PERFORMANCE ANALYSIS OF LEAF IMAGE CLASSIFICATION USING MACHINE LEARNING ALGORITHMS ON DIFFERENT DATASETS

Dr. V. Rajesh¹, Bhanu Prakash Dudi²

¹ Professor, Dept. of ECE, KL Deemed to be University, Vaddeswaram, Guntur, Andra Pradesh
² Research Scholar, Dept. of ECE, KL Deemed to be University, Vaddeswaram, Guntur, Andra Pradesh

ABSTRACT

Differentiating and understanding various plants species is one of the most significance for keeping up biodiversity so that they will be able to recognize it by using its leaf image features. Leaf image classification is a pivotal errand since leaves are showing high intra class variety in the leaf features. There is a need to build a classifier to recognize the plants adequately. In this paper, we have taken five distinct classifiers specifically Random Forest, Naive bayes, Support Vector Machines, K-Nearest Neighbors, and Logistic Regression. To assess the performance of these classifiers we tested on the three publicly accessible datasets, for example, Flavia, Swedish, and Folio. The performance comparison of these models are done and implemented by using python language in Google Colab environment. From the experimental results it is shown that Random Forest and Logistic Regression models gives accuracy of 84%.

Keywords: Leaf features, Flavia dataset, Swedish dataset, Folio dataset, Google Colab, Accuracy;

I. INTRODUCTION

Plants are one of the essential segments of the earth answerable for ensuring the world’s current circumstance [1]. They give food, air, medicines, fuel and shelter for animals and birds .Nevertheless, starting late a steadily expanding number of plants are at the very edge of the end due to constant de-forestation [2] [3]. Plants can be known by their external features such as the kind of plant, its natural surroundings, anomaly of root, stem, leaves, flowers and seeds.[4] In the field of ayurved, Medicine and Pharmacy it is essentially a easy tool to identify plants on their external features, in automated system manner [5]. A long list of external features of plants can be prepared but usually only few important features are enough to know most of the plants. In this context leaf of the plant plays important role to know the plant type. Leaf of the plants has many external features to distinguishing the plants such as leaf arrangement, types of venation, various shapes of leaf blade, types of leaf margins [5].

II. RELATED WORK

Sandeep Kumar E., et all [6],center around the plant identification system and finding the plant ID utilizing leaf features. In this paper, the authors proposed a remarkable plant identifier is registered and this identifier calculation is tried on ten diverse medicinal plants for precise ID. This paper utilizes an alternate leaf features to ascertain the plant ID number for each plant utilizing the weighted averaging method.

Pradeep Kumar., et all [7], projected a leaf recognition system employing orthogonal moments as shape descriptors, Histogram of Oriented Gradients (HOG) and gabor features as textures descriptors. A Support Vector Machine classifier is employed for classification and recognition of leaf images.

Petre Lameke et.all [8], projected a novel descriptor based on the angles between points of the leaf contour classification. Mei-Fem Bong et.all [9], Projected a absolutely unique approach to cluster the plants on their lobes, sinuses and margin. Finally all the boundary points and centre point. The quantity of peak and valley is calculated to cluster the plants in step with the rule- based method.

H.X.Kan et.all [10], proposed an automatic classification techniques dependent on leaf images of medicinal plants. This paper centres around ten shape features and five texture features for identifying of plants utilizing Support Vector Machine classifier. Naresh Y.G. et all. [11] proposed the classification of plant leaves
dependent on texture features. The Modified Local Binary Pattern (MLBP) is proposed to extract texture features from plant leaves. The nearest neighbor classifier is utilized for classification.

Jyotismita Chaki et al. [12] proposed plant classification based on combining the leaf texture and shape features. For getting texture features authors used Gabor filter as gray level co-occurrence matrix (GLCM) while shape of the leaf is captured using set of curvelet transforms. A Neuro-Fuzzy Controller (NFS) and Feed forward back propagation multi layer perceptron (MLP) are utilized as classifiers. Huisi Wu et al. [13], proposed binary gabor pattern and Extreme Learning Machine for Automatic plant leaf recognition. These methods require only one parameter for tuning the classifier while others require more.

The rest of the paper is structured as follows: Section 3 describes about datasets, machine learning models and proposed system description and experimental results and section 4 concludes the paper with future scope.

III. MATERIALS AND METHODS

3.1 Dataset

In this work the accompanying datasets have been considered to evaluate the performance of the distinctive ML models.

a. Flavia leaf image dataset: It is a publicly accessible dataset, made by Stephen Gang Wu, Forrest sheng Bao [14]. This dataset comprises top-notch leaf pictures with a white background. A sum of 1907 images were available in the dataset has a place with 33 plant species. A large portion of the Flavia dataset is gathered from basic plants in the Yangtze delta, China.

b. Swedish leaf image dataset: The Swedish leaf image dataset has images of 15 types with 75 pictures for every species. An aggregate of 1125 pictures are accessible in the dataset. This dataset made by Oskar Soderkvist [15], all leaf pictures are gathered from Swedish trees.

c. Folio leaf image dataset: This dataset built with 32 distinctive leaves, each plant has 20 images [16]. The leaves are gathered from the University of Mauritius, Mauritius. The images consist of a white background. The pictures were taken in sunshine conditions.

3.2 Machine Learning Models

In this work the following models are considered for the study on leaf image datasets [17].

i. Random Forest Algorithm: It uses the tree based decision algorithm. It eliminates over fitting problems by creating multiple decision trees and let them vote on how to classify inputs.

ii. Naive Bayes Algorithm: This algorithm is used to classify the objects with very high-dimensional features. This algorithm is very fast and simple it consists of tuneable parameters. This algorithm uses bayes theorem to classify the objects.

iii. Support Vector Machine Algorithm: It is a general straight forward supervised machine utilized for classification. Fundamentally, SVM finds a hyper-plane that makes a limit between the various classes. In SVM, we plot every information attribute in an N-diemensional space, where N is the number of features in the data. Next, finding the optimal hyper plane to separate the data [18].

iv. K-nearest neighbors: This algorithm works on principle that nearest points have the same label and all features are equally important [18].

v. Logistic Regression: This method is a variation of the linear regression, explicit for the circumstance when the needy variable is binary. Rather than straightforwardly fitting the needy variable, its likelihood of event is utilized to make a decision threshold.

3.3 System Description
Figure 1 shows the block diagram of the proposed framework in which images are labeled and preprocessed to boost the performance of the model. Leaf features such as margin shape, color, and margin texture are important for classification tasks. The feature extraction can be performed by contour-based extraction. It is articulated in terms of leaf length, width, aspect ratio, area, rectangularity, circularity, and perimeter. These feature extraction is implemented in Python OpenCV library. Leaf color features are extracted as mean and standard deviation of RGB channels. Texture features are determined using Mahotas framework [19]. It is a computer vision and image processing library for Python. The extracted features of each image forms a dataset file for machine learning model to train and test the model. Performance evaluation of the model is done using Accuracy, Precision, Recall, and F1-Measure [20]. The table I shows different features extracted for the proposed work.

### Table I: Leaf Image Features

<table>
<thead>
<tr>
<th>SI NO</th>
<th>Feature of Leaf</th>
<th>Description of the feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Area (A)</td>
<td>Number of pixels in region of the leaf.</td>
</tr>
<tr>
<td>2</td>
<td>Perimeter (P)</td>
<td>Summation of the distances between each connecting pair of pixels around the border of the leaf.</td>
</tr>
<tr>
<td>3</td>
<td>Physical Length (L)</td>
<td>A line segment associating the base and the tip of the leaf.</td>
</tr>
<tr>
<td>4</td>
<td>Physical Width (W)</td>
<td>The most extreme width is opposite to the actual length.</td>
</tr>
<tr>
<td>5</td>
<td>Aspect ratio (AR)</td>
<td>Proportion of the length to width A = L/W</td>
</tr>
<tr>
<td>6</td>
<td>Rectangularity (Rect)</td>
<td>Rect = A_{Leaf}/A_{Boundary}</td>
</tr>
<tr>
<td>7</td>
<td>Circularity (Cir)</td>
<td>Delineates the distinction between a leaf shape and a circle. Cir = 4πA/P^2</td>
</tr>
<tr>
<td>8</td>
<td>Mean</td>
<td>It is a angular second moment.</td>
</tr>
<tr>
<td>10</td>
<td>Contrast</td>
<td>It is the standard deviation of the intensity image.</td>
</tr>
<tr>
<td>11</td>
<td>Correlation</td>
<td>It is the measure of the intensity randomness.</td>
</tr>
<tr>
<td>12</td>
<td>Entropy</td>
<td>A measure of the intensity randomness.</td>
</tr>
</tbody>
</table>
Inverse di
ference_ moments

\[
\sum_{i=0}^{N-1} (P_i)(P_0) (P_1)
\]

\[
\sum_{i} \sum_{j} \frac{1}{1 + (i-j)} P(i,j)
\]

3.4 Experimental Results:

The machine learning pipeline model for all explained above models are implemented using python language in Google Colab environment and models are trained and tested on three publically available leaf image datasets (flavia, Swedish, and Folio).

In our experiment, 70% of the data are used for training and 30% of the data for testing purpose. Accuracy, precision, recall and F-1 measure are calculated as performance factor for each classifier [21].

Experimental results from the five classifiers namely, Random Forest, Naïve Bayes, Support Vector Machine, K-nearest neighbors, Logistic Regression for the flavia, Swedish and Folio leaf image dataset are tabulated as shown in Table II, III and IV correspondingly.

From Table II it is observed that Logistic Regression and Random Forest classifier are works well on Flavia leaf image dataset. We achieved accuracy of 84.11% with Random Forest and accuracy of 84.16% with Logistic Regression. Figure 2 shows the bar graph of classifiers performance on Flavia dataset.

From Table III it is seen that Logistic Regression and Random Forest is functioning admirably on Swedish leaf dataset. We accomplished an accuracy of 84.61% with the Random forest and an accuracy of 84.02% with the Logistic regression. Figure 3 shows the reference chart of classifier performance on Swedish dataset.

We achieved accuracy of 84.04% with Random forest and accuracy of 77.65% with the Logistic regression on Folio image dataset. Figure 4 shows the bar graph of classifier performance on Folio image dataset and Table IV gives accuracy, recall, precision and f1-measure for five classifiers.

### Table II: Performance analysis on Flavia Leaf Image dataset

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy in %</th>
<th>Precision in %</th>
<th>Recall in %</th>
<th>F1- Measure in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest Algorithm</td>
<td>84.11</td>
<td>85.25</td>
<td>84.11</td>
<td>84.23</td>
</tr>
<tr>
<td>Support Vector Machines Algorithm</td>
<td>79.05</td>
<td>82.04</td>
<td>79.05</td>
<td>79.36</td>
</tr>
<tr>
<td>Logistic regression Algorithm</td>
<td>84.11</td>
<td>85.87</td>
<td>84.11</td>
<td>84.26</td>
</tr>
<tr>
<td>K-nearest neighbours Algorithm</td>
<td>80.10</td>
<td>81.39</td>
<td>80.10</td>
<td>80.04</td>
</tr>
<tr>
<td>Naïve Bayes Algorithm</td>
<td>72.25</td>
<td>75.14</td>
<td>72.25</td>
<td>72.91</td>
</tr>
</tbody>
</table>

### Table III: Performance analysis on Swedish Leaf Image dataset

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy in %</th>
<th>Precision in %</th>
<th>Recall in %</th>
<th>F1- Measure in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest Algorithm</td>
<td>84.61</td>
<td>85.56</td>
<td>84.61</td>
<td>84.54</td>
</tr>
<tr>
<td>Support Vector Machines Algorithm</td>
<td>79.28</td>
<td>82.04</td>
<td>79.28</td>
<td>79.83</td>
</tr>
<tr>
<td>Logistic regression Algorithm</td>
<td>84.02</td>
<td>85.45</td>
<td>84.02</td>
<td>83.57</td>
</tr>
<tr>
<td>K-nearest neighbours Algorithm</td>
<td>76.03</td>
<td>76.16</td>
<td>76.03</td>
<td>75.29</td>
</tr>
<tr>
<td>Naïve Bayes Algorithm</td>
<td>73.07</td>
<td>75.36</td>
<td>73.07</td>
<td>73.57</td>
</tr>
</tbody>
</table>
### Table IV: Performance analysis on Folio Leaf Image dataset

<table>
<thead>
<tr>
<th></th>
<th>Random Forest Algorithm</th>
<th>Support Vector Machines Algorithm</th>
<th>Logistic regression Algorithm</th>
<th>K-nearest neighbours Algorithm</th>
<th>Naïve Bayes Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy in %</td>
<td>84.04</td>
<td>60.63</td>
<td>77.65</td>
<td>70.21</td>
<td>72.34</td>
</tr>
<tr>
<td>Precision in %</td>
<td>87.53</td>
<td>82.56</td>
<td>84.22</td>
<td>79.09</td>
<td>76.06</td>
</tr>
<tr>
<td>Recall in %</td>
<td>84.04</td>
<td>60.63</td>
<td>77.65</td>
<td>70.21</td>
<td>72.34</td>
</tr>
<tr>
<td>F1-Measure in %</td>
<td>84.45</td>
<td>65.03</td>
<td>78.45</td>
<td>71.68</td>
<td>73.20</td>
</tr>
</tbody>
</table>

![Classifier performance on Flavia leaf dataset](image1)

Figure 3: Classifier performance on Flavia dataset

![Classifier performance on Swedish dataset](image2)

Figure 4: Classifier performance on Swedish leaf dataset
From the examination, it is evident that no single model can be considered to be the best in all cases. Among every one of these classifiers, Random forest and Logistic regression are having more steady outcomes. In the future work, accuracy of the classifier can be increased by using feature selection methods and automating the feature extraction by using auto encoders and Convolution Neural Network (CNN) to classify the leaf images effectively.

REFERENCES