iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** V **Month of publication:** May 2020

DOI: http://doi.org/10.22214/ijraset.2020.5421

Medicinal Plant Identification using Image Processing Technique

Nayana G. Gavhale[1], Dr. A. P. Thakare[2]
[1, 2]*Department of Electronics and Telecommunication , Sipna College of Engineering and Technology Amravati, Maharashtra*

Abstract: From Vedic times plants have been used as a source of medicine in ayurveda. In the preparation of ayurvedic medicine, identification of correct plant is the most important step, which have been done manually. Due to demand of mass production, Identification of these plants automatically using image processing is important. In this paper we have been implement a technique for medicinal plant identification using random forest algorithm, an ensemble supervise machine learning algorithm based on colour, texture and geometrical features.
Keywords: **Medicinal Plant, Image Processing, Identification, Automatic, Ayurveda.**

I. INTRODUCTION

Ayurveda is an ancient system of medicine practiced in India and has its roots in the Vedic times, approximately 5000 years ago. The main constituents of ayurvedic medicines are plant leaves and other parts of plants like root, bark etc. More than 8000 plants of Indian origin have been found to be of medicinal value. Combinations of a small subset amounting to 1500 of these plants are used in Herbal medicines of different systems of India. Specifically, commercial Ayurvedic preparations use 500 of these plants. Over 80% of plants used in ayurvedic formulations are collected from the forests and wastelands whereas the remaining are cultivated in agricultural lands

In the ancient past, the Ayurvedic physicians themselves picked the medicinal plants and prepared the medicines for their patients. Most of the plants are identified using their leafs the common steps to classify the leaf of a plants are Capturing image, noise removal and resizing ,Extracting features, use proposed methodology and finally identify or recognized the plant.

II. SURVEY ON EXISTING METHODOLOGIES

A.Gopal et.al [1] implement a system using image processing with images of the plant leaves as a basis of classification. The software returns the closest match to the query. The proposed algorithm is implemented and the efficiency of the system is found by testing it on 10 different plant species. The software is trained with100 (10 number of each plant species) leaves and tested with 50 (tested with different plant species) leaves. The efficiency of the implementation of the proposed algorithms is found to be 92%.

Umme Habiba et.al [2] In this paper, for automatically classifying medicinal plants, they present a Multichannel Modified Local Gradient Pattern (MCMLGP), a new texture-based feature descriptor that uses different channels of color images for extracting more significant features to improve the performance of classification. Auther have trained their proposed approach using SVM classifier with arious kernels such as linear, polynomial and HI. In addition, used different feature descriptors for comparative experimental analysis with MCMLGP by conducting the rigorous experiment on our own medicinal plants dataset. The proposed approach gain higher accuracy (96.11%) than other techniques, and significantly valuable for exploration and evolution of medicinal plants classification.

R.Janani et.al[3] have proposed a method for the extraction of shape, color and texture features from leaf images and training an artificial neural network (ANN) classifier to identify the exact leaf class. The key issue lies in the selection of proper image input feature to attend high efficiency with less computational complexity. They tested the accuracy of network with different combination of input feature. the test result on 63 leaf images reviles that this method gives 94.4% accuracy with a minimum of 8 input features. this approach is more prominent for leaf identification system that have minimum input and demand less computational time.

Vijayashree.T et.al [4] has created database with 127 herbal leaves. For creating a database 11 texture parameters are taken into account. The parameters are Sum of Variance, Inverse Difference Moment, Aspect ratio, Correlation, Sum Entropy, Mean, and Sum Average. Gray level co-occurrence matrix (GLCM) is used for determining the parameters like entropy, homogeneity, contrast and energy. A test image is taken and compared with the database; the dissimilarity is calculated with the extracted parameters. The one with least dissimilarity is identified as the leaf and the output is displayed.

Venkataraman et.al [5] a system is developed which would provide a solution for identifying the plant and providing it's medicinal values, thereby helping in the cure of many ailments in a natural way. This paper discusses about the dataset collection, feature extraction using texture and HOG and thereby classifying based on Support Vector Machine algorithm.

Shitala Prasad et.al [6] These paper presents a new and efficient technique for leaf acquisition.The image is transformed to device independent lαβ color space that is further used to compute VGG-16 feature map. This feature map is re-projected to PCA subspace to optimize the performance for species recognition. To prove the robustness, the paper uses two different types of plant leaf datasets.

Dileep M.R et.al [7] This work proposes AyurLeaf, a Deep Learning based Convolutional Neural Network (CNN) model, to classify medicinal plants using leaf features such as shape, size, color, texture etc. This research work also proposes a standard dataset for medicinal plants, commonly seen in various regions of Kerala, the state on southwestern coast of India. The proposed dataset contains leaf samples from 40 medicinal plants. A deep neural network inspired from Alexnet is utilised for the efficient feature extraction from the dataset. Finally, the classification is performed using Softmax and SVM classifiers. Our model, upon 5-cross validation, achieved a classification accuracy of 96.76% on AyurLeaf dataset. AyurLeaf helps us to preserve the traditional medicinal knowledge carried by our ancestors and provides an easy way to identify and classify medicinal plants.

C.Amuthalingeswaran et.al [8] had built a model (Deep Neural Networks) for the identification of medicinal plants. To train the model auther used around 8,000 images belonging to four different classes. Finally, arrived with good accuracy of 85% when testing with images taken from the open field landareas.

Manojkumar P. et.al [9]This paper explores feature vectors from both the front and back side of a green leaf along with morphological features to arrive at a unique optimum combination of features that maximizes the identification rate. A database of medicinal plant leaves is created from scanned images of front and back side of leaves of commonly used ayurvedic medicinal plants. The leaves are classified based on the unique feature combination. Identification rates up to 99% have been obtained when tested over a wide spectrum of classifiers. The above work has been extended to include identification by dry leaves and a combination of feature vectors is obtained, using which, identification rates exceeding 94% have been achieved.

Amala Sabu et.al [11] The proposed system uses a combination of SURF and HOG features extracted from leaf images and a classification using k-NN classifier. Our experiments show results which seem to be sufficient for building apps for real life use.

III. PROPOSED METHODOLOGY

The proposed methodology, have been used Bagging or Random Forest Ensemble Algorithms for Machine Learning to identify the medicinal plant as shown in the fig.1 .

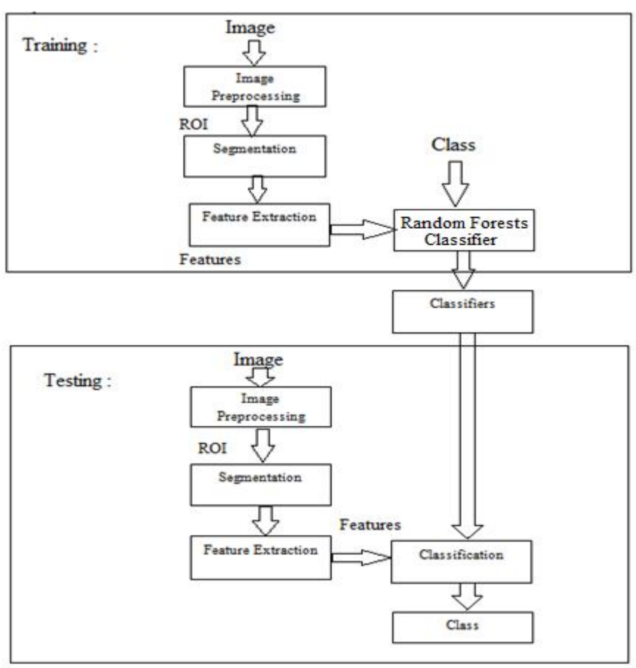


Fig.1. Block Diagram for Medicinal Plant identification

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

A. Image Pre-Processing

Due to various noises present in the raw images which are captured by camera are not suitable for direct processing. Therefore it is necessary to pre-process it before examining. The pre-processing involved as gray scale conversion which will convert RGB image into Gray scale image, image resize, noise removal and image enhancement and then this image is given to the feature extraction process in conversion.

B. Image Segmentation

Segmentation procedure partition an image into its constituent parts or objects. The grain components are separated by using image segmentation process.The field of mathematical morphology contributes to a wide range of operators to image processing. we are only handling binary images. For a binary image, black pixels (“0”) are normally taken to represent background regions, while white pixels (“1”) denote foreground. The two most basic operations in mathematical morphology are dilation and erosion. These operations can be considered as morphological non-linear filters. The segmented image undergoes a series of morphological operations to detect the exact shape of the object. Morphological image processing is a collection of non-linear operations related to the shape or morphology of features in an image using erosion, dilation and image opening, closing.

C. Image Feature Extraction

1) *Color Feature*

a) *Color Histograms:* Color histograms are defined as a set of bins where each bin denotes the probability of pixels in the image being of a particular color. A color histogram for a given image is defined as a vector:

H □ {H[0], H[1], H[2],□, H[i], □, H[N]}

where i represents a color in the color histogram and corresponds to a sub cube in the RGB color space, H[i] is the number of pixels in color i in that image, and N is the number of bins in the color histogram, i.e., the number of colors in the adopted color model

b) Color Moment: Color moments are measures that can be used differentiate images based on their features of color. Once calculated, these three moments that is mean, standard deviation and Skewness provides a measurement for color similarity between images. These values of similarity can then be compared to the values of images indexed in a database for tasks like image retrieval. An image therefore is characterized by 9 moments 3 moments for each 3 color channels. We will define the ith color channel at the jth image pixel as pij . The three color moments, mean, standard deviation and skewness can then be defined as:

$$\text{Mean} = \sum_{N}^{j=1} \frac{1}{N} p_{ij}$$

$$\text{Standard Deviation} = \sqrt{\left(\frac{1}{N} \sum_{N}^{j=1} (p_{ij} - E_i)^2\right)}$$

$$\text{Skewness} = \sqrt[3]{\left(\frac{1}{N} \sum_{N}^{j=1} (p_{ij} - E_i)^3\right)}$$

2) *Texture Features:* A statistical method of examining texture that considers the spatial relationship of pixels is the gray-level co-occurrence matrix (GLCM), also known as the gray-level spatial dependence matrix. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix.

a) *Statistical method as*

$$\text{Contrast} = \sum_{i,j} |i-j|^2 p(i,j) \qquad \text{Energy} = \sum_{i,j} p(i,j)^2$$

$$\text{Correlation} = \sum_{i,j} \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i \sigma_j} \qquad \text{Homogeneity} = \sum_{i,j} \frac{p(i,j)}{1+|i-j|}$$

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

3) *Geometric Features:* Convex area specifies the number of wide pixels in the convex image. a convex image is a binary image that specifies the smallest convex polygon that can content the region, with all pixels within the polygon filled in.

$$\text{Equivalent Diameter} = \sqrt{\frac{4 * Area}{\pi}}$$

$$\text{Solidity} = \frac{Convex\ Area}{Original\ Area}$$

Eccentricity is the characteristics of any conic section and it is given as

$$\text{Eccentricity} = \sqrt{1 - \frac{(Minor\ axis\ length)^2}{(Major\ axis\ length)^2}}$$

$$\text{Extent} = \frac{Number\ of\ pixels\ in\ the\ region}{Number\ of\ pixels\ in\ the\ bounding\ box}$$

Compactness is the ratio of 4π times Area to the square of perimeter

$$\text{Compactness} = \frac{4\pi * Area\ of\ leaf}{Perimeter^2}$$

Aspect ratio is the slimness, it is given by

$$\text{Aspect ratio} = \frac{Maximum\ axial\ length}{Minimum\ axial\ length}$$

$$\text{Entirety} = \frac{Convex\ Area - Area}{Area}$$

$$\text{Perimeter ratio of length and width} = \frac{P}{L+W}$$

Where P,L,W are the Perimeter , Length and Width respectively

D. Random Forests Classifier

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes(classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of over fitting to their training set To understand and use the various options, further information about how they are computed is useful. Most of the options depend on two data objects generated by random forests. When the training set for the current tree is drawn by sampling with replacement, about one-third of the cases are left out of the sample. This oob (out-of-bag) data is used to get a running unbiased estimate of the classification error as trees are added to the forest. It is also used to get estimates of variable importance. After each tree is built, all of the data are run down the tree, and proximities are computed for each pair of cases. If two cases occupy the same terminal node, their proximity is increased by one. At the end of the run, the proximities are normalized by dividing by the number of trees. Proximities are used in replacing missing data, locating outliers, and producing illuminating low-dimensional views of the data.

IV. RESULT

In this section, the experiments conducted to test the quality of the results generated by the proposed method are discussed. The collected dataset consists of images of the leaves of aloevera, Coriander, drumstick, hibiscus, Mint, neem, papaya, tulsi-Basil and etc. Under the guidance of domain expert we created dataset for training and testing. In the carried experiments, random forest classifiers, an ensemble supervised machine learning algorithm were tested. The proposed technique has been implemented by using the MATLAB R2018b environment on Core i3 with processor speed of 2.2 GHz. We used image processing, statistics and machine learning toolbox and guide functions to implement the algorithm in MATLAB. The user defined interface for project development is as shown in fig.

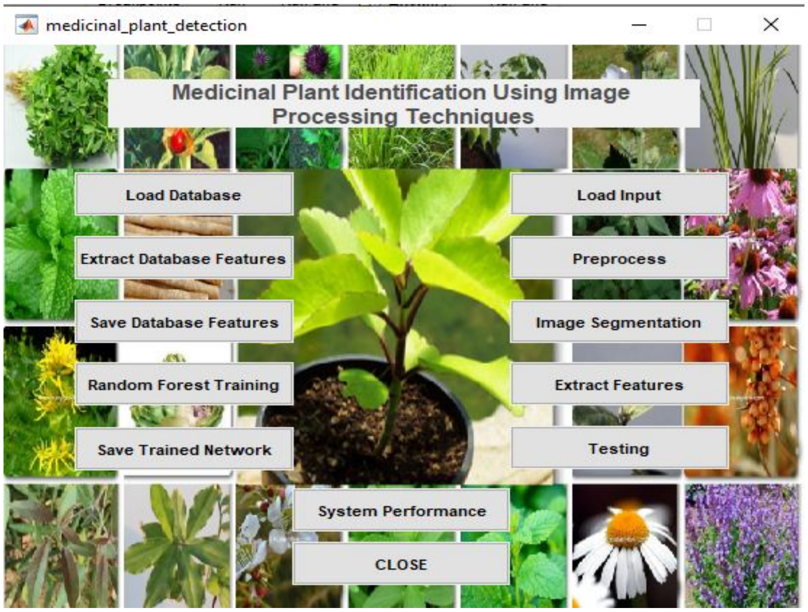


Fig: UI for project development

The images were initially pre-processed and then the features were extracted from each leaf image with the help of the above mentioned equations. Separate pre-processing steps were employed for color, texture based GLCM, GRLM and geometrical feature based analysis. Sample testing procedure images as shown below in fig.

Fig: a) Test Leaf Image (Hibiscus) b) Gray Scale Image c) Leaf Image after morphological operations

Different methods are used to measures and evaluate the performance. The measure used here is the Classification Accuracy. These values are calculated from the confusion matrix. Confusion matrix is the table that describes the performance of the classifier or classification model. It is mainly applied on a set of test data for which the true values are known. Confusion matrix holds the information or the data about the actual and predicted values. The following table shows the confusion matrix for the binary classification.

		Predicted	
		Positive	Negative
Actual	Positive	TP	FP
	Negative	FN	TN

Table : confusion matrix

TP (true positives) - Correct prediction of abnormal.
TN (true negative) - Correct prediction of normal.
FN (false negative) - Incorrect prediction of normal.
FP (false positive) - Incorrect prediction of abnormal.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

During the testing stage, the dataset for testing is provided to the proposed method to detect or classify the medicinal plant. Then evaluation metrics (specificity, accuracy, sensitivity) are used to evaluate the acquired results. The following formulas are used to find the values of these metrics in medicinal images.

Accuracy can be calculated using confusion matrix.It is defined as the ratio of correctly classified images to the total images. The classification accuracy is calculated as

$$Accuracy = (TN + TP)/(TN + TP + FN + FP)$$

Performance tests are evaluated using Sensitivity, Specificity and F-score.

1) *Sensitivity (or Recall):* It is a true positive rate (TPR). It is the ratio of correctly classified true images to the total test images. The sensitivity is calculated as

$$Sensitivity = TP/(TP + FN)$$

2) *Specificity:* It is a true negative rate (TNR). It is the ratio of correctly classified false images to the total false test images.

The specificity is calculated as

$$Specificity = TN/(TN + FP$$

3) *F-score:* is the harmonic mean of the precision and recall, precision provides the number of positive class predictions that actually belong to the positive class and recall quantifies the number of positive class predictions made out of all positive examples in the dataset.

$$Precision = TP / (TP + FP)$$
$$F\text{-}score = (2 * Precision * Recall) / (Precision + Recall)$$

The software is trained with a total of 150 leaves across 10 categories. A total of 110 leaves from the test database were tested, 104 out of these total were identified correctly giving a success rate of 94.54 % to the program and remaining parameter according to above equation listed in table below.

Table : Testing Performance

Parameters	Testing Performance (Percentage)
Accuracy	94.54
Sensitivity	96.23
Specificity	50.00
F-score	97.70

V. CONCLUSIONS

In this paper we have implemented a technique for medicinal plant identification using random forest algorithm, an ensemble supervise machine learning algorithm based on color ,texture and geometrical features to identify the correct species of medicinal plant. the combination of shape, color and texture features result in correct leaf identification accuracy of 94.54 %.The results shown in this technique are very promising and thus indicate the aptness of this algorithm for medicinal plant identification systems. this work can be extended to a larger number of Plants species with improved accuracy in future.

VI. ACKNOWLEDGMENT

I would like to express my sincere thanks to my Guide, Head of the department and Principal of Sipna college of Engineering and Technology Amravati for their constant encouragement and motivating me in my research work.

REFERENCES

[1] A.Gopal , S. Prudhveeswar Reddy and V. Gayatri " Classification of Selected Medicinal Plants Leaf Using Image Processing ” Published in 2012 International Conference on Machine Vision and Image Processing (MVIP)

[2] Umme Habiba, Md. Rasel Howlader, Md. Aminul Islam, Rahat Hossain Faisal and Md. Mostafijur Rahman"Automatic "Medicinal Plants Classification using Multi-channel Modified Local Gradient Pattern with SVM Classifier" published in Joint 2019 8th International Conference on Informatics, Electronics & Vision (ICIEV) & 3rd International Conference on Imaging, Vision & Pattern Recognition (IVPR)

[3] R.Janani, A.Gopal "Identification of selected medicinal plant leaves using image feature and ANN" published in 2013 international conference on advanced electronic system (ICAES)

[4] Vijayashree.T, Dr. A. Gopal "Leaf Identification for the Extraction of Medicinal Qualities Using Image Processing Algorithm" published in 2017 International Conference on Intelligent Computing and Control (I2C2)

[5] D Venkataraman, Mangayarkarasi N "Support Vector Machine Based Classification of Medicinal Plants Using Leaf Features" published in 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)

[6] Shitala Prasad, Pankaj P Singh"Medicinal Plant Leaf Information Extraction Using Deep Features" published in Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017

[7] Dileep M.R., Pournami P.N."AyurLeaf: A Deep Learning Approach for Classification of Medicinal Plants" published in TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)

[8] C.Amuthalingeswaran, Mr. M.Sivakumar, Dr. P. Renuga" IDENTIFICATION OF MEDICINAL PLANT’S AND THEIR USAGE BY USING DEEP LEARNING" Published in Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8

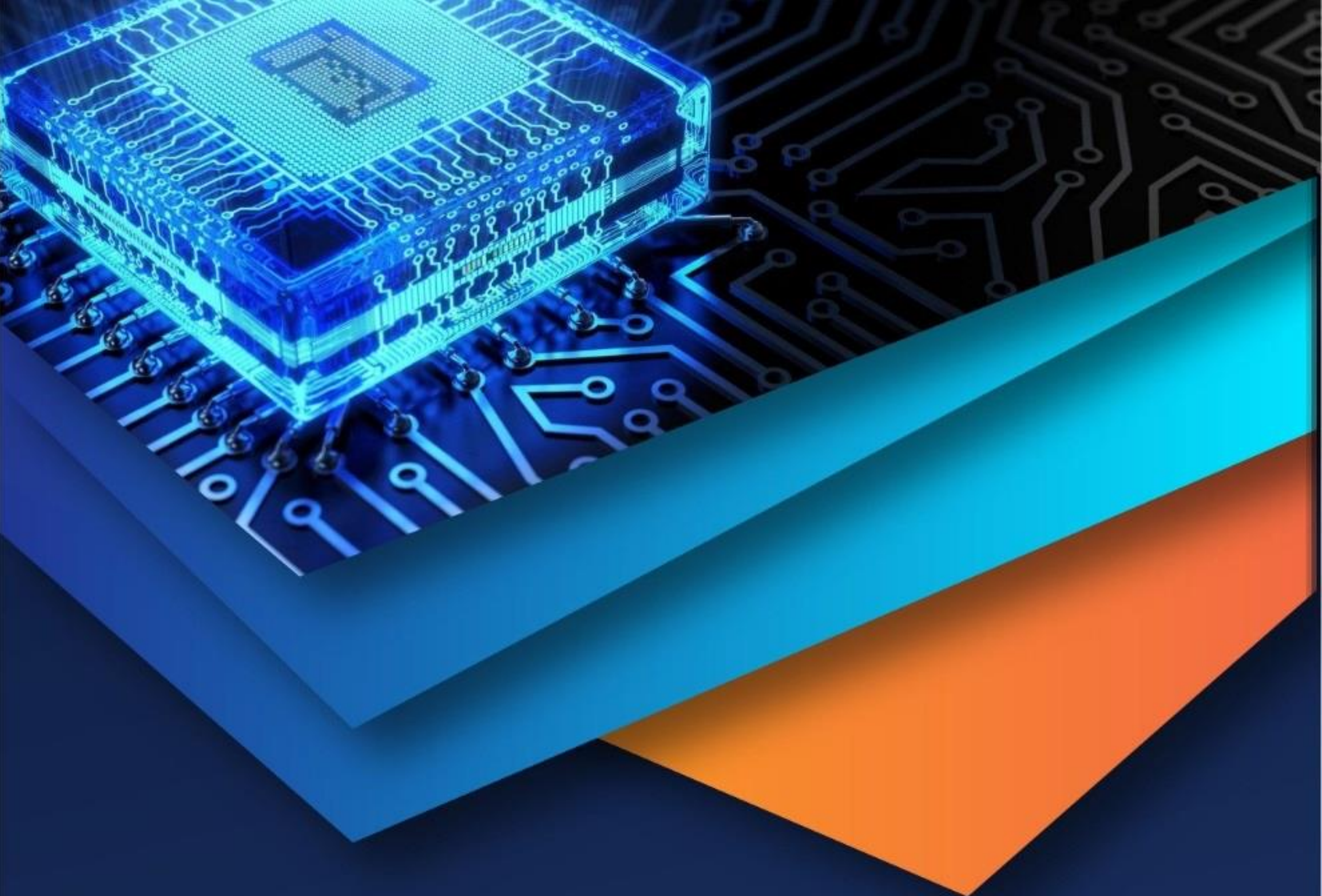
[9] Manojkumar P., Surya C. M., Varun P. Gopi "Identification of Ayurvedic Medicinal Plants by Image Processing of Leaf Samples" Published in 2017 Third international conference on research in computational Intelligence and communication network (ICRCICN)

[10] Sujit A. ,Aji. S "An Optimal Feature set With LPB for Leaf Image classification" Published in Proceedings of the Fourth International Conference on Computing Methodologies and Communication (ICCMC 2020) IEEE Xplore Part Number:CFP20K25-ART; ISBN:978-1-7281-4889-2

[11] Amala Sabu, Sreekumar K, Rahul R Nair "Recognition of Ayurvedic Medicinal Plants from Leaves: A Computer Vision Approach" Published in 2017 Fourth International Conference on Image Information Processing (ICIIP)

[12] S.Perumal1 and T.Velmurugan2 ” Preprocessing by Contrast Enhancement Techniques for Medical Images" published in International Journal of Pure and Applied Mathematics Volume 118 No. 18 2018, 3681-3688

[13] Nourhan Zayed and Heba A. Elnemr " Statistical Analysis of Haralick Texture Features to Discriminate Lung Abnormalities" Published in International Journal of Biomedical Imaging / 2015 Article ID 267807



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY