A Feature Engineering Approach To Improve Plant Species Classification From Leaf Images

Apurv Patel av9patel@uwaterloo.ca
Lakshmi Samhitha lsyadlap@uwaterloo.ca
Jayagn Modh jmodh@uwaterloo.ca
Sravani Subraveti ssubrave@uwaterloo.ca

Abstract— **This paper illustrates the importance and contribution of features in identifying Plant Leaf Species. For identifying various features in a leaf like shape, color and size, the approach is to calculate the accuracy rate for different feature combinations.To implement the idea, we have taken image classification algorithms like Support Vector Machine(SVM)[1], Random Forest(RF)[2],K-Nearest Neighbours(K-NN)[3]etc., To improve the accuracy rate of the Algorithms in identifying the plant leaf species, we have explored other features of a Leaf.One such feature inclusion is about Hu Moments[8], which are Shape Descriptors of the leaf which is one of the focus of this research project. The results obtained by training the algorithms are explained and analysed.**

Keywords: **Image Classification, Plant Leaf Species, Accuracy rate, Feature extraction, Hu Moments**

I. Introduction

One of the most important tasks for researchers, field guides, and others is plant identification, as plants play an important role in everyone's life.It is time consuming to manually collect information on unknown plants, such as going to the local nursery, reading books, or writing on the Internet. Automatic plant identification system is therefore of great importance to facilitate and speed up the process[4].Plants are identified by their organic characteristics, such as flowers, leaves, roots, or the whole plants. After some research, it is understood that leaves is origin of characteristics to identify a plant species[5]. A collection of features must be extracted to use a leaf as a discriminating feature of a plant. It is possible to extract features from a leaf image from its form, colour and texture[4]. Our interests in identifying the leaf species using an image by applying image processing techniques drove us to understand the techniques, the cruxes of leaf features.

The current state of art in Image Classification is Convolution Neural Networks(C-NN)[12]. Many papers that were published illustrated the accuracy rate for identifying the leaf image is quite high using CNN, however, we have studied other Machine Learning Algorithms in this project because they are still the current state of art in the field. Algorithms like Support Vector Machine(SVM)[1],K-Nearest Neighbour(K-NN)[3],Random Forest(RF)[2], Probabilistic Neural Network(PNN)[6] are implemented in this project. We have studied the review project titled "Image Classification Techniques in Plant Leaf Species Recognition" using the above classification algorithms. These Algorithms provide the accuracy percentage (%) using Flavia Dataset[6], LeafImage Dataset from UCI Machine Learning repository[7]. We have analyzed and interpreted the limitations and advantages of each algorithm. The maximum and minimum accuracy rate is observed for SVM[1] and RF algorithms[2] with 91.2% and 88.0% accuracy rate respectively.

We learned that accuracy recognition of a leaf depends on Feature extraction from the image.In our research project, the features extracted by processing an image is obtained from the shape, color, and texture of the leaf. Here the overall features extracted from a leaf image are 17 features. Under these features, the shape has 7 geometrical features, color and texture have 6 and 4 features respectively.Here the shape features consists of area, perimeter, circularity, rectangularity, Physiological Length, Physiological Width, aspect ratio. Colour features are mean_r, mean _g, mean_b, stddev_r, stddev_g, stddec_b. Tex-

ture features are entropy,constrast, correlation, inverse_difference_moments[1] We have used different combinations of these 17 features and trained with the image classifiers. Later, we compared the accuracies achieved on the test dataset by using every possible combination of those extracted features. These obtained accuracies enabled us to identify which feature or combination of features are more important and contribute strongly to classify one leaf species from another species. However, we explored other features that we could train in the SVM Algorithm in Python development environment, we found other features like Hu Moments[8], Zernike moments[10], vein features and other local features like Scale Invariant Feature Transform (SIFT)and Speeded Up Robust Features (SURF)[11]. Out of all these features we studied the Hu Moments which are similar to Shape features in a leaf like area, perimeter, circularity, etc which are called Shape-descriptors for Hu Moments.
Hu moments are robust representation of shape as they are invariant of rotation, reflection,scale, and translation[8]. We calculated Hu moments for every image in the data set and analysed the results.We used the Flavia dataset which consists of 1907 leaf images belonging to 32 species and we also used OpenCV for doing all the image processing tasks and extracting the features.Flavia data is published in Dec'2007[6]

II. Approach for Feature Engineering

For any problem statement to solve, we need to follow few steps to arrive at a result. To understand the "How and What" aspects of the project i.e, which features contribute primarily in classifying a leaf from an image, The Approach for feature engineering is illustrated in this section. We implemented the SVM Algorithm in the project in Python 3 Environment. To achieve the accuracy of the leaf image, pre-processing of the image is required and following are the steps:

- Conversion of RGB to Grayscale image
- Smoothing image using Guassian filter
- Adaptive image thresholding using Otsu's thresholding method
- Closing of holes using Morphological Transformation
- Boundary extraction using contours

After the pre-processing of an image, the following is the approach to calculate the accuracy rate. Firstly, the accuracy rate for each set of features is calculated separately for Shape features, Color Features and Texture Features and recorded the results. Then, we have calculated taking two combinations i.e., Shape and Colour, Shape and texture, Texture and Colour and noted the results Further, we trained the algorithm to take k-sized feature vector where k=5,6,7,.,17 we found every possible combination for a single k i.e., we had $^{17}C_k$ combinations for each value of k and recorded the results.
Hu Moments: Hu Moments are a set of 7 numbers calculated using central moments that are invariant to image transformations. The first 6 moments have been proved to be invariant to translation, scale and rotation and reflection.While the 7th moment's sign changes for image reflection.[8] The block diagram of Hu moments as mentioned below, is the approach to get the 7 Hu Moments of a Leaf Image. Initially, the original Image is converted to Gray scale Image, then it is transformed to a Binary Image using Thresholding technique. Later, Image Moments of the binary image is calculated i.e., weighted averages of Image pixel Intensities. Then, Hu Moments of a leaf image is achieved by implementing the SVM Algorithm. The training of Hu Moments is explained in below diagram.[8] To get the 7 Hu moments the following formulae are used to calculate the numbers[9]:

$$h_0 = \eta_{20} + \eta_{02} \tag{1}$$

$$h_1 = (\eta_{20} - \eta_{02})^2 + 4(\eta_{11})^2 \tag{2}$$

$$h_2 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{3}$$

$$h_3 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{4}$$

$$\begin{aligned} h_4 = {} & (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 \\ & - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})[3(\eta_{30} + \eta_{12})^2 \\ & - (\eta_{21} + \eta_{03})^2] \end{aligned} \tag{5}$$

$$\begin{aligned} h_5 = {} & (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ & + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})] \end{aligned} \tag{6}$$

$$\begin{aligned} h_6 = {} & (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 \\ & - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \\ & \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \end{aligned} \tag{7}$$

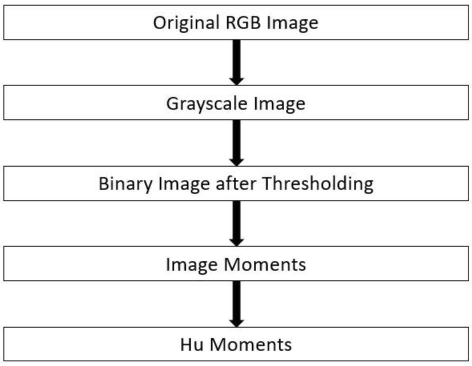

Fig. 1: Block diagram for Hu Moments

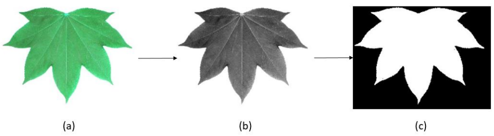

Fig. 2: Binary Image of a Leaf

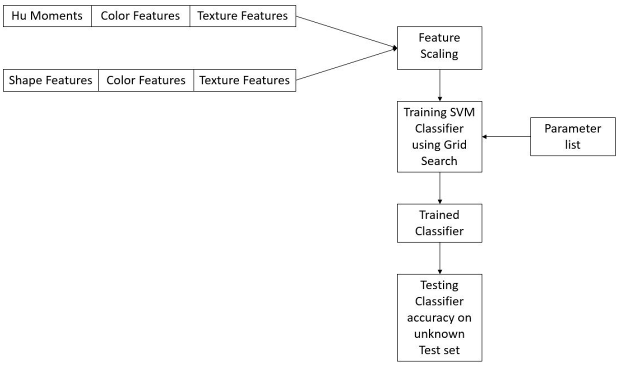

Fig. 3: Training of Hu Moments

Additionally, since the shape and Hu moments are similar, we have removed the Shape features in SVM Algorithm and included the Hu moments without grid search to check the accuracy rate of recognising a leaf image.

Furthermore, by including the grid search for Hu moments as well, along with Colour and texture features, we have recorded the results.

III. Experiment Results & Analysis

Coming to the "Why" aspect of the project, this section explains the experiment results and Analysis. The results obtained from the above explained approach are presented as graphs and tables. For the initial calculation of Accuracy rate for features is presented in a graph i.e., in Fig 1. The graph below illustrates the obtained results by taking combinations of features. The accuracy rate for Shape, Colour and Texture is 89.5% and for Shape and Texture is 89%İt is observed that significant results are obtained for Shape and Texture feature combinations.

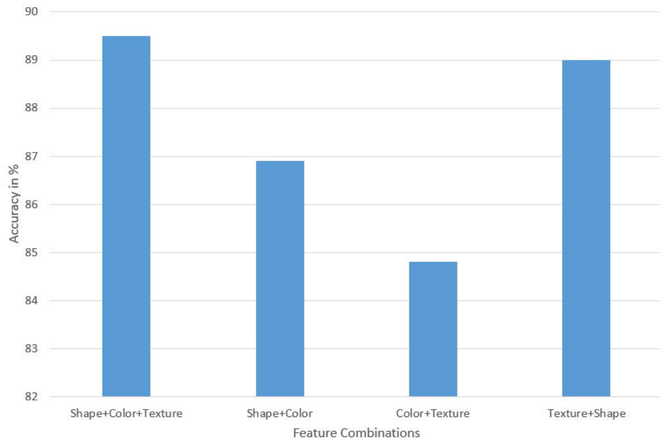

Fig. 4: Graph for Feature Combinations

Accuracy Calculation	
Size of Feature Combination	Maximum Accuracy rate
5	79.4
6	84.9
7	87.7
8	88.7
9	89.7
10	89.5
11	90.5
12	91.0
13	91.8
14	91.8
15	91.6
16	90.4
17	89.5

TABLE I: Accuracy Calculation of Feature Combinations

We can observe that the obtained results in the below table,for 13 features, maximum accuracy rate is recorded. These 13 features are mostly from the combination of Shape and Texture features. Therefore, from the above graph and the results in the table, we can interpret that Shape and Texture plays an important role in identifying Plant Leaf Species from a given Image. Here, Colour Features do not contribute to the classifier performance is because as every leaf has similar colour with minimum difference in the shades.Hence, as the size of feature combinations increases, we can observe from the table that the Maximum Accuracy rate obtained by training the algorithm is decreasing.

As explained in the approach section, Furthermore, We have applied the Hu Moments to the Algorithm and recorded the results. The results portrait that the accuracy rate for Hu Momemts is 93 % , which is an improvement from the previous.

After gathering the results for SVM [1] Algorithm, we have applied the approach to other image classification algorithms like K-NN [3], Naive Bayes, Random Forest [2] and Probabilistic Neural Network [6]. The graph below illustrates the comparison results of Plant leaf Species identification for a given leaf image.

Overall, The Hu Moments for every algorithm gives more accuracy percentage% when compared with Shape, Colour, Texture features.The maximum Accuracy is obtained for Hu Moments which is for SVM and Probabilistic Neural Network(P-NN).

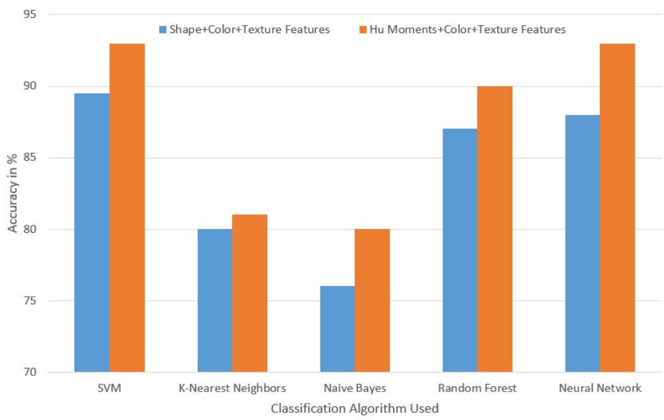

Fig. 5: Comparison Graph for Algorithms with and without Hu Moments

The below table represents the exact results obtained by running the various algorithms with and without Hu Moments.

IV. Conclusion

After performing the experiments and training the classifier on various combinations of extracted features, we got to know from our analysis that the shape and texture features of leaf proves to be the most contributing features to classify plant species. We also found out that using Hu Moments as shape descriptors instead of the geometrical shape features yield better performance in identifying plant species because of its consideration of pixel intensities which proves to be more effective. From our results, we can deduce that there is still a room for improvement in the plant species classification task by focusing more on feature engineering. Further work can be carried out to introduce more meaningful features based on plant texture, veins and leaf margin that can enable us to train a robust classifier with state-of-the-art performance. Even bigger and more diverse data-sets in this classification task could help us to train models that can generalise well and is effective in real-time leaf pictures.

Comparision of Accuracy Percentages for Various Algorithms		
Algorithm	Shape+Color+Texture	HuMoments,Color, Texture
SVM	89.5	93
K-Nearest Neighbors	80	81
Naive Bayes	76	80
Random Forest	87	90
Neural Network	88	93

TABLE II: Accuracy Calculation for Various Algorithms

References

[1] ArunPriya C and Balasaravanan T, 21-23 March 2012. (*International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012)*) An Efficient Leaf Recognition Algorithm for Plant Classification Using Support Vector Machine

[2] Esraa Elhariri, Nashwa El-Bendary and Aboul Ella Hassanien, 26-28 Aug. 2016. (*2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*)Plant classification system based on leaf features

[3] Pullela S V V S R Kumar,Konda Naga Venkateswara Rao and Akella S 26-28 Aug. 2016. (*2016 2nd International Conference on Contemporary Computing and Informatics (ICCCI)*)Leaf classification based on Shape and Edge feature with k-NN Classifier

[4] Aparajita Sahay and Min Chen, 26-28 Aug. 2016. (*2014 9th International Conference on Computer Engineering Systems (ICCES)*) Leaf Analysis for Plant Recognition

[5] Sigit Adinugroho and Yuita Arum Sari,3-5 May 2018(*2016. 2018 5th International Conference on Electrical and Electronic Engineering (ICEEE)*) Leaves Classification Using Neural Network Based on Ensemble Features

[6] Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, Yi-Fan Chan and Qiao-Liang Xiang, (*2007 IEEE International Symposium on Signal Processing and Information Technology*). A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network.

[7] Pedro F.B. Silva, Andre R.S. Marcal, Rubim M. Almeida da Silva. *(2013). Springer Lecture Notes in Computer Science,*

Vol. 7950, 197-204. Evaluation of Features for Leaf Discrimination.

[8] Zhihu Huang, and Jinsong Leng, 2017. (*2010 2nd International Conference on Computer Engineering and Technology*) Analysis of Hu's moment invariants on image scaling and rotation

[9] Mallick, Satya and Krutika Bapat."Shape Matching using Hu Moments (C++/Python)" Learn OpenCV, 10 December 2018, https://www.learnopencv.com/shape-matching-using-hu-moments-c-python/.

[10] B.S.Harish Aditi Hedge, OmPriya Venkatesh, D.G.Spoorthy and D.Sushma. (*2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*) Classification of plant leaves using Morphological features and Zernike moments

[11] Dr. Sandeep Kumar and Samrity (*2015 International Journal of Science and Research (IJSR)*) Plant Species Identification using SIFT and SURF Technique

[12] Wang-Su Jeon and SamSang-Yong Rheer (*International Journal of Fuzzy Logic and Intelligent SystemsVol. 17, No. 1, March 2017, pp. 26-34*) Plant Leaf Recognition Using a Convolution Neural Network