iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** IV **Month of publication:** April 2019

DOI: https://doi.org/10.22214/ijraset.2019.4350

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

Plant Species Classification and Disease Detection using Convolutional Neural Network

Himanshu Joshi[1], Reeya Saxena[2], Snehal Bhise[3], Diksha Sawant[4], Prof.Anand Bhosale[5]
[1, 2, 3, 4, 5]*Information Technology, Savitribai Phule Pune University*

Abstract: This paper presents a survey on detection and classification of leaf spices. It is di□cult for human eyes to identify the exact type of leaf spices. Thus, in order to identify the leaf spices accurately, the use of image processing and machine learning techniques can be helpful. The images used for this work were acquired from the digital camera. In pre-processing step, background removal technique is applied on the image in order to remove background from the image. Then, the background removed images are further processed for image segmentation. Di□erent segmented images will be used for extracting the features such as color, shape and texture from the images. At last, these extracted features will be used as inputs of classifier. The goal of proposed work is to diagnose the spices using image processing of plant leaf. In the proposed system, leaf image with complex background is taken as input. Threshold is deployed to mask green pixels and image is processed to remove noise using an isotropic di□usion. Then leaf spices segmentation and then classification using classifier will be done.
Keywords: 1. Preprocessing,2. Feature Extraction, 3. Classifier, 4. Threshold,5. Isotrofic-Difiusion.

I. INTRODUCTION

Plants are an integral part of ecosystem. Due to deforesting many plant species are under the risk of annihilation. Plants are useful for human being and other living things. They are useful as foodstu□, as medicine and also in many industries. Identifying plants helps ensure the protection and survival of all-natural life. Plant identification can be performed using many di□erent techniques using the planta's leaves. Leaves are useful to classify plants since they are more readily available than the other bio-metric components like flowers which are available for a short period. Plant classification by using leaves requires different bio-metric features of leaf like colour, shape, texture. This identification manually is time consuming and expensive. Leaves can be classified based on colour that include similarity between two images with the help of colour histogram, but the colour based classification is depend on seasonal e□ect of sunlight. So, by using classifier we can classify the Species of leaf.

A. Problem Statement

One of the important sectors of Indian Economy is Agriculture. Employment to almost 50 percent of the countries workforce is provided by Indian agriculture sector. India is known to be the world's largest producer of pulses, rice, wheat, spices and spice products. Farmers' economic growth depends on the quality of the products that they produce, which relies on the planta's growth and the yield they get. Therefore, in field of agriculture, detection of spices in plants plays an instrumental role. In order to detect a plant spices at very initial stage, use of automatic spices detection technique is advantageous. Manual detection of plant spices using leaf images is a tedious job. Hence, it is required to develop computational methods which will make the process of leaf species detection and classification using leaf images automatic.

B. Background Study

Below figure shows the fundamental steps of an automated plant classification system, that is nothing but a system architecture. Initially, the leaf images would be acquired using digital camera, scanner or some other equipment's. The images were then pre-processed to remove noise and improve the quality. Noise occurs as pixel values which do not represent the true intensities of an image during the image acquisition. Image enhancement is a process that is used to emphasise the features of an image. It is a necessary step to remove the image noises in order to highlight or enhance the important features of an image. Subsequently, the region of interest (ROI) was segmented from the images, followed by feature extraction[1]. Finally, the extracted features are fed to classification or recognition system. Leaves are commonly used in plant species recognition due to their availability throughout the year, especially, in the tropical areas. [1,2,5] Shape is the most common feature that have been used to develop plant identification systems. [1,2,5] Texture is one of the important features of the plant identification system, which can be used to characterize the leaves based on the surface structure of the leaves. [1,2,5]

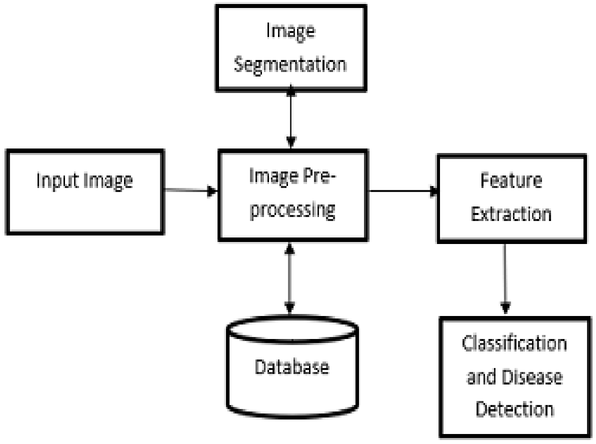


Fig -1: Proposed System Architecture [5]

II. PREPROCESSING

It is common practice to have the pre-processing of Cotton leaf images before it has been extracted and classified.

There are five main steps used for the detection of plant leaf diseases as shown in fig. The processing scheme consists of image acquisition through digital camera or web, image pre-processing includes image enhancement and image segmentation where the affected and useful area are segmented, feature extraction and classification. Finally, the presence of diseases on the plant leaf will be identified. [8]

In the initial step, RGB images of leaf samples were picked up. The step-by-step procedure as shown below:

1) RGB image acquisition;
2) Pre-processing of image using Histogram equalization;
3) Resize the image;
4) CNN for Image Segmentation;
5) Computing features extraction;
6) Classification & Recognition using neural networks.
7) Statistical analysis.

A. Pre-Processing of Leaf Image

The input image has to be pre-processed because images are corrupted by a type of multiplicative noise like light intensity and shadow on a cotton leaf images that may contain useful information about the leaf spot that can be used in the diagnosis. The pre-processing is done with the contrast enhancement using Histogram equalization.

B. Leaf Segmentation

The leaf spot in the capture image generally contains reflection from source, which forms some intense spot in the cotton leaf, but pixel value within the cotton leaf is over a particular threshold (20) then it is replaced by pixel value of some neighbourhood pixel. This operation fills all intense leaf spot present in cotton leaf area as shown in the figure below.

Fig -2 a) Captured Image

Fig -2 b) Image After Segmentation.

C. Feature Extraction

Feature extraction is the process of transforming the raw pixel values from an image, to a more meaningful and useful information that can be used in other techniques, such as point matching or machine learning. In the proposed system we are using the Scale Variant Feature Transform (SIFT) algorithm. Features of the leaf like colour, texture, shape, vein. Numerous methods can be employed for feature extraction such as Histogram of Oriented Gradient (HOG), Zernike Moments, Hu's Moment and others.[1] In this research we used SIFT & Canny Edge Detection Algorithm (CEDA). Where CEDA is used for feature extraction of leaf.

The following information can be extracted from the Image.

1) *Convex hull Information:* It is formed using boundary point of leaf. Convex hull is approximated and number of vertices is extracted. The area of Hull and Perimeter is also Calculated.[7]
2) *Morphological Information:* The length and width of leaf is calculated by finding minimum and maximum x and y coordinates. The perimeter and area of leaf is also calculated using boundary points.[7]
3) *Distance Maps:*

a) *Vertical and Horizontal maps*: Lines are drawn on segmented image and each line consist of minimum and maximum image which intersect with leaf and its respected axis. The distance of each line is estimated and stored. This is above in Fig 3a.[7]

b) *Centroid radial map:* The centroid of leaf is found by intersecting the diagonal axes though the bounding box along the leaf. Sixteen point radiating from centroid are taken on boundary box. The Euclidian distance between this point and centroid are calculated as well as that between the centroid and the point which intersect the leaf boundary. This is shown in Fig 3.b[7]

The Information extracted are then used to create ratio which will be used in pattern matcher.

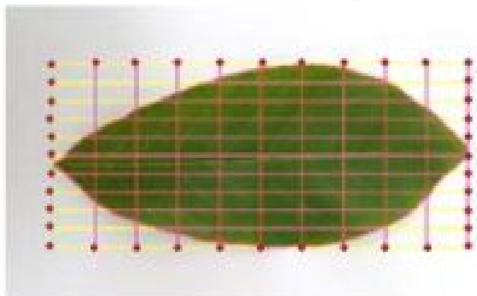


Fig -3 a) Distance Map X and Y[7]

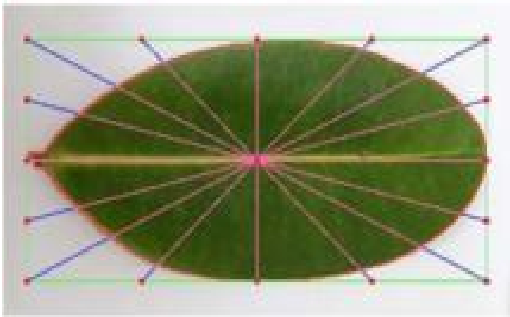


Fig -3 b) Centroid Radical Distance Map[7]

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

4) *Color Histogram:* It is a computed for a cropped part of image since if the whole is used, white spaces surrounding the leaf would affect the histogram. To Crop the image, the length and width of bounding box are used as markers to crop the central part of leaf image.[7]
5) *Matcher:* It consisted of 2 stages of CNN algorithm. All ratio is normalized to 0 or 1 before any comparison are made.

a) *Stage 1*: The leaf to be recognized undergoes the same process as the ones in database. The new leaf is then compared to training set one by one. The sum of all the Euclidian distance between the leaf and those in database are calculated. The 3 closest distance are returned. Each Ratio then feature in CNN classifier. For the distance and centroid map all the distance in set are considered as individual feature. [1,7]
b) *Stage II*: For instance, where the result set from stage 1 consist of different plants, color histogram of the new leaf compared those from result set. The correlation is calculated. Its value lies between -1 to 1. Value close to 1 means that very high correlation which means that the image is very similar. The closest match is calculated using CNN algorithm. [1,7]

More are the images greater is the efficiency of CNN algorithm.

Table below shows the ratio that have been calculated from previous measurement.[7]

Aspect Ratio	White area ratio	Perimeter to area	Perimeter to hull	Hull area ratio	Distance map x	Distance map y	Centroid radical distance
Width/Length	Area of Leaf/ (Length*width)	Perimeter of Leaf/ Area of Leaf	Perimeter of hull/ Perimeter of leaf	Area of leaf/Area of hull	Distance of lines parallel to x-axis/Length	Distance of the lines parallel to y-axis/Width	Distance from centroid to intersecting points/distance from centroid to boundary point

Table1: Ratio

III. CLASSIFICATION OF PLANT SPECIES AND DISEASE DETECTION

The processing scheme consists of test RGB image acquisition from database or web. Image pre-processing includes image enhancement and image segmentation where the affected and useful area are segmented each filter having size of 512 X 512 pixels. Here the size of feature vector is the size of image 512 X 512 pixels. Fig -4b shows that Enhance Test image using histogram equalization. Preprocessing the test image using histogram equalization is applied to increase the contrast in low contrast image where, leaf spot is highlight in Fig -4b. [2,4,5,7]

Fig -4 a) Test RGB Image

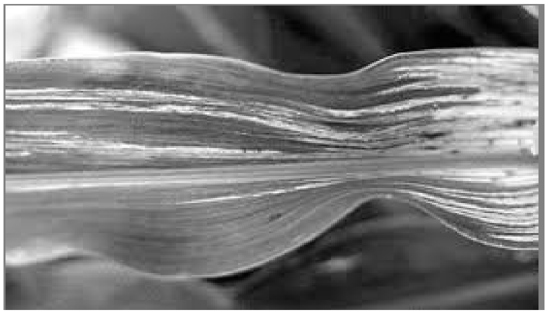


Fig -4 b) Histogram Equitization of Image

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

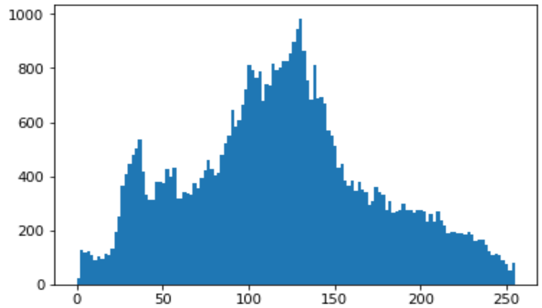


Fig -4 c) Histogram of Concentrated Pixels of image

A. Leaves Disease Detection Parameter

Recognition Accuracy Comparisons, Execution Time Comparisons, False Accept and False Reject Rates for D-Leaf datasets are compared below in respective table. Table -2 shows that recognition accuracy for detecting diseases on D-Leaf dataset. It shows that CNN algorithm has highest accuracy up to 97%.

Sr. No	Feature Extraction method/No of dataset	CNN Classifier Accuracy
1	100	85.6
2	200	86.9
3	300	87.9
4	400	90
5	500	91.3
6	600	93.6
7	700	96.8

Table 2: Recognition accuracy comparison

B. Execution Time

Table -3 shows that execution time in second for detecting diseases on D-Leaf dataset. Out of which CNN Classifier algorithm takes less execution time.

Sr. No	Feature Extraction method/No of dataset	CNN Classifier Execution or Compiling time in sec
1	100	45.6
2	200	60.6
3	300	70.5
4	400	80.6
5	500	100.4
6	600	120.6
7	700	144.5

Table 3: Execution Time

IV. RESULTS

The method of testing used was to use every photo of leaves in the database as an input image to the system, compare it to all other leaves and calculate the percentage accuracy of the system. This technique has an advantage of testing all images in database rather than just a small percentage of it. Every time the system applies the matcher to a leaf, it will create in a CSV file with the actual plant name and the predicted plant name.[5]

Testing was done on 740 leaves coming from different species of plants. In Particular we noted a 100% accuracy for seven different type of leaves. Only few plant species recognitions were below 85%.

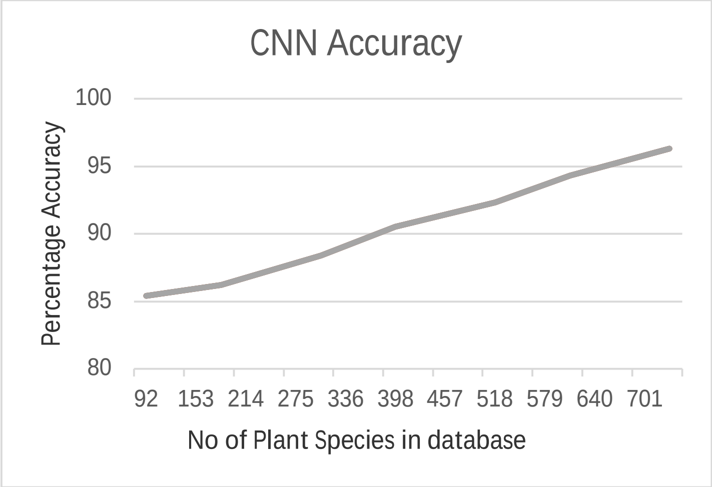


Fig. 5 Effect of increasing in number of plants species on classifier.

From above fig 5 it can be noted that with only eight species, we have a very low classifier accuracy of 85.4% which rapidly increases with the increase in number of species and touches to 96.3% in our system. We usually expect that the recognition accuracy to go down for more variety of dataset. However, in this case the accuracy is going up rapidly and it is still very less for 740 species. As the no of leaves of the plant increases the efficiency of classifier also increases in proposed algorithm. However, there is no increase in the accuracy of the classifier once the saturation limit is been reached. Overall accuracy follows similar trend but there is increase in almost (+-1%) overall accuracy using CNN classifier, after adding each additional leaf considering the starting amount of leaves. Thus, it is possible to obtain high number of accuracies by using relatively large number of species but with small no of leaves per species. We also demonstrated how accuracy varies with number of plant species and number of leaves. The accuracy obtained are comparable with existing work. However, as our approach is based on CNN classifier it is expected to run faster than PNN and KNN classifier. The effect of varying number of species and number of leaves had been successfully tackled in this literature. The Overall accuracy at the first phase was close to 90%. As the dataset increased the accuracy for the classifier also got increased significantly. The Color histogram matching operation was then applied on result from the first phase and accuracy rose. Moreover, the same test was carries on entire D leaf dataset the much higher accuracy close to 95% where obtained in first and second stages respectively. Thus, we can see this approach is very effective in classification of plant and leaves. Further experiment was conducted in order to assess the effect of increasing the number of plant species and the number of leaves on classification accuracy. We used CEDA and SIFT as we have greater amount of dataset and the other algorithm like SEDA and PCA are applicable only for short amount of dataset. We found that by using above algorithm our efficiency on out has significantly increased and does not vary time to time. Thus, we Successfully tested various species of plant in the database where we came to conclusion that as the number of species increases the efficiency of the Convolutional Neural network classifier also increases.

V. CONCLUSIONS

Plants play an important role in our lives, without plants there will not be the existence of the ecology of the earth. The large amount of leaf types now makes the human being in a front of some problems in the specification of the use of plants, the first need to know the use of a plant is the identification of the plant. The proposed system is an automated client-server system and is capable of classifying plants by extracting all morphological features of leaf from binary masks of leaf and predicting plant species using a SIFT feature detection algorithm and by canny edge detection algorithm giving the highest accuracy of 90 and 97.5 percent respectively. Hence saving the loss of biodiversity and reducing the dependency on the expert to a certain extent is possible. It can provide the help for a person having less knowledge about the plant species. So, we can conclude that these algorithms are the most suitable for this task. The proposed system is web-based system with user interface. In future, we can design an application for plant species classification with some improvement in technology support.

VI. DATASET

The leaf images dataset used in this research and a graphical user interface of D-Leaf is available for download from: JW Tan & Siow-Wee Chang (2017), D-Leaf dataset. fig share. https://figshare.com/articles/D-Leaf_Dataset/5732955.[1]

VII. ACKNOWLEDGMENT

We welcome this opportunity to express my heartfelt gratitude and regards to Prof Anand Bhosale, Department of Information Technology, International Institute of Information Technology, Pune, Maharashtra, India, for his unconditional guidance.

REFERENCES

[1] Jing Wei Tan, Siow-Wee Chang, Sameem Abdul-Kareem, Hwa Jen Yap, Kien-Thai Yong, “Deep Learning for Plant Species Classification using Leaf Vein Morphometric”, : DOI 10.1109/TCBB.2018.2848653, IEEE/ACM Transactions on Computational Biology and Bioinformatics

[2] Pawan P. Warne1, Dr. S. R. Ganorkar2 , “Detection of Diseases on Cotton Leaves Using K-Mean Clustering Method”, Volume 2 April 2015,IRJET.

[3] Krithika, N., Selvarani, “An Individual Grape Leaf Disease Identification Using Leaf Skeletons and KNN Classification,” ICIIECS March 2017.

[4] Dias, R. O. Q., Borges, D. L., “Recognizing Plant Species in the Wild: deep learning results and a new database,” 2016 IEEE ISM.

[5] Vishnu S and A. Ranjith Ram, “Plant Disease Detection Using Leaf Pattern: A Review,” IJISET Volume 2 Issue 6 June 2015.

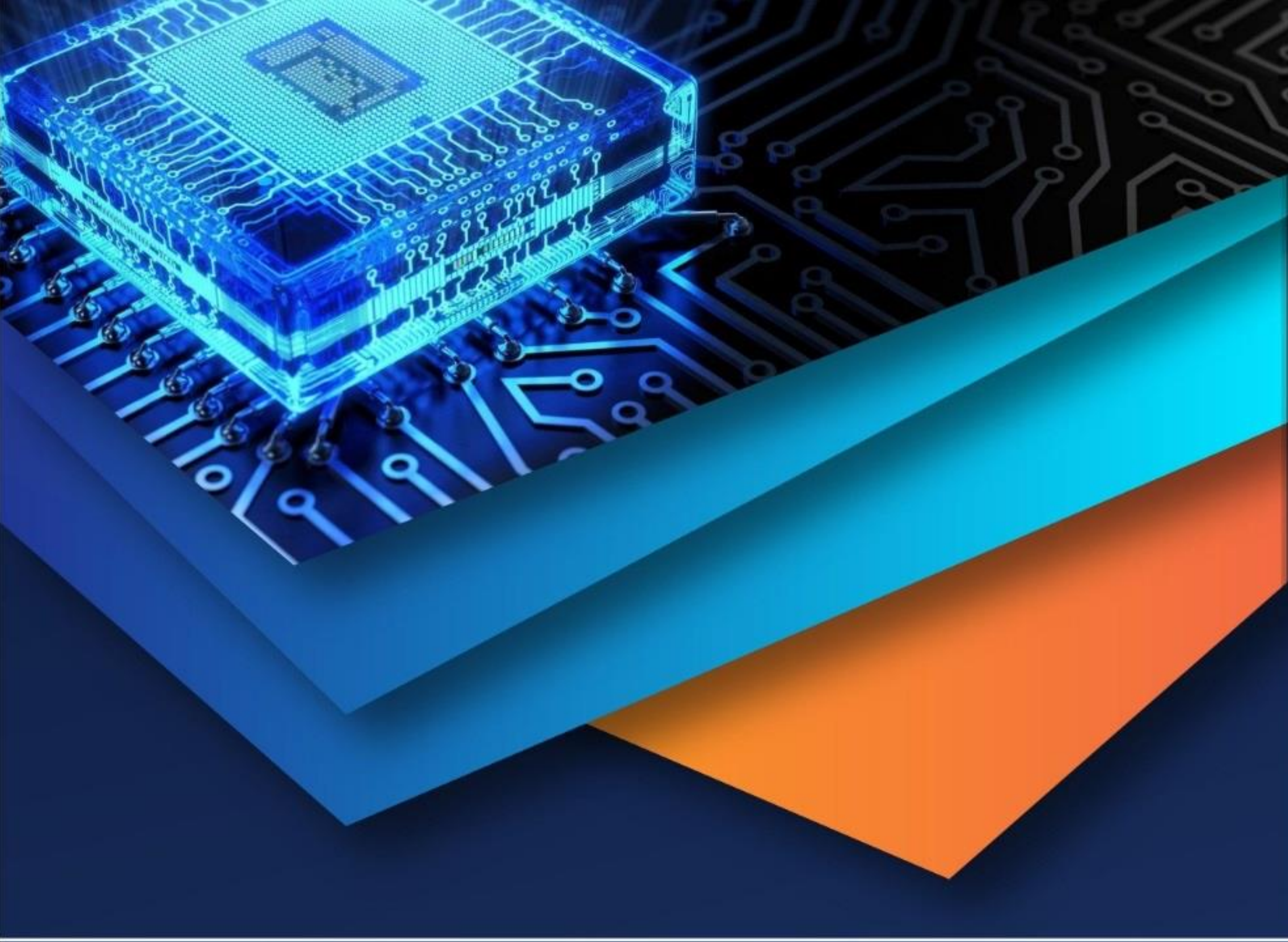
[6] Elhariri, E., El-Bendary, N., Hassanien A. E.,” Plant Classification System based on Leaf Features”,ICCES December 2014

[7] Trishen Munisami, Mahesh Ramsuran, Somveer Krishnah, SameerchandPudaruth,” Plant leaf recognition using shape feature and color histogram with k nearest classifier”.August 2015,Procedia Science direct.

[8] J. Fan, N. Zhou, J. Peng, L. Gao, “Hierarchical Learning of Tree Classifiers for Large-Scale Plant Species Identification,” IEEE TRANSACTIONS ON IMAGE PROCESSING, 2015.

[9] Prof.Anand Bhosale ,” Efficient Estimation of Complex Queries”,International Journal of Advance Research in Computer Science and Software Engineering,Volume 4 April 2016.

[10] Sagar P. Jaikar, Ramkrushna C. Maheshwar, Sameer P. Mamadapure Anand A. Bhosle ,” Secure Data Distribution using Secret Splitting Over Cloud”,Global Journal of Computer Science and Technology :B Cloud and Distributed.,Volume 17 Issue 1,2017.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY