



Classification of Pepper Seeds Using Data Mining Algorithms

Jesusimo L. Dioses Jr.

College of Computing Science, Information and Communication Technology, Isabela State University,
Echague, Isabela, Philippines, jesusimo.l.dioses@gmail.com

ABSTRACT

This paper is about separating the two classes of pepper seeds namely the bell pepper and the chili pepper. This paper used the same dataset from the previous study obtained using image processing. The image processing process was not repeated instead the extracted features namely the area, equivalent diameter, perimeter and roundness were used. The four features were subjected for classification using the 23 classifiers in the MATLAB Classification Learner App. Among the 23 classifiers, the best in terms of accuracy and speed is the weighted K nearest neighbor which achieved 100% accuracy in 0.06 seconds.

Key words : Bell Pepper, Chili Pepper, Classification Learner App, K Nearest Neighbor, Weighted KNN, Boosted Trees, Bagged Trees, Subspace KNN

1. INTRODUCTION

The bell pepper and chili pepper's fruit can easily be differentiated in terms of size and shape. The bell pepper's fruit is obviously larger and round in shape whereas the chili pepper is small and elongated in shape. But in terms of their seeds, the untrained human eye will have difficulty differentiating the seeds of the two species of pepper. What is difficult in human vision is achievable if machine vision is applied. This paper is the continuation of the paper entitled "Bell Pepper And Chili Pepper Classification: An Application Of Image Processing And Fuzzy Logic". In that paper the morphological features of the two pepper species were extracted using image processing and were classified using fuzzy logic. In the same paper, the area, equivalent diameter, perimeter and roundness percentage were extracted using image processing and the extracted features were used to develop a fuzzy logic classifier. Using fuzzy logic, the features of bell pepper and chili pepper were discriminated with 85% accuracy.

The author saw the need to continue his previous research by using the data on his first research on classification of pepper seeds and apply data mining algorithms as classifiers in the MATLAB's Classification Learner App (CLA)[1]. Objects can be grouped or clustered based on their features using data

mining or machine learning classifiers[2-5]. The major advantage of this approach is that the CLA used 23 classifiers. If one classifier has low accuracy, there is a high chance of obtaining high accuracy on other classifiers. The CLA is a collection of 23 data mining classifiers where in a single file of data set can be analyzed simultaneously by the 23 classifiers. By using CLA, the performance of the data mining classifiers can be compared without going through each classifier one by one through manual coding. CLA has been used in classifying green coffee beans in which the highest classification accuracy was achieved using the course tree algorithm [6]. CLA had been used in classifying civet coffee with the highest accuracy of 86.5% obtained by both the quadratic support vector machine and the fine Gaussian support vector machine[7]. CLA has been used in common sports recognition using inertial sensor with 91.2% accuracy[8]. CLA has been used for classification of coffee bean species and 95.6% accuracy was achieved using the Bagged Trees classifier [9].

2. METHODOLOGY

Image processing is an effective approach in extracting features of an object[6-20]. The image processing done in the author's previous research [1] which is the method for feature extraction of the pepper seeds was not repeated in this paper; however the data from that research were utilized. The difference from the previous study was the classifier was changed from fuzzy logic to data mining algorithms in the MATLAB's Classification Learner App. The data is composed of four features of bell and chili pepper which are the area, equivalent diameter, perimeter and roundness. There are 60 seeds for bell and chili pepper used in this research. The data set was exported into the MATLAB's classification learner app workspace and was subjected to the 23 classifiers. There were 120 observations. There were four predictors which are the four features and there are two responses which are the two pepper species. The CLA was set up to no validation in the settings.

This paper uses the data in a research done previously by the author [1]. The dataset used in the study is available in the previous study [1].

3. RESULTS AND DISCUSSION

Table 1 shows the result of classification using the CLA.

Table 1: Classification results for each classifier using the Classification Learner App without using the PCA

Classifier	Classifier Type	Accuracy (%)
Decision Trees	Fine Tree	90
	Medium Tree	90
	Coarse Tree	77.5
Discriminant Analysis	Linear Discriminant	75
	Quadratic Discriminant	69.2
Logistic Regression	Logistic Regression	73.3
Support Vector Machines (SVM)	Linear SVM	73.3
	Quadratic SVM	75
	Cubic SVM	75
	Fine Gaussian SVM	89.2
	Coarse Gaussian SVM	74.2
K Nearest Neighbor (KNN)	Fine KNN	100
	Medium KNN	78.3
	Coarse KNN	72.5
	Cosine KNN	76.7
	Cubic KNN	77.5
Ensemble Classifiers	Weighted KNN	100
	Boosted Trees	100
	Bagged Trees	100
	Subspace Discriminant	74.2
	Subspace KNN	100
	RUS Boosted Trees	93.3

Table 2 shows the classifiers with 100 % accuracy and their training time.

Table 2: Classifier with 100% Accuracy and their Training Time

Classifier Type	Accuracy (%)	Training Time (s)
Fine KNN	100	0.24764
Weighted KNN	100	0.060582
Boosted Trees	100	0.55509
Bagged Trees	100	0.40112
Subspace KNN	100	0.33857

It can be seen in Table 1 that there are five classifiers that obtained 100 % accuracy in classifying bell pepper seeds from chili pepper seeds. These classifiers are Fine K-Nearest Neighbor, Weighted K Nearest Neighbor, Boosted Trees and Bagged Trees and Subspace KNN.

As shown in Table 2, among 5 classifiers with 100% accuracy the fastest training time is the weighted KNN with 0.060582 seconds followed by Fine KNN with 0.24764 seconds. Subspace KNN comes as third with 0.33857 seconds ; Bagged Trees comes at fourth with 0.40112 seconds and Boosted Trees as fifth with 0.55509 seconds. The two fastest and highest accuracy were from the same family which is the K nearest neighbor.

Figure 1 shows the scatter plot of weighted KNN classifier.

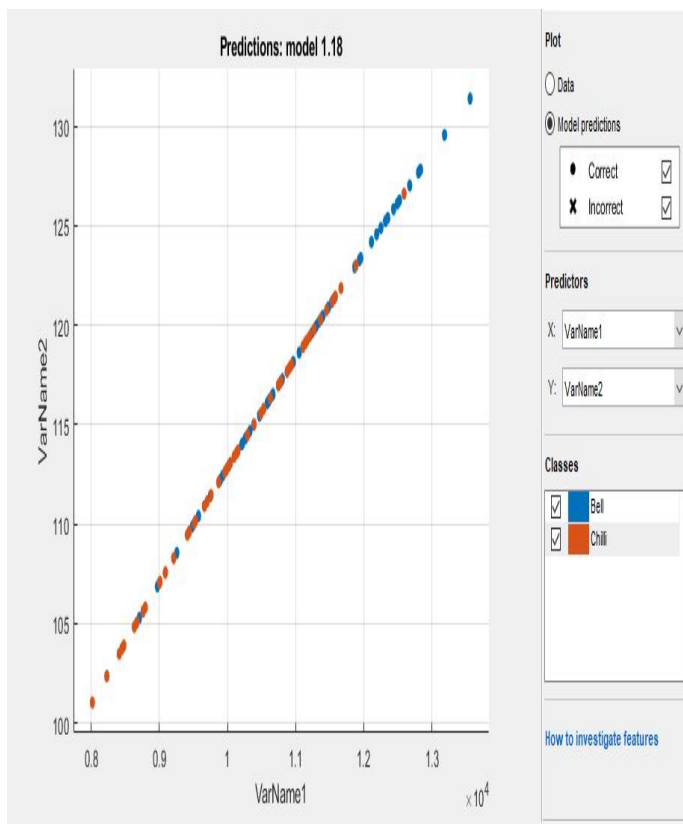


Figure 1: Scatter Plot of Weighted KNN Classifier

It can be seen in Figure 1 that the four features of bell and chilli pepper namely the area, equivalent diameter, perimeter and roundness are overlapping to each other. This means that the untrained eye of human will find it difficult to classify one from the other.

Figure 2 shows the confusion matrix of weighted KNN.



Figure 2: Confusion Matrix of Weighted KNN Classifier

It can be seen in Figure 2, the confusion matrix that the weighted KNN Classifier was able to differentiate the two classes of pepper seeds with 100 % accuracy.

Figure 3 shows the Receiver Operating Characteristic (ROC) curve of the weighted KNN.

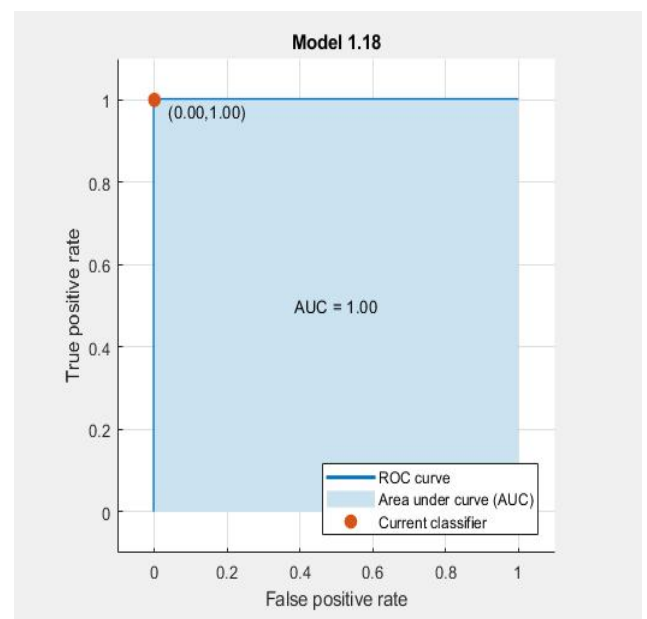


Figure 3: The ROC Curve of Weighted KNN Classifier

In Figure 3 the area under the curve (AUC) is equal to 1. It is the overall measure of the quality of classifier, large AUC indicates better classifier performance. The AUC of the weighted KNN is equal to 1 which means that it has a better performance compared to other classifiers.

Figure 4 show the parallel coordinates plot of weighted KNN.

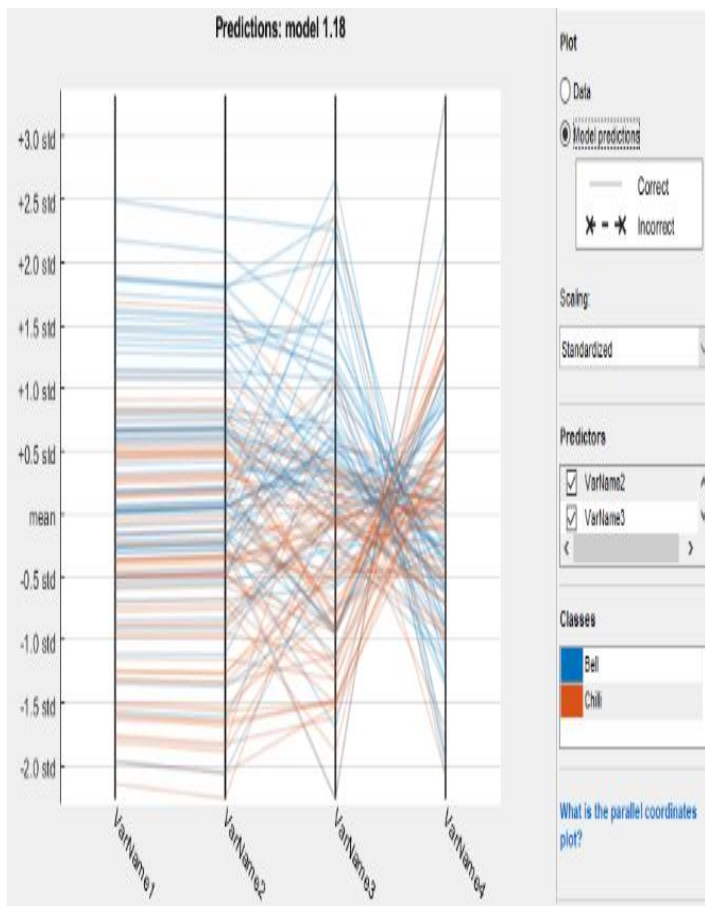


Figure 4: The Parallel Coordinates Plot of Weighted KNN Classifier

The parallel coordinates plot shows the useful variables or features of the bell and chili pepper that contributed to their discrimination. It can be seen in the plot that the features area and equivalent diameter are the two most important features in separating the bell from chili pepper. The third most useful features is the perimeter and the fourth is the roundness. This means that the two classes can be separated in terms of size using weighted KNN.

Table 3 shows the comparison of the previous study from this study.

Table 3: Comparison of the Results of this Study from the Previous Study [1]

Previous Study	Fuzzy Logic Classifier	85 %
This Study	Weighted KNN Classifier	100 %

As shown in Table 3, the classifier in this study which is the weighted KNN performed much better than the classifier in the previous study which uses fuzzy logic in classifying bell pepper seeds from the chili pepper seeds.

4. CONCLUSION

This paper used the different data mining algorithms in separating the seeds of the bell pepper from the chili pepper. The features of the two classes of peppers were extracted using the Classification Learner App of MATLAB. It can be concluded, that among the 23 classifiers the weighted KNN is the best classifier in terms of accuracy and in terms of speed. Also the best features to use is the area and the equivalent diameter.

REFERENCES

1. E. J. Olaes, E. R. Arboleda, J. L. D. Jr, and R. M. Dellosa. **Bell Pepper And Chili Pepper Classification: An Application Of Image Processing And Fuzzy Logic**, *Int. J. Sci. Technol. Res.*, vol. 9, no. 02, pp. 4833–4839, 2020.
2. M. A. Arasi and S. Babu. **Survey of machine learning techniques in medical imaging**, *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 5, pp. 210–2116, 2019. <https://doi.org/10.30534/ijatcse/2019/39852019>
3. H. D. Gadade and D.K.Kirange. **Machine Learning Approach towards Tomato Leaf Disease Classification**, *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 1, pp. 490–495, 2020. <https://doi.org/10.30534/ijatcse/2020/67912020>
4. J. L. D. Mercaral, A. J. P. Delima, and R. N. Vilchez. **Prediction of employees’ lateness determinants using machine learning algorithms**, *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 1, pp. 779–783, 2020. <https://doi.org/10.30534/ijatcse/2020/111912020>
5. M. A. Ottom, N. A. Alawad, and K. M. O. Nahar. **Classification of Mushroom Fungi Using Machine Learning Techniques**, *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 5, pp. 2378–2385, 2019. <https://doi.org/10.30534/ijatcse/2019/78852019>
6. E. R. Arboleda. **Comparing Performances of Data Mining Algorithms for Classification of Green Coffee Beans**, *Int. J. Eng. Adv. Technol.*, vol. 8, no. 5, pp. 1563–1567, 2019.
7. E. R. Arboleda. **Discrimination of civet coffee using near infrared spectroscopy and artificial neural network**, *Int. J. Adv. Comput. Res.*, vol. 8, no. 39, pp. 324–334, 2018.
8. N. F. Ghazali, N. Shahar, N. A. Rahmad, N. A. J. Sufri, M. A. As’ari, and H. F. M. Latif. **Common sport activity recognition using inertial sensor**, in *Proceedings - 2018 IEEE 14th International Colloquium on Signal Processing and its*

- Application, CSPA 2018*, 2018, no. March, pp. 67–71.
<https://doi.org/10.1109/CSPA.2018.8368687>
9. D. L. Bersabal, E. R. Arboleda, and E. M. Galas. **Coffee Bean Recognition Using Shape Features Using Decision Trees And Ensemble Classifiers**, *Int. J. Sci. Technol. Res.*, vol. 9, no. 02, 2020.
 10. A. E. Andaya, E. R. Arboleda, A. A. Andilab, and R. M. Dellosa. **Meat Marbling Scoring Using Image Processing with Fuzzy Logic Based Classifier**, *Int. J. Sci. Technol. Res.*, vol. 8, no. 08, pp. 1442–1445, 2019.
 11. E. R. Arboleda, A. C. Fajardo, and R. P. Medina. **An Image Processing Technique for Coffee Black Beans Identification**, in *2018 IEEE International Conference on Innovative Research and Development (ICIRD)*, 2018, no. May, pp. 1–5.
 12. R. A. Magbayao, E. R. Arboleda, and E. M. Galas. **Identification Of Asian Green Mussel Perna Viridis ' Sex Using Image Processing , Fuzzy Logic And K – Nearest Neighbor**, *Int. J. Sci. Technol. Res.*, vol. 9, no. 01, 2020.
 13. V. M. D. Manalo, E. R. Arboleda, J. L. D. Jr, and R. M. Dellosa. **Differentiation Among Lettuce (L . Sativa) Seed Varieties Grown In Gourmet Farms , Silang Cavite , Philippines Using Image Processing With Fuzzy Logic And Knn As Classifiers**, *Int. J. Sci. Technol. Res.*, vol. 8, no. 10, pp. 8–11, 2019.
 14. K. E. N. T. Rabe, E. R. Arboleda, and R. M. Dellosa, **Fuzzy Logic Based Vehicular Congestion Estimation Monitoring System Using Image Processing And KNN Classifier**, *Int. J. Sci. Tehnol. Res.*, vol. 8, no. 08, pp. 1377–1380, 2019.
<https://doi.org/10.1109/ICIRD.2018.8376325>
 15. J. G. C. Rancapan, E. R. Arboleda, J. L. D. Jr, and R. M. Dellosa. **Egg Fertility Detection Using Image Processing And Fuzzy Logic**, *Int. J. Sci. Technol. Res.*, vol. 8, no. 10, pp. 3228–3230, 2019.
 16. F. B. Vidad III, E. R. Arboleda, J. L. Dioses Jr, and R. M. Dellos. **Morphological Based Seedling Comparison Of Natural And Hybrid Lemon,”** *Int. J. Sci. Technol. Res.*, vol. 9, no. 02, 2020.
 17. E. R. Arboleda, A. C. Fajardo, and R. P. Medina. **Classification of Coffee Bean Species Using Image Processing , Artificial Neural Network and K Nearest Neighbors**, in *2018 IEEE International Conference on Innovative Research and Development (ICIRD)*, 2018, no. May, pp. 1–5.
<https://doi.org/10.1109/ICIRD.2018.8376326>
 18. E. R. Arboleda and E. M. Galas. **Image Processing Of Clean And Dirty Dishes To Design And Construct A Fuzzy Logic Dishwasher**, *Int. J. Sci. Technol. Res.*, vol. 8, no. 12, pp. 8–11, 2019.
 19. R. E. T. Bae, E. R. Arboleda, A. Andilab, and R. M. Dellosa. **Implementation Of Template Matching , Fuzzy Logic And K Nearest Neighbor Classifier**
 20. C. I. A. Baluran, E. R. Arboleda, M. G. Dizon, and R. M. Dellosa. **Crab Gender Classification Using Image Processing , Fuzzy Logic And K Nearest Neighbor (KNN) Classifier**, *Int. J. Sci. Technol. Res.*, vol. 8, no. 10, pp. 1458–1462, 2019.
 21. N. O. Delgado, E. R. Arboleda, J. L. Dioses Jr, and R. M. Dellosa. **Identification Of Mango Leaves Using Artificial Intelligence**, *Int. J. Sci. Technol. Res.*, vol. 8, no. 12, pp. 2864–2868, 2019.
 22. J. A. A. Garcia, E. R. Arboleda, and E. M. Galas. **Identification Of Visually Similar Vegetable Seeds Using Image Processing And Fuzzy Logic**, *Int. J. Sci. Technol. Res.*, vol. 9, no. 02, pp. 4925–4928, 2020.
 23. A. J. G. Gonzales, J. A. D. Sosa, E. R. Arboleda, and E. M. Galas. **Filtering Of Faded Coffee Beans Using Image Processing**, *Int. J. Sci. Technol. Res.*, vol. 8, no. 12, pp. 3634–3637, 2019.
 24. C. L. Macalalad, E. R. Arboleda, A. A. Andilab, and R. M. Dellosa. **Morphological Based Grain Comparison of Three Rice Grain Variety**, *Int. Journal Sci. Technol. Res.*, vol. 8, no. 08, pp. 1446–1450, 2019.