Performance Evaluation of Classification Techniques for Computer Vision based Cashew Grading System

Mayur Thakkar
Charotar University of
Sci. and Technology,
Changa, India, - 388421.

Malay Bhatt
Dharmsinh Desai University,
Nadiad, India - 387001.

C. K. Bhensdadia
Dharmsinh Desai University,
Nadiad, India - 387001.

ABSTRACT

The performance of various classification algorithms greatly depends on the characteristics of the data to be classified. There is no single classifier that works best on all given problems. The purpose of this study is to develop the computer vision based cashew grading system in conjunction with most accurate classification technique. The performance of different classification techniques including Multi-Layer Perceptron, Naive Bayes, K-Nearest Neighbor, Decision tree, Support Vector Machine are evaluated using WEKA toolbox to have most suitable classification technique for the cashew grading system. Subsequently, the classification technique that has the potential to significantly improve the performance of the system is suggested to be utilized in cashew grading system.

Keywords
Cashew grading system, Decision tree, k-Nearest Neighbors, Multi-Layer Perceptron (MLP), Naïve Bayes, Support Vector Machine (SVM).
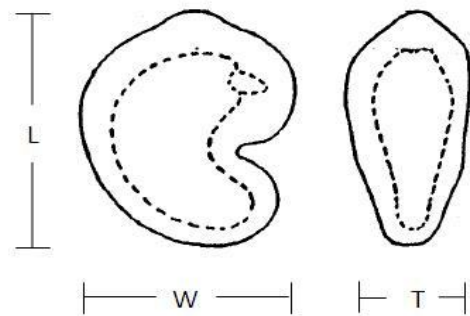
1. INTRODUCTION

Cashew is one of the most popular tree nuts. It is an expensive agricultural product and the prices depend on its quality. Today, various kinds of cashews are available in the market with different qualities. To ascertain the quality, grade standard have been designed by considering the color and the size (weight) of the cashew kernel as important characteristic as shown in Table 1 and Table 2.

Several attempts have been made to mechanize the grading of the kernels, with limited success. Power driven rotary sieves are one mechanical method, another being the use of two outwardly rotating rubber rollers aligned at a diverging angle. Because of direct contact, which can cause the damage to the cashew kernel, mechanical grading system is not appropriate for the cashew kernel grading.

With exception of few mechanical methods, grading of the cashew kernel is still labor intensive manual process. Cashew kernels are mostly graded manually by skilled labor, employed only for grading, but the quality decisions may vary among the graders and are inconsistent. This way of grading presents many quality problems and grading is the last opportunity for the quality control.

Computer vision system has proven successful for the objective, online measurement of several agricultural products [2]. Computer vision based cashew grading system is an alternative to the manual, mechanical and optical methods. This method offers automated, high speed, non-destructive and cost effective technique for classification. Designing such system without taking the physical properties of cashew kernel into consideration may yield poor results. In [1], the physical properties of the raw cashew nut and cashew kernel have been evaluated. Length (L), Width (W) and Thickness (T) of the cashew kernel plays vital role in deciding the grade of the cashew kernel which are measured as shown in Fig. 1.

Fig 1: Length (L), Width (W) and Thickness (T) of the cashew kernel.

2. METHODS AND MATERIAL
The samples of whole cashews of different grades, used in this study were collected from Orbitta Exports, one of the cashew production companies of Gujarat. Initially the different samples of the each grade are taken and weight of each cashew kernel is measured individually with accuracy of 0.001 gm.

A Computer Vision System is developed which consists of two digital cameras placed in front and top of cashew sample under investigation at distance of 15 cm from the sample position as well as perpendicular to each other, an image capturing box, fluorescent lamp and computer system. Image processing toolbox in the MATLAB is used as image analysis and processing software to extract the features from the image. Fig. 2 shows the general operations for the cashew grading system.

Table 1. Color characteristic of the whole cashew kernel

Cashew Kernel Type	Color Characteristic
White Whole (W)	Cashew kernels are white and free from damage.
Scorched Whole (SW)	Cashew kernels are light brown and free from damage.
Dessert Whole (DW)	Cashew kernels are dark brown, it may show deep black spot and free from damage.

Table 2. Weight characteristic of the whole cashew kernel

White Whole Grades	Number of Kernels Per 454 gms.	Scorched Whole Grades	Number of Kernels Per 454 gms.	Dessert Whole Grades	Number of Kernels Per 454 gms.
W180	170-180	SW180	170-180		
W210	200-210	SW210	200-210		
W240	220-240	SW240	220-240		
W280	260-280	SW280	260-280		
W320	300-320	SW320	300-320	DW	No sepcification
W400	350-400	SW400	350-400		
W450	400-450	SW450	400-450		
W500	450-450	SW500	450-450		
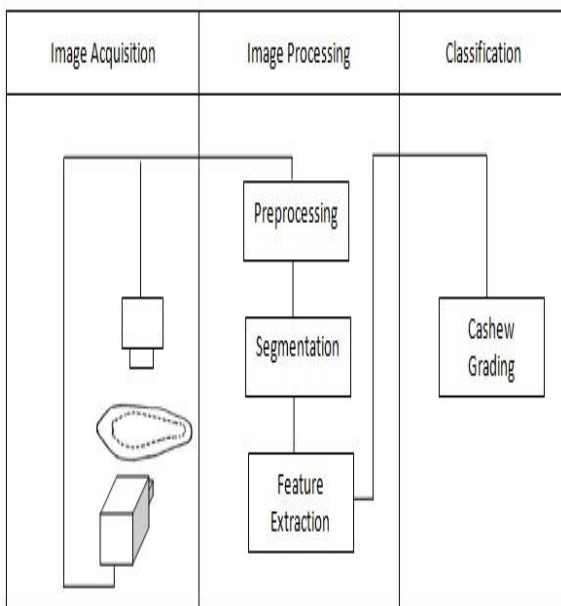
Fig 2: Schematic representation of cashew grading system

Image acquisition involves capturing of RGB front and top view images of each cashew kernel under study. During Preprocessing phase image is smoothed using 3x3 average filter. In this study, black color background is utilized to have bimodal histogram. Threshold segmentation technique differentiates the cashew kernel region from background and converts the gray-scale image into the binary image. To estimate the grade of the cashew kernel, Length, Width, Thickness and Color of the cashew kernel are considered as important features. These Features are extracted using image analysis and image processing algorithms.

The dataset for the cashew is designed from the extracted features as shown in Fig. 3, with a total data of 6800 and a dimension of 1700 rows, each row contains the data of single cashew and 4 columns, each column contains data for one of the extracted features value of color, length, width, thickness. Waikato Environment for Knowledge Analysis (WEKA) toolbox is used to evaluate and find most suitable classification techniques for the cashew grading system among various multi-class classification techniques including Multi-Layer Perceptron (MLP), Decision Tree, k-Nearest Neighbors, Naïve bayes and Support Vector Machine.

Fig 3: Data set of cashew kernels

3. CLASSIFICATION TECHNIQUES

3.1 Multi-Layer Perceptron

Multilayer Feed forward Neural Networks [5] provide a natural extension to the multiclass problem. Instead of just having one neuron in the output layer, with binary output, we could have N binary neurons. The output codeword corresponding to each class can be chosen as either one-per-class coding or distributed output coding:

• *One-per-class coding*: Each output neuron is designated the task of identifying a given class. The output code for that class should be 1 at this neuron, and 0 for the others. Therefore, we will need N = K neurons in the output layer, where K is the number of classes. When testing an unknown example, the neuron providing the maximum output is considered the class label for that example.

• *Distributed output coding*: Each class is assigned a unique binary codeword from 0 to $2N - 1$, where N is the number of output neurons. When testing an unknown example, the output codeword is compared to the codewords for the K classes, and the nearest codeword, according to some distance measure, is considered the winning class. Usually the Hamming distance is used in that case, which is the number of different bits between the two codewords. For instance, for a 4 class problem, and using N = 5 bit codewords, the coding can be as shown in table 2. The hamming distance between each pair of classes is equal to 3 i.e. each two codes differ in three bits. If we got a codeword for an unknown example as 11101, we compute its distance from the four codewords shown above. The nearest codeword is that for class 3 with a distance of 1, so class label assigned to that example will be class 3.

3.2 Decision Tree

Decision trees are a powerful classification technique. Two widely known algorithms for building decision trees are Classification and Regression Trees [6] and ID3/C4.5 [7]. The tree tries to infer a split of the training data based on the values of the available features to produce a good generalization. The split at each node is based on the feature that gives the maximum information gain. Each leaf node corresponds to a class label. A new example is classified by following a path from the root node to a leaf node, where at each node a test is performed on some feature of that example. The leaf node reached is considered the class label for that example. The algorithm can naturally handle binary or multiclass classification problems. The leaf nodes can refer to either of the K classes concerned.

3.3 K-Nearest Neighbors

K-Nearest Neighbors [8] is considered among the oldest non-parametric classification algorithms. To classify an unknown example, the distance (using some distance measure e.g. Euclidean) from that example to every other training example is measured. The k smallest distances are identified, and the most represented class in these k classes is considered the output class label. The value of k is normally determined using a validation set or using cross-validation.

3.4 Naive Bayes

Naive Bayes [9] is a successful classifier based upon the principle of Maximum A Posteriori (MAP). Given a problem with K classes {C1, . . . ,CK} with so-called prior probabilities P(C1), . . . , P(CK), we can assign the class label c to an unknown example with features x = (x1, . . . , xN) such that c = argmaxcP(C = ckx1, . . . , xN), that is choose the class with the maximum a posterior probability given the observed data. This aposterior probability can be formulated, using Bayes theorem, as follows: P(C = ckx1, . . . , xN) = P(C=c)P(x1,...,xNkC=c) P(x1,...,xN). As the denominator is the same for all classes, it can be dropped from the comparison.

Now, we should compute the so-called class conditional probabilities of the features given the available classes. This can be quite difficult taking into account the dependencies between features. The naive bayes approach is to assume class conditional independence i.e. x1, . . . , xN are independent given the class. This simplifies the numerator to be P(C = c)P(x1kC = c) . . . P(xNkC = c), and then choosing the class c that maximizes this value over all the classes c = 1, . . . ,K. Clearly this approach is naturally extensible to the case of having more than two classes, and was shown to perform well in spite of the underlying simplifying assumption of conditional independence.

3.5 Support Vector Machine

Support Vector Machines are among the most robust and successful classification algorithms [10,11]. They are based upon the idea of maximizing the margin i.e. maximizing the minimum distance from the separating hyperplane to the nearest example. The basic SVM supports only binary classification, but extensions have been proposed to handle the multiclass classification case as well. In these extensions, additional parameters and constraints are added to the optimization problem to handle the separation of the different classes. The formulation can result in a large optimization problem, which may be impractical for a large number of classes.

4. EXPERIMENTAL RESULTS

The performance of the different classification techniques are evaluated with respect to whole cashew kernel data set. The experimental results are as per Table 3.

Table 3. Performance evaluation of different classification techniques

Sr. No.	Classification Techniques	Classification Accuracy
1	ML-Perceptron	86%
2	Decision tree	79%
3	k-Nearest Neighbors	76%
4	Naive Bayes	81%
5	SVM	77%

5. CONCLUSION

The aim of this study is to investigate the performance of different multiclass classification techniques against whole cashew data set and to find the most appropriate technique for cashew grading system. Performance of the various classifiers including Multi-Layer Perceptron (MLP), Decision Trees, k-Nearest Neighbors, Naive Bayes, Support Vector Machine are evaluated and it is observed that Multi-Layer Perceptron classification technique is more feasible to be used in cashew grading system as it possesses comparatively higher classification accuracy of 86% than other classifiers.

6. FUTURE WORK

In this study we evaluate the performance of the several classifier techniques to identify the efficient classifier for the automatic cashew grading system. In cashew kernel grading specification, there is presence of imprecision in weight based characteristic. If number of white cashew kernels per pound (454 gms) are in between 170 to 180 then cashew kernel grade is W180 and if number of white cashew kernels per pound are in between 200 to 210 then cashew kernel grade is W210. But it is observed that there is no specification for in-between ranges (181-199 cashew kernels per pound). Because of this kind of the imprecision it is possible that the fuzzy logic can be more effective for the decision control. Therefore in the future work, for the same dataset of the whole cashew kernel, Fuzzy Inference System will be designed to improve the classification accuracy.

7. REFERENCES

[1] Balasubramanian, D. 2001. Physical properties of raw cashew nut. Journal of agricultural Engineering Research, 78(3), pp. 291-297.

[2] Brosnan, T., and Sun, D. W. 2004. Improving quality inspection of food products by computer vision. Journal of Food Engineering, 61, pp. 3-16.

[3] F. Ibrahim, N.A. Abu Osman, J. Usman and N.A. Kadri, 2007. Comparison of Different Classification Techniques Using WEKA for Breast Cancer.

[4] Tavel, P. 2007 Modeling and Simulation Design. AK Peters Ltd. Rafael C. Gonzalez and Richard E. Woods, 2002. Digital Image Processing. Prentice-Hall International.

[5] Bishop, M. 1995. Neural networks for Pattern Recognition. Oxford University Press.

[6] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, 1984. Classification and Regression Trees. Chapman and Hall.

[7] J. R. Quinlan, 1993. C4.5:Programs for Machine Learning. Morgan Kaufmann.

[8] Bay, S., 1998. Combining nearest neighbor classifiers through multiple feature subsets. International Conference on Machine Learning, 37–45.

[9] Rish, I., 2001. An empirical study of the naive bayes classifier", In IJCAI Workshop on Empirical Methods in Artificial Intelligence.

[10] Burges C., 1998. A tutorial on support vector machines for pattern recognition. In Data Mining and Knowledge Discovery, 1–47.

[11] Dietterich T., Bakiri G., 1995. Solving multiclass learning problems via error correcting output codes. Journal of Artificial Intelligence Research, 1–38.

[12] MATLAB documentation on Image Processing Toolbox. www.mathworks.com.