International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 3 ǁ 2022 ǁ PP. 39-43

Greenery Classification Using Machine Learning

Kush Patel [1]
Department of Computer Science CHARUSAT UNIVERSITY M.TECH
Changa, Anand, Gujarat

Dhaval Bhoi [2]
Assistant Professor
U & P U. Department of Computer Engineering CHARUSAT UNIVERSITY
Changa, Anand, Gujarat

Riddhi Shukla [3]
Megha Shah[4]
Mogar, Anand, Gujarat

Abstract—*There are many methods to identify plants have been proposed by several researchers. Commonly, the methods did not capture margin ,texture and shape information,because this factor play important role for classification. This paper presents a good method for plant species classification using train dataset. The various kind of supervise classification algorithm are apply to same dataset, also check accuracy and log loss. Linear discriminate algorithm gives high accuracy and lower logloss which is approximately more than 95% and 0.86 respectively. This algorithm are classify 99 kind of leaf species.*

Keywords: *classification,dataset,accuracy,logloss, visuliza- tion,texture,margin,shape,supervise learning*

Date of Submission: 28-02-2022 Date of acceptance: 09-03-2022

I. INTRODUCTION

Plant is very useful source for human to living on earth. Plants maintain the stability of oxygen and carbon dioxide of earth's atmosphere. The association between plants and human beings are also very close. Moreover, plants are powerful means of livelihood and production of human beings and around half a million species of plant in the world.

The system providing a one new data science project agriculture field user. It provides machine learning based communication for students who study in agriculture field. We will develop a project on machine learning. In our project we have use the dataset and this dataset are split into two parts train and test. In dataset that contain 99 leaf species , each contain 10 sample so 990 rows, 64 attribute of texture ,64 attribute of shape and 64 attribute of margin available for each samples.

After splitting dataset, we will create model. Using our project know the status about leaf belong to which species. Automating plant recognition might have many applications, including: Species population tracking and preservation, Plant- based medicinal research, Crop and food supply management. This project is useful in real life scenario and it will be helpful to many botanist and students that will be study in agriculture field and form our career perspectives this field have wide opportunity.

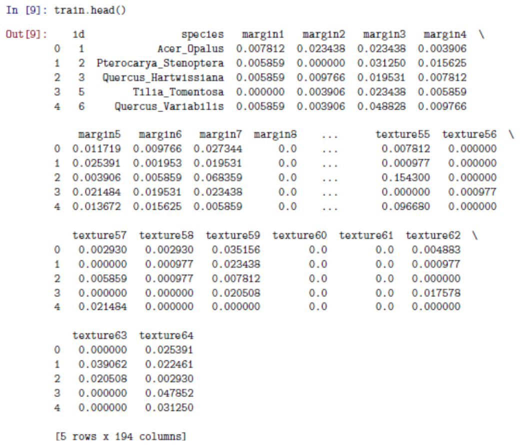
II. RELATED WORK

Lately, plant classification became one of major re- searches. Shanwen et al.[6] used a combination between semi-

Supervised locally linear embedding (semi-SLLE) and KNN algorithms for plant classification based on leaf images and showed its performance. James Cope et al.[4] presented plant texture classification using Gabor co-occurrences; where joint distributions for the responses from applying different scales of the Gabor filter are calculated. The difference among leaf textures is calculated by the Jeffrey divergence measure of corresponding distributions. Also Kadir et al. in [5] incorpo- rates shape and vein, colour, and texture features to classify leaves using probabilistic neural network and proves that it gives better result with average accuracy of 93.75%. Plant leaf images corresponding to three plant types, are analysed using two different shape modelling techniques in Chaki et al.[3], authors proposed an automated system for recognizing plant species based on leaf images. The author Bhardwaj in[2], that presented a simple computational method in computer vision to recognize plant leaves and to classify it using Knearest neighbours. Anang Hudaya also worked on plant classification in his paper[1], presenting a scalable approach for classifying plant leaves using the 2-dimensional shape feature, using distributed hierarchical graph neuron (DHGN) for pattern recognition and k-nearest neighbours (kNN) for pattern classification.

III. DATASETS

The ’Leaves’ dataset contains 99 various kind of species of leaves each species represented by three 64 element vector for each of three core features collecting from particular leaf i.e texture ,shape and margin. This dataset contains 10 samples of each species as shown in 1.



```
In [9]: train.head()

Out[9]:   id                species   margin1   margin2   margin3   margin4  \
       0   1            Acer_Opalus  0.007812  0.023438  0.023438  0.003906
       1   2  Pterocarya_Stenoptera  0.005859  0.000000  0.031250  0.015625
       2   3   Quercus_Hartwissiana  0.005859  0.009766  0.019531  0.007812
       3   5        Tilia_Tomentosa  0.000000  0.003906  0.023438  0.005859
       4   6     Quercus_Variabilis  0.005859  0.003906  0.048828  0.009766

          margin5   margin6   margin7  margin8    ...      texture55  texture56  \
       0  0.011719  0.009766  0.027344      0.0    ...       0.007812   0.000000
       1  0.025391  0.001953  0.019531      0.0    ...       0.000977   0.000000
       2  0.003906  0.005859  0.068359      0.0    ...       0.154300   0.000000
       3  0.021484  0.019531  0.023438      0.0    ...       0.000000   0.000977
       4  0.013672  0.015625  0.005859      0.0    ...       0.096680   0.000000

         texture57  texture58  texture59  texture60  texture61  texture62  \
       0  0.002930   0.002930   0.035156        0.0        0.0   0.004883
       1  0.000000   0.000977   0.023438        0.0        0.0   0.000977
       2  0.005859   0.000977   0.007812        0.0        0.0   0.000000
       3  0.000000   0.000000   0.020508        0.0        0.0   0.017578
       4  0.021484   0.000000   0.000000        0.0        0.0   0.000000

         texture63  texture64
       0  0.000000   0.025391
       1  0.039062   0.022461
       2  0.020508   0.002930
       3  0.000000   0.047852
       4  0.000000   0.031250

       [5 rows x 194 columns]
```

Fig. 1. Leaf dataset for classification

The above dataset shows top five rows and all columns out of 990 rows. In this dataset each species gives the unique id, this is used for classification result.

I. PROPOSED APPROACHES

The present work shows a comparison of classification of 99 different species of plant leaves using three features from the leaf dataset; Fig 2 shows the architecture of proposed approaches:

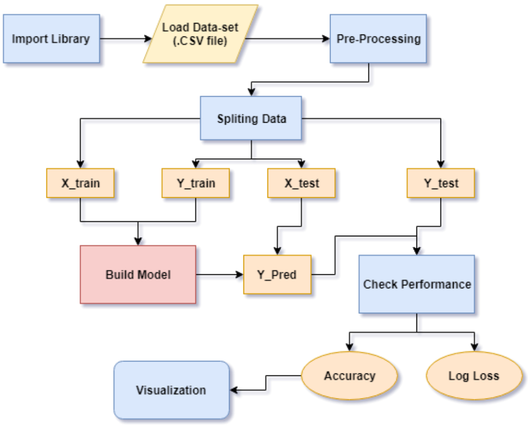


Fig. 2. The architecture of proposed approaches

A. Import library and load dataset

Firstly, import some library which can used entire process of classification. we import numpy for Ndimansional array processing ,panda for data structure and analysis, matplotlib for 2D /3D plotting,seaborn for data visualization also import sklearn library that provide range of supervise classification algorithm via consistent interface in python. After that load dataset for build model that can be used for classification purpose.

B. Pre-processing data

The real word data is noisy and inconsistent. Inconsistent is occur when similar data is kept in different format. So we can perform the cleaning prepare and manipulation Data. If data is consistent format and clean then data is used for further task otherwise can’t used for further task. In our dataset we have encode species column using label encoder and convert into numeric form.

C. Splitting dataset

In this section our dataset is splitting into two parts namely training and testing. The training part contain 80% data and testing part contain 20% data out of whole dataset. This splitting are done using stratified shuffle split because of this split does not face problem of over fitting. In this section decide dependent and independent variable namely X and Y respectively. The X contain 64 attribute of texture ,shape and margin ,all are dependent on label of species. The Y contain label of species which are independent of other variable.

D. Build model

In this phase we have build model using training data as well dependent and independent variable of our data set. We apply various kind of supervised machine learning classification algorithms to same dataset. In each case, we used three differ- ent approaches for classification: linear discriminate analysis , K neighbour classifier, support vector classifier, numeric support vector classifier, decision tree classifier ,random forest ,adaptive boosting, gradient boosting, gusssian NB, quadratic discriminate analysis

E. Check Performance

The performance can be evaluated based on their accuracy and log loss.If the accuracy will be more and log loss is less then algorithm is best otherwise does not efficient for this kind of dataset.

1) Accuracy : The accuracy can define as ratio of number of correct prediction to total number of sample . A higher accuracy value means better predictions. Accuracy = Number of correct prediction / Total number of available sample

2) Logloss : Log Loss is classification metric based on probabilities. It is hard to interpret raw log-loss values, but log-loss is still a good metric for comparing models. A lower log-loss value means better predictions.

IV. RESULTS AND DISCUSSION

The following section shows the different results obtained for various kind of classification algorithm

Sr. No.	Algorithms	Accuracy	Logloss
1	Linear Discriminate Analysis	0.984848	1.0138
2	Kneighbors Classifer	0.848485	1.7042
3	NuSVC	0.848485	2.5285
4	Decision Tree Classifier	0.641414	11.1640
5	Random Forest Classifier	0.893939	0.9568
6	SVC	0.787879	4.6107
7	Adaboost Classifier	0.035354	4.3949
8	Gradient boosting Classifier	0.575758	3.0122
9	Gaussian NB	0.580808	14.1294
10	Quadratic Discriminate Analysis	0.020202	34.0154

TABLE I
COMPARATIVE ACCURACY & LOGLOSS OF GREENERY CLASSIFICATION

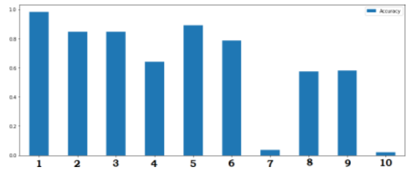
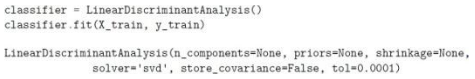


Fig. 3. comparative visualization of accuracy

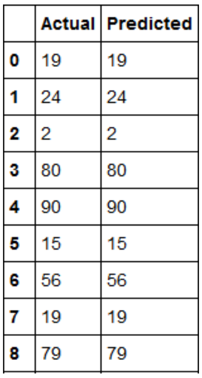
Table 1 and Figure 3 show the obtained results of the classification of species represented by the Margin, texture and shape information. The table 1 prove the linear discrim- inate algorithm is best for leaf dataset this algorithm is give approximate 98% accuracy compare than other algorithm. Additionally, adaptive boosting classifier give less accuracy i.e 3%



```
classifier = LinearDiscriminantAnalysis()
classifier.fit(X_train, y_train)

LinearDiscriminantAnalysis(n_components=None, priors=None, shrinkage=None,
              solver='svd', store_covariance=False, tol=0.0001)
```

Fig. 4. parameters of LDA



	Actual	Predicted
0	19	19
1	24	24
2	2	2
3	80	80
4	90	90
5	15	15
6	56	56
7	19	19
8	79	79

Fig. 5. Result of LDA algorithm based on their id

estimation in SVD solver. The figure 5 shows the result of LDA algorithm for actual and predict by their unique id. The figure 6 shows visualization output graph of LDA algorithm that display fluction between actual and predicted values. The adaptive boosting algorithm gives less accuracy and more log loss , the rule of this algorithm is to convert weak dataset into strong dataset , applying this rule in our dataset that time face problem of over fitting and performance will be decrease.

V. CONCLUSION

Plants play an important role in our routine life, without plants there will not be the existence of the ecology of the earth and human life. The variety 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 leaf. This work proposed a comparative study of supervised classification of plant leaves, where we used to represent ten classification algorithm and this classification is totally based on the dataset using texture, shape and margin information. Overall, machine learning techniques are being widely used to solve real-world problems by storing, manipulating, extracting and retrieving data from large sources. Various kind of supervised machine learning classification algorithms are applied and compare their accuracy as well as log-loss for same dataset. Base on comparison Linear Discriminate Analysis algorithm is suitable for this kind of data. This all algorithm are classified 99 various kind of leave species based on their core feature i.e shape, texture and margin information available in our dataset.

ACKNOWLEDGMENT

The authors would like to thank the all people for their comments and suggestions to improve this paper. We also wish to express our sincere thanks to all faculty members of the Computer Science Department for their Cooperation and valuable suggestions in each stage, which lead to proceed of our work.

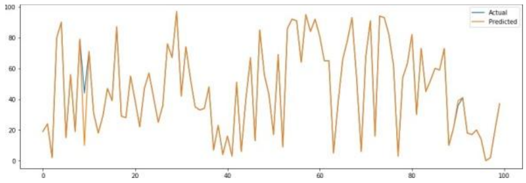


Fig. 6. Visulization output of LDA algorithm

The linear discriminate algorithm is best because our dataset is multi class in which 99 different class are available, this algorithm perform dimensional reduction with higher attribute. The fig 4 shows the paramatars of LDA algorithm number of component is used for dimensional reduction, prior is default none, solver is two types 'svd' singular value decomposi- tion not compute covariance matrix ,so store covariance and shrinkage is false. Another solver is 'lsqr' least square solution that use shrinkage. The tol means threshold used for rank

REFERENCES

[1]. Anang Hudaya Muhamad Amin and Asad̄I Khan. One-shot classification of 2-d leaf shapes using distributed hierarchical graph neuron (dhgn) scheme with k-nn classifier. *Procedia Computer Science*, pages 84–96, 2013.

[2]. Anant Bhardwaj, Manpreet Kaur, and Anupam Kumar. Recognition of plants by leaf image using moment invariant and texture analysis. *International Journal of Innovation and Applied Studies*, 3(1):237–248, 2013.

[3]. Jyotismita Chaki and Ranjan Parekh. Plant leaf recognition using shape based features and neural network classifiers. *International Journal of Advanced Computer Science and Applications*, 2(10), 2011.

[4]. James S Cope, Paolo Remagnino, Sarah Barman, and Paul Wilkin. Plant texture classification using gabor co-occurrences. In

International Symposium on Visual Computing, pages 669–677. Springer, 2010.

[5]. Abdul Kadir, Lukito Edi Nugroho, Adhi Susanto, and Paulus Insap Santosa. Leaf classification using shape, color, and texture features. *arXiv preprint arXiv:1401.4447*, 2013.

[6]. Shan-Wen Zhang and Jing Liu. Weighted locally linear embedding for plant leaf visualization. In *International Conference on Intelligent Computing*, pages 52–58. Springer, 2010.