ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 8, Issue 1, January 2019

Distinguishing of Rice Varieties by Using Machine Learning Models

Puneet Dheer

Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India

Abstract: A large number of studies have been executed for classifying plant types and identifying diseases of various crops particularly using images. The plant type identification problem is further complicated by common object recognition difficulties mainly due to light, pose and orientation. The present study was undertaken to distinguish the four different Indian rice varieties by utilizing the respective collected features and applying machine learning methods. There are various methods available from simple to complex model. However, the study was carried out with simple models like Linear Discriminant Analysis, Logistic Regression, K-Nearest Neighbours and Naïve-Bayes method. K-NN out performed over the other methods with an Accuracy and Precision of 99.16% and 99% respectively.

Keywords: Distinctness, Rice Varieties, Machine Learning, K-NN

I. INTRODUCTION

Rice (Oryza sativa L.) is an essential staple food for more than half of the global population. The classification of rice varieties into a specific class is the most important interest for domain of specific professionals. Typically to distinguish the different varieties requires various sampling by inspection on the agricultural field by skilled workers. Rice grains of different varieties can be mixed during the cultivation, harvesting, and processing may reduce the quality of the products. However, the existence of a large amount of various varieties makes it quite difficult to analyze and classify it by the novice worker. The varieties can be distinguished by recognizing their plant height, panicle density, grain types and color etc.

Genetic marker-based methods have been applied for identifying rice varieties. Steele et al. [1] Selected insertion and deletion markers to distinguish Basmati rice grains from some other fragrant rice varieties. Cirillo et al [2] Applied Random Amplified Polymorphic DNA (RAPD) approach to fingerprint rice grains of 13 Italian accessions. Becerra et al [3] determined the genetic variability of certain Chilean and foreign commercial rice cultivars using Simple Sequence Repeat (SSR) markers. In another study, Chuang et al. [4] reported using SSR markers to distinguish 36 varieties of rice grains from different countries. These genetic marker-based methods are accurate but often time-taking and/or costly for the real time applications.

Image-based methods, Hobson et al [5] proposed image processing techniques for identifying the different varieties of rice based on their size, shape and color. Camelo-Méndez et al [6] characterized the rice grains of 9 Mexican cultivars by performing Principle Component Analysis (PCA) and hierarchical analysis. Kong et al [7] classified the rice seeds of 4 accessions using a near-infrared hyperspectral imaging system and various machine learning algorithms. Image based results of these studies have been promising, they have included a limited number of varieties for discrimination, requires high end imaging processing techniques and the respective test dataset that makes the method too costly and not frequently available to the consumer.

The current study aimed to differentiate the rice grains of 4 Indian varieties using machine learning methods. The following were the specific objectives of the study: (1) Pre-processing the acquired data; (2) Different machine learning classifiers are evaluated including Linear Discriminant Analysis, Logistic Regression, K-Nearest Neighbors and Naïve Bayes method (3) Tested on the selected model after cross validation.

II. MATERIALS AND METHODS

A. Sample Collection and Pre-processing

The present investigation embodied four Indian promising varieties namely, Sarjoo 52, Taraori Basmati, Kasturi and Jal Lahari. One hundred random samples comprising six features of each variety were acquired during field inspection. All the collected data were further divided into training and testing data set in 70:30 ratio and then normalized. All these

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 8, Issue 1, January 2019

samples were collected when each variety reached their respective optimum stages. Six different features selected are: plant height, number of effective tillers, panicle length, number of grains/panicle, grain length and grain breadth.

B. Models used

- K-Nearest Neighbors: The K-NN classifies test sample based on the majority of its K-Nearest Neighbors with minimum distance signifies most common attributes. Here, K distance was selected as 20 after cross validation (Duda et al. [8]; Bishop [9]).
- Naïve Bayes Classifier: The Naïve Bayes is a statistical classifier (Mitchell) [10] which is based on Bayes theorem. This method predicts probabilities of a given samples belonging to a specific class, which means that it provides the probability of occurrence of a given sample or data points within a particular class.
- Fisher's Linear Discriminant Analysis: LDA determines the discriminant dimension in response-pattern space, on which the ratio of between-class over within-class variance of the data is maximized (Duda et al. [8]; Bishop [9]).
- Logistic Regression: A traditional statistical procedure, separates two classes by an S-shaped discriminant function through the decision space (Agresti [11]).

C. Evaluation measure

The Accuracy of distinctness of the rice varieties under study has been computed using the following expression which uses numerical details of correctly classified class from total samples of rice in the dataset.

$$\text{Accuracy} = \frac{\text{no. of correctly indentified samples}}{\text{total no. of samples}} * 100$$

The Precision and Recall are also the important measure to consider for system evaluations which are calculated as follows:

$$\text{Precision} = \frac{\sum \text{True Positive}}{\sum \text{Predicted Condition Positive}} * 100$$

$$\text{Recall} = \frac{\sum \text{ True Positive}}{\sum \text{ Condition Positive}} * 100$$

III. RESULTS AND DISCUSSION

The different machine learning models were tested on the self-collected dataset of four different rice varieties with 100 samples each. Each sample accompanied with six features. These data were trained and tested for four different classifiers (K-NN, LR, LDA, NB). It applied 10-fold cross-validation and selected the suitable model for further classification on unseen test data set. The Table 1 shows that the Precision and Recall results for all varieties only with a K-Nearest Neighbors classifier (K-NN). The confusion matrix Table 2 contains the number of correctly and incorrectly classified grains against every variety and then the Accuracy score, Precision and Recall values calculated for analysis of the selected model. Here, incorrect classification contains both false positive and false negative test samples and correct classification includes all true positive and true negative values after application of K-NN and selected features. K-NN classifier gives an Accuracy of 99.16%. The Precision and Recall on test dataset are 99% and 99%, respectively. The above-mentioned other classifiers show different accuracies in comparison to each other and K-NN outperforms all others for rice distinctness.

Table 1: Precision and Recall of Rice Varieties Under K-NN Model

No.	Varieties	Precision	Recall
1	**Jal Lahari**	100%	100%
2	**Kasturi**	97%	100%
3	**Sarjoo-52**	100%	100%
4	**Taraori Basmati**	100%	97%

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 8, Issue 1, January 2019

Table 2: Confusion Matrix of Rice Varieties Under K-NN Model.

	Predicted				
Actual	**Sarjoo-52**	**26**	**0**	**0**	**0**
	Taraori Basmati	**0**	**31**	**0**	**0**
	Jal Lahari	**0**	**0**	**33**	**0**
	Kasturi	**0**	**1**	**0**	**29**
		Sarjoo-52	**Taraori Basmati**	**Jal Lahari**	**Kasturi**

CONCLUSION

In this study, previous work of different approaches was presented by their pros and cons. The results obtained based on employing feature normalization and K-NN model are quite promising, which suggest that this method can provide an accurate solution to the rice varieties for their distinguishing and/or identification problem alternative to sophisticated image segmentation techniques as discussed earlier. Future work is being more focused towards our self-collected dataset to include more rice varieties to address automatic identification of rice varieties in particular and other field crops varieties in general.

REFERENCES

[1]. Steele, K.A., Ogden, R., McEwing, R., Briggs, H., Gorham, J., 2008. InDel markers distinguish Basmatis from other fragrant rice varieties. Field Crops Res. 105, 81–87.
[2]. Cirillo, A., Del Gaudio, S., Di Bernardo, G., Galderisi, U., Cascino, A., Cipollaro, M., 2009. Molecular characterization of Italian rice cultivars. Eur. Food Res. Technol. 228, 875–881.
[3]. Becerra, V., Paredes, M., Gutiérrez, E., Rojo, C., 2015. Genetic diversity, identification, and certification of Chilean rice varieties using molecular markers. Chilean J. Agric. Res. 75, 267–274.
[4]. Chuang, H., Lur, H., Hwu, K., Chang, M., 2011. Authentication of domestic Taiwan rice varieties based on fingerprinting analysis of microsatellite DNA markers. Botanical Stud. 52, 393–405.
[5]. D. M. Hobson, R. M. Carter and Y. Yan, "Characterisation and Identification of Rice Grains through Digital Image Analysis," 2007 IEEE Instrumentation & Measurement Technology Conference IMTC 2007, Warsaw, 2007, pp. 1-5.
[6]. Camelo-Méndez, G.A., Camacho-Díaz, B.H., del Villar-Martínez, A.A., Arenas-Ocampo, M.L., Bello-Pérez, L.A., Jiménez-Aparicio, A.R., 2012. Digital image analysis of diverse Mexican rice cultivars. J. Sci. Food Agric. 92, 2709–2714.
[7]. Kong, W., Zhang, C., Liu, F., Nie, P., He, Y., 2013. Rice seed cultivar identification using near-infrared hyperspectral imaging and multivariate data analysis. Sensors 13 (7), 8916–8927.
[8]. Duda, R.O., Hart, P., Stork, D.G., 2000. Pattern Classification 2nd ed. John Wiley and Sons, New York.
[9]. Bishop, C.M., 2007. Pattern Recognition and Machine Learning. Springer, New York.
[10]. Mitchell, T., 1997. Machine Learning. McGraw Hill, NY.
[11]. Agresti, A., 1996. An Introduction to Categorical Data Analysis. Wiley, New York.