Leaf Classification by Random Forest Classifier Based On Leaf Features Fusion & Zernike Moment

[1] Jignesh B. Jethva, [2] Dr. C. R. Singla
[1] PhD Scholar, [2] Professor
[1, 2] Electronics and Communication Department
[12] Madhav University, Abu Road, Pindwara, Dist. Sirohi, Bharja, Rajasthan, INDIA

Abstract: This paper presents a classification approach by Random Forest Classifier algorithm for classifying the different types of plants. This Proposed Approach consists of four phases that are image pre-processing, feature extraction, features fusion and classification phases. Most types of plants have unique leaves. There are many features of leaf such as Color features, Vein features, GLCM features, Shape features and Gabor features. Also calculate Zernike moments such as amplitudes and phase. These all features are fused by concatenating of two vectors. So classification approach presented in this research depends on plant's leaves. Experimental results show accuracy and other parameters measured in this approach with fusion of all these features or their different combinations.

Keywords **- Random Forest, Zernike Moment, Gabor Filters, GLCM.**

I. INTRODUCTION

Customary acknowledgment of plant species is done by manual coordinating of the plant's highlights, identifying with segments of the plant, for example, leaves, blossoms, and bark, against a map book (Meeta, 2012). Endeavors to computerize this procedure have been made, utilizing highlights of plants extricated from pictures as info parameters to different classifier frameworks (Cope, 2011). Since plant leaves are frequently more accessible than the foods grown from the ground, and in light of the fact that leaves are likewise for the most part two-dimensional (2D) fit as a fiddle, the greater part of the current work on PC construct plant acknowledgment are based with respect to the leaves of plants. This work looks at existing frameworks of PC based robotized framework for ID of plant species and different methods utilized in that.

It is outstanding that plants assume a critical part in safeguarding earth's nature and condition by keeping up a solid climate and giving sustenance and asylum to complex bug and creature species. Plants are additionally imperative for their therapeutic properties, as elective vitality sources like bio-fuel and for meeting our different household necessities like timber, apparel, sustenance and beauty care products. Leaves assume a central part in recognizable proof of a plant because of its handiness almost consistently, effectiveness to access, convey and process in PC. As the state of plant leaves is a standout amongst the most critical highlights for portraying different plants outwardly, the investigation of leaf picture recovery plans will be an essential stage for building up a plant distinguishing proof framework.

A standout amongst the most critical errands for researchers, field aides, and others is order of plants, since plants have a vital part in the normal hover of life. They are essential to relatively every other type of life, as they frame the biggest piece of the living beings that can change over the daylight into sustenance. Likewise, as all oxygen noticeable all around that people and different creatures inhale is delivered by plants, henceforth without plants it is hard to consider presence of human life on earth. Grouping plants helps at guaranteeing the insurance and survival of all regular life. The procedure of plant grouping can be performed utilizing diverse courses, for example, cell and sub-atomic science and in addition utilizing the plants' takes off.

Most sorts of plants have one of a kind leaves that are not the same as each other in view of various attributes, for example, shape, shading, surface, and the edge. The considerable data conveyed by each can be utilized to distinguish and arrange the source or the kind of plant, so leaf acknowledgment/order is imperative undertaking at the procedure of plant grouping.. As of late, using PC vision in agribusiness field has turned out to be extremely across the board, particularly for objects where appearances convey a ton of data.

II. MOTIVATION

The human visual system has no problem interpreting the subtle variations in translucency and shading in this Figure 1. Photograph and correctly segmenting the object from its background.

Figure 1.Lotus flower seen as to the naked eye.

We should envision a man taking a field excursion, and seeing a hedge or a plant on the ground, he or she might want to know whether it's a weed or some other plant however has no clue about what sort of plant it could be. With a decent computerized camera and an acknowledgment program, one could get some valuable data. "Plants accept an imperative part in our environment. Without plants there will be no nearness of the world's temperament. In any case, starting late, numerous sorts of plants are at the peril of end. To guarantee plants and to list distinctive sorts of greenery assorted varieties, a plant database is a basic walk towards insurance of earth's biosphere. There are countless species around the globe. To deal with such volumes of information, change of a smart and viable portrayal system has transformed into a district of dynamic research. Despite the conservation edge, affirmation of plants is also imperative to utilize their remedial properties and using them as wellsprings of choice essentialness sources like bio-fuel. There are a couple of ways to deal with see a plant, like blossom, root, and leaf, natural item et cetera".

Objectives of System

- To confirm the importance of leaf length, width, area and perimeter.
- To protect plants which are at the verge of extinction.
- To improve learning of plant species.
- To easy diagnoses of plant diseases.
- To improve effective use of medicine properties of plant.
- To increase accuracy of Plant Identification based on Feature Fusion.
- To be suitably depicted with the assistance of consistent shape descriptors like circularity, linearity thus on in light of Zernike Moment.

III. PROPOSED WORK

In Our Proposed Approach, First step is Leaf Image Acquisition. In this step digital leaf image is captured. Then apply pre-processing step on these leaf image. Prior to the operations, a portion of the leaf pictures are turned physically to help the program to mastermind leaf zenith bearing to the correct side. A short time later, programmed pre-handling procedures are connected to the greater part of the leaf pictures. Pre-processing steps involves converting RGB to Grayscale Image, then apply Median Filtering on it, then converting into binary and apply segmentation on it. After pre-processing, the important and essential task is to measure the properties of an object which is called Feature Extraction because objects have to be detected based on these computed properties. In Feature extraction, I will extract Features such as Color Features, Texture Features, Shape Features, and Vein Features and also apply Zernike Moments of Leaf Image. After Feature Extraction, next step is Feature Fusion to combine more than one Feature to get more accuracy for classification.

Once the features have been fused, then these features vectors are to be used to classify and identify plant using RF (Random Forest) classifier to classify plants. A brief explanation on the proposed system is given in the Figure 2.

Figure 2. Proposed Work

A. Algorithm for Proposed System

Steps:

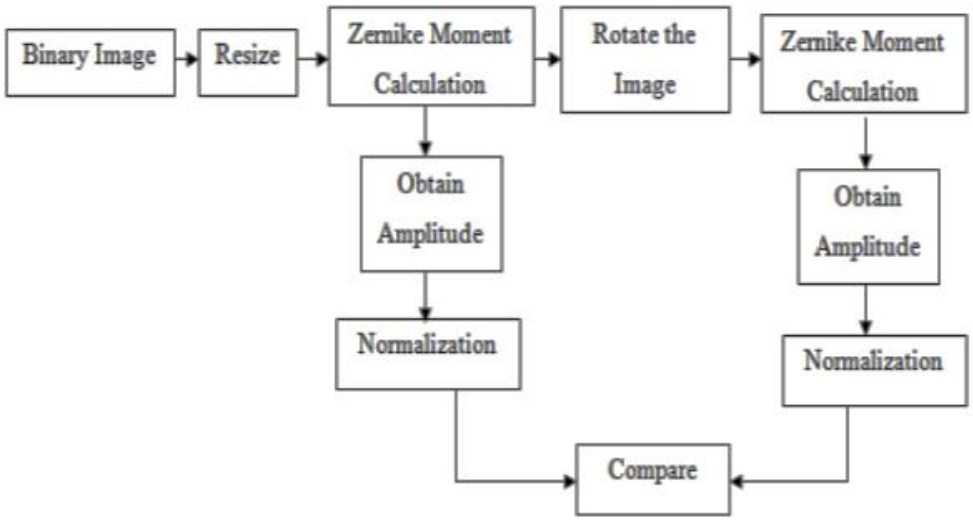
1. Prepare Training Dataset
   - 1.1. Collect Plant Leaf Samples.
   - 1.2. Acquisition of plant leaf images
   - 1.3. Apply Preprocessing on each plan leaf image includes Gray conversion, median filtering and then binarization and segmentation.
   - 1.4. Extract Features of plant leaf such as shape, color, vein, texture etc. and apply Zernike moment.
   - 1.5. Fuse the features based on combination.
   - 1.6. Prepare features vector.
2. Read the testing plant leaf image
3. Apply Pre-processing on test image including same steps in step 1.3
4. Extract Features specified in Step 1.4 and Fuse them based on combination.
5. Train the training dataset and predicate testing image by using Random Forest Classifier.
6. Finally, identify the plan leaf.
7. Stop.

B. Features Extraction In Proposed Work

Features	Sub Features	Description
Shape Features	**Eccentricity**	"It is defined as the ratio of the distance between the foci of the ellipse and its major axis length. It is used to differentiate rounded and long leafs"
	Solidity	"It is the ratio between object's area and area of the object's convex hull. It may be considered as a certain measure of convexity" Solidity=$\frac{A(I)}{A(H(I))}$ Where A(I) is the object area and A(H(I)) is the area of object's convex hull.
	Aspect Ratio (AR)	"It is the ratio between the maximum length D_{MAX} and the minimum length D_{MIN} of the minimum bounding rectangle". $AR = \frac{D_{MAX}}{D_{MIN}}$
	Width Ratio (WR)	"It is the ratio of width at half of major axis to maximum width".
	Perimeter	"It is scalar that specifies the distance around boundary of the region"
	Area	"It is scalar that specifies the actual number of pixels in the region"
	Roundness or Circularity	"It is the ratio of 4*PI*Area of the leaf to the square of perimeter"
	EquivDiameter	"It is scalar that specifies the diameter of a circle with same area as the region computed as sqrt (4*Area/pi)"
	Centroid	"It is 1 by Q vector that specifies the center of mass of region. First element of centroid is horizontal coordinate of center of mass and second element is the vertical coordinate"
	Convex Area	"It is scalar that specifies the number of pixels in convex Image"
	Convex Hull	"It is p by 2 matrixes that specify smallest convex polygon that contain the region"
Color Features	**Mean**	$\bar{X}_i = \frac{\sum_{j=1}^{MN} X_{i,j}}{M \cdot N}$
	Standard deviation	$\vartheta_i = \sqrt{\frac{1}{M \cdot N}\sum_{j=1}^{MN}(X_{i,j} - \bar{X}_i)^2}$
	Skewness	$\boldsymbol{S_i} = \sqrt[3]{\frac{1}{M \cdot N}\sum_{j=1}^{MN}(X_{i,j} - \bar{X}_i)^3}$
	Kurtosis	$\boldsymbol{K_i} = \sqrt[4]{\frac{1}{M \cdot N}\sum_{j=1}^{MN}(X_{i,j} - \bar{X}_i)^4}$
	Where $X_{i,j}$ is the value of image pixel j of color channel i. $\bar{X}_i$ is the mean for each channel i. ϑ_i is the standard deviation , $\boldsymbol{S_i}$is skewness and $\boldsymbol{K_i}$is kurtosis for each channel	
Vein Features	"Vein features are features derived from vein of the leaf. There are four kinds of vein features, definedas follows: V1=A1/A, V2=A2/A, V3=A3/A, V4=A4/A WhereA1, A2, A3	

	and A4 are pixel number that constructs the vein and A is area of the leaf".	
TextureFeatures	**GrayLevel Co-occurrence Matrix (GLCM)**	$$Angular\ Second\ Moment = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} {P_{ij}}^2$$ $$Contrast = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1}(i-j)^2 P_{ij}$$ $$Correlation = \frac{1}{\vartheta_x \vartheta_y}\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}[ijP_{ij} - \mu_x \mu_y]$$ $$Entropy = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} P_{ij} \log P_{ij}$$ $$variance = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1}(i-\pi)^2 P_{ij}$$ $$Homogeneity = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\frac{1}{(i-j)^2} P_{ij}$$ $$Sum\ of\ Entropy = -\sum_{i=2}^{2G-2} P_{x+y}(i) \log P_{x+y}(i)$$ $$Cluster\ Shade = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1}(i+j-\pi_x-\pi_y)^3 P_{ij}$$ $$Prominence = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1}(i+j-\pi_x-\pi_y)^4 P_{ij}$$ Where $\mu_x, \mu_y, \vartheta_x$ and $\vartheta_y$mean and standard deviation of corresponding distribution and G are is number of Gray levels
	Gabor Filter	A complex Gabor filter is defined as the product of a Gaussian kernel and a complex sinusoid. A 2D Gaussian curve g with a spread of σ in both x and y directions", is represented as below: $$g(x,y,\sigma) = \frac{1}{2\pi\sigma^2}\exp(-\frac{x^2+y^2}{2\sigma^2})$$ $(x,y,u,\theta,\varphi) = \exp\{j2\pi(x.u\cos\theta + y.u\sin\theta)+\varphi\}$

C. Zernike Moment



V. PERFORMANCE PARAMETERS

The performance of the proposed system is tested with Random Forest classifiers by using the feature set extracted in the dataset. Confusion matrix, sensitivity, accuracy, kappa statistics, RMSE and AUROC metrics are measured

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$RMSE = \sqrt{\frac{\sum(y' - y)^2}{n}}$$

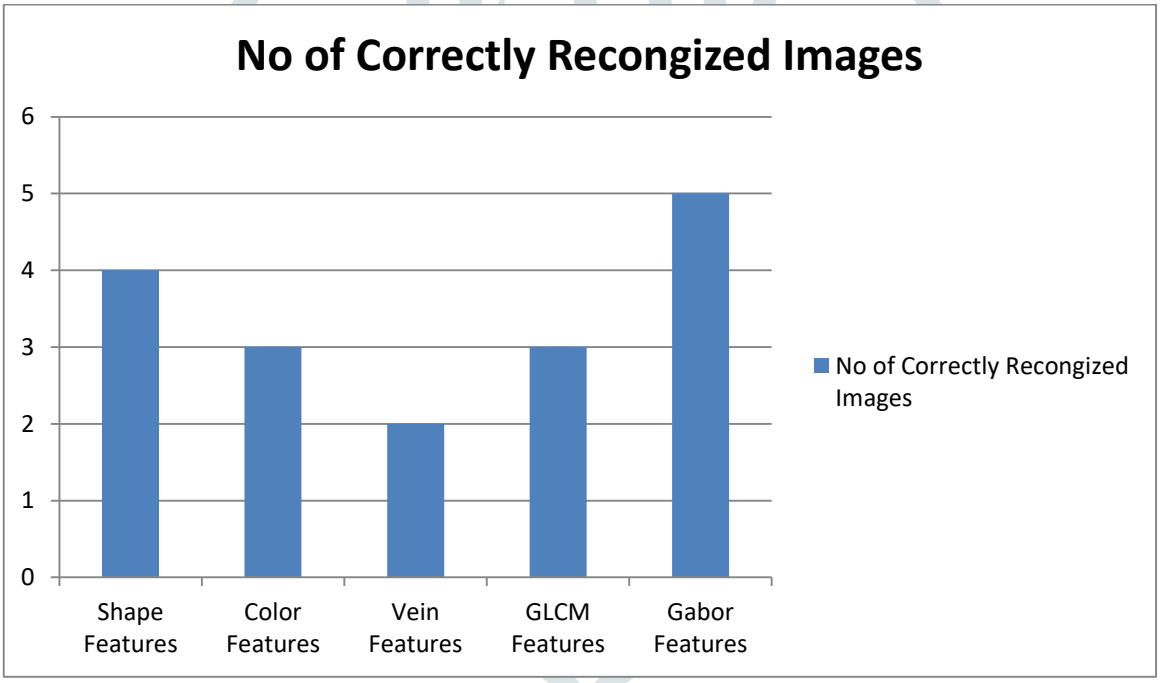
Here *TP*, *TN*, *FP* and *FN* denote the number of Leaf Images classified as true positive, true negative, false positive, and false negative, respectively. In the root mean squared error (*RMSE*), *y* and *y'* depict actual and predicted values. *n*is the number of Leaf images.

TP (True Positives) refers to the positive images that were correctly labelled by the classifier.
TN (True Negatives) refers to the negative images that were correctly labelled by the classifier.
FP (False Positives) refers to the negative images that were incorrectly labelled as positive by the classifier.
FN (True Positives) refers to the positive images that were incorrectly labelled as negative by the classifier.

VI. EXPERIMENTAL RESULTS

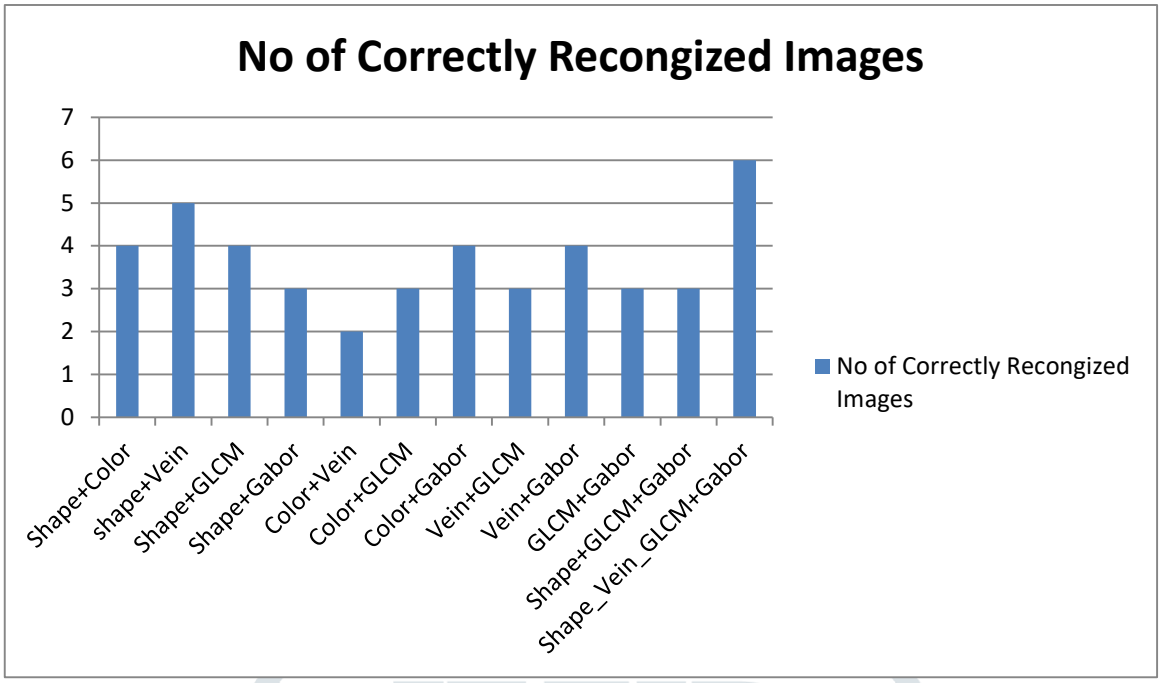
Experiment 1:
Aim: Leaf Recognition by Different Types of Features

No of Test Images	No of Correctly Recognized Images				
	By Shape Features	By Color Features	By Vein Features	By GLCM Texture Features	By Gabor Features
8	4	3	2	3	5



Experiment 2:
Aim: Leaf Recognition by Different Types of Features Fusion

No of Test Images	6
Shape+Color Features Correctly Recognized	4
Shape+Vein Features Correctly Recognized	5
Shape+GLCM Features Correctly Recognized	4
Shape+Gabor Features Correctly Recognized	3
Color+Vein Features Correctly Recognized	2
Color+ GLCM Features Correctly Recognized	3
Color+ Gabor Features Correctly Recognized	4
Vein + GLCM Features Correctly Recognized	3
Vein + Gabor Features Correctly Recognized	4
GLCM+ Gabor Features Correctly Recognized	3
Shape+ GLCM+ Gabor Features Correctly Recognized	3
Shape+Vien+GLCM+Gabor	6



Experiment 3:

Aim: Performance Parameters for Feature Fusion

No of test Images	2
Confusion Matrix	1 0 0 1
Accuracy	1.0000 0.7500
Sensitivity	1.0000 0.5000
Specificity	1.0000 1.0000
Precision	1 1
Recall	1.0000 0.5000
Fscore	1.0000 0.6667
Gorder	1 2

VII. CONCLUSION

The proposed classification approach was implemented by applying background removal, and extracting color components for each image. Then, feature extraction was applied to each pre-processed image, color, moments, shape, first order texture, GLCM and vein features are obtained as features vector and fused features vectors based on different combinations. Finally, RF models is developed for plant classification.We have use features fusion with Zernike moments to recognize plant leaf with accuracy more than 95%.

REFERENCES

1. "Leaf Recognition Based on Feature Extraction and Zernike Moments", Pallavi P, V.S Veena Devi, International Journal of Innovative Research in Computer and Communication Engineering, May 2014.
2. "A survey of computer-based vision systems for automatic identification of plant species", OluleyeBabatunde, Leisa Armstrong, Dean Diepeveen, JinsongLeng, Journal of Agricultural Informatics. 2015.
3. "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network", Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, Yi-Fan Chang and Qiao-Liang Xiang, International Symposium on Signal Processing and Information Technology, 2007
4. "A Review and a Comparative Study of Various Plant Recognition and Classification Techniques using Leaf Images", AnandHanda And RashiAgarwal, InternationalJournal of Computer Applications, August 2015
5. "Leaf Shape Extraction For Plant Classification", M. M. Amlekar, A. T. Gaikwad, R. R. Manza, P. L. Yannawar, 2015 International Conference on Pervasive Computing (ICPC), 2015

6. "Classification of selected medicinal plants leaf using image processing", Gopal, A, S. Prudhveeswar Reddy, and V. Gayatri. 2012 International Conference on Machine Vision and Image Processing (MVIP), 2012.
7. "Plant classification system based on leaf features", Elhariri, Esraa, Nashwa El-Bendary, and Aboul Ella Hassanien, 2014 9th International Conference on Computer Engineering & Systems (ICCES), 2014
8. "A review on plant recognition and classification techniques using leaf images", AnantBhardwaj, ManpreetKaur, International Journal of Engineering Trends and Technology- Volume 4 Issue 2- 2013
9. "Leaves recognition system using a neural network", BoranSekeroglua, YucelLnan, 12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS, 2016
10. "Multiscale Distance Matrix for Fast Plant Leaf Recognition", Hu, Rongxiang, Wei Jia, Haibin Ling, and De- Shuang Huang. IEEE Transactions on Image Processing, 2012.
11. "Plant leaf recognition using shape features and colour histogram with k-nearest neighbour classifiers", TrishenMunisami, MahessRamsurn, SomveerKishnah, SameerchandPudaruth, Second International Symposium on Computer Vision and the Internet, Elsevier, 2015
12. "Automated Tool for Plant Leaf Identification using Morphological Features", AjinkyaGawade, AnandSartape, AamodChemburkar, Prasad Somawanshi, ReenaPagare, JayshreeGhorpade, International Journal of Emerging Research in Management &Technology, May 2015
13. "Plant Identification Using Leaf Images.", Sachin D. Chothe, V.R.Ratnaparkhe, International Journal of Innovative Research in Science, Engineering and Technology, May 2015
14. "A Review on Off-line Leaf Recognition Using Neural Network", Miss. NeedaSamreenI.Khan, Prof. Rajesh B.Pandhare, International Journal of Computer Science and Mobile Computing, Jan 2015
15. "Leaf recognition and segmentation by using depth image", Shao, Xiaowei, Yun Shi, Wenbing Wu, Peng Yang, Zhongxin Chen, and RyosukeShibasaki, 2014 The Third International Conference on Agro- Geoinformatics, 2014
16. "An Effective Tea Leaf Recognition Algorithm for Plant Classification Using Radial Basis Function Machine", ArunpriyaC., Antony SelvadossThanamani, International Journal of Modern Engineering Research (IJMER), Mar. 2014
17. "Detection and classification of plant leaf diseases", KshitijFulsoundar, TusharKadlag, SanmanBhadale, Pratik Bharvirkar, Prof S.P.Godse, International Journal of Engineering Research and General Science Volume 2, Issue 6, October-November, 2014
18. "Plant Leaf Recognition using Texture and Shape features with Neural Classifiers", JyotismitaChakia, RanjanParekha, and Samar Bhattacharyaa, Pattern Recognition Letters, 2015
19. "Study of Various Techniques for Medicinal Plant Identification", ShyamVijayraoPundkar, Prof. M.M. Waghmare, International Journal on Recent and Innovation Trends in Computing and Communication, November 2014
20. "Plant Leaf Disease Identificationand Prevention by Mobile Vision System", Deore Nikita R., KhadeChhaya A., KotwalPooja R.&PatilKalyani J, Imperial Journal of Interdisciplinary Research (IJIR) , 2016