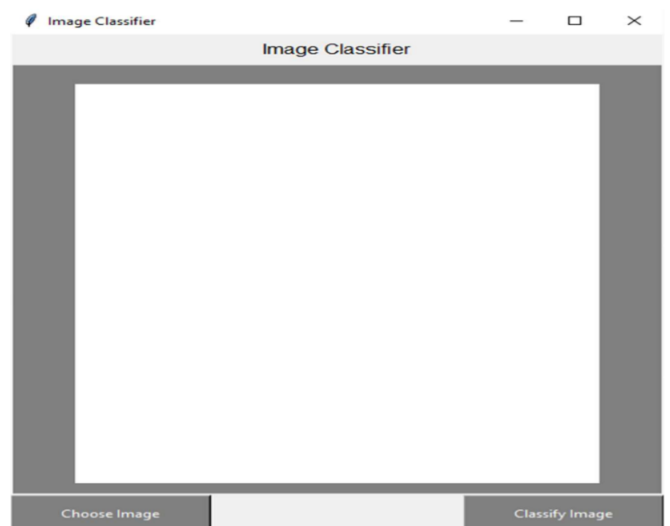

Fig. 8. GUI

The user clicks on the 'Choose Image' button. He is given the option of choosing a photo from the local system.

Fig. 9.   Choosing a photo

After choosing the photo, when the user clicks on the 'Classify Image' button, the leaf is predicted by the trained model and respective class name is printed. The following result obtained.

Fig. 10. Predicting Leaf

If the leaf in the uploaded picture is Medicinal, this is the result obtained.
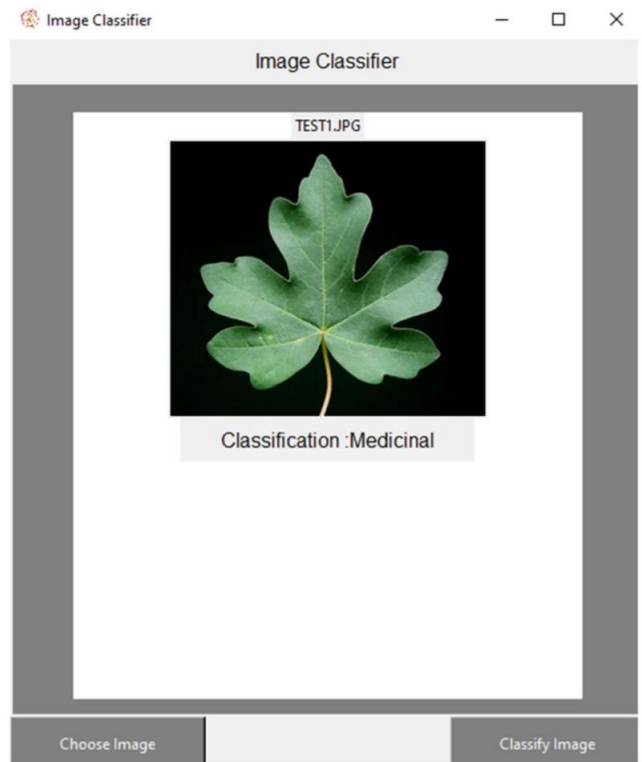

Fig. 11. Predicting Medicinal

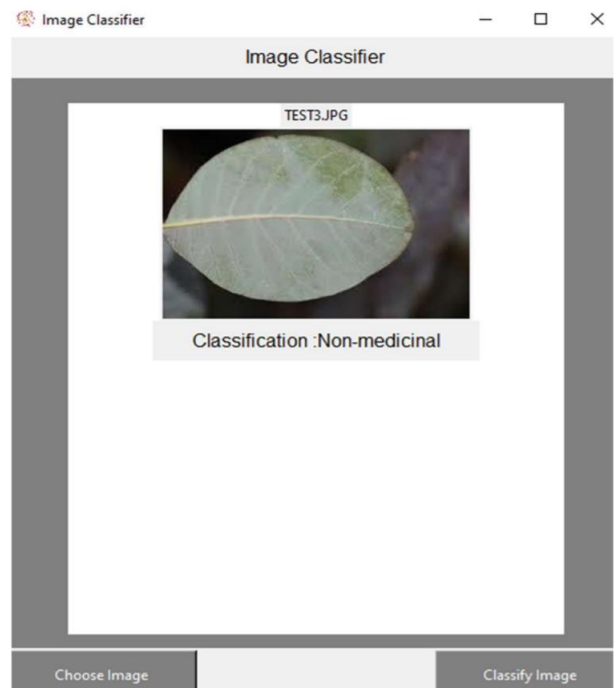If the leaf in the picture uploaded is Non-medicinal, this is the result obtained.

Fig. 12. Predicting Non-Medicinal

V. CONCLUSION AND FUTURE WORK

This research highlights the enormous potential of picture classification algorithms and how they can be utilized to outperform humans in a range of classification tasks. With advances being made daily in the domains of picture

classification and computer vision, this suggested method, if implemented on a wider scale, has the potential to dramatically alter the way traditional medicine such as Ayurveda, Unani, and others is performed in the nation.

One of the primary disadvantages of traditional medicine that is impeding its popularity and expansion is a lack of information about the requisite plants/plant extracts among the urban people. This initiative may assist in overcoming this disadvantage by offering an effective and user-friendly interface for identifying and using these plants.

Even though the suggested system is fully functional in its current form, it cannot be used as a real-time application without a few critical enhancements: -

• To begin, a beautiful and simple-to-use user interface must be built so that the system can be used without difficulty and provide the intended outcomes.

• Additionally, if networking capabilities are included, the user will be able to access enormous web resources to learn more about the plant specimen in question, rather than relying on pre-loaded data.

• Moreover, the CNN method may be improved by hyper-parameter tweaking, data redesign, and model optimization.

REFERENCES

[1] Manojkumar P., Surya C. M., and Varun 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), pp 978- 1-5386-1931-5.

[2] Mr. K.Nithiyanandhan and Prof.T.Bhaskara Reddy, "Analysis of the Medicinal Leaves by using Image Processing Techniques and ANN", Vol 8, No. 5, ISSN No. 0976-5697, MayJune 2017.

[3] Adams Begue, VenithaKowlessur, FawziMahomoodally, Upasana Singh and Sameerchand, "Automatic Recognition of Medicinal Plants using Machine Learning Techniques", International Journal of Advanced Computer Science and Applications, Vol. 8, No. 4, 2017.

[4] H. X. Kan, L. Jin, and F. L. Zhou," Classification of Medicinal Plant Leaf Image Based on Multi-Feature Extraction", Pattern Recognition and Image Analysis, Vol. 27, No. 3, 2017, pp. 581–587, 1054-6618. © Pleiades Publishing, Ltd.

[5] Riddhi H. Shaparia, Dr. Narendra M. Patel and Prof. Zankhana H. Shah," Flower Classification using Texture and Color Features", International Conference on Research and Innovations in Science, Engineering &Technology, Volume 2, 2017, Pages 113–118.

[6] Marco Seeland, Michael Rzanny, NedalAlaqraa, Jana Wäldchen, Patrick Mäder, "Plant species classification using flower images—A comparative study of local feature representations", PLOS ONE | DOI:10.1371/journal.pone.0170629 February 24, 2017.

[7] Pradeepkumar Choudhary, Rahul Khandekar, AakashBorkar, and PunitChotaliya, "Image processing algorithm for fruit identification", International Research Journal of Engineering and Technology (IRJET), Vol 4 Issue 3, e-ISSN: 2395 -0056, p-ISSN: 2395- 0072, Mar -2017. [8] D Venkataraman and Mangayarkarasi N, "Computer Vision Based Feature Extraction of Leaves for Identification of Medicinal Values of Plants", IEEE International Conference on Computational Intelligence and Computing Research, 978-1-5090-0612-0/16/$31.00 ©2016 IEE.