CLASSIFICATION OF SELECTED APPLE FRUIT VARIETIES USING NAIVE BAYES

Misigo Ronald
School of Computing and Informatics, University of Nairobi
P.O Box 30197 – 00100, Nairobi, Kenya
misigoronald@gmail.com

Miriti Evans
School of Computing and Informatics, University of Nairobi
P.O Box 30197 – 00100, Nairobi, Kenya
eamiriti@uonbi.ac.ke

Abstract

Manual sorting of apple fruit varieties results to high cost, subjectivity, tediousness and inconsistency associated with human beings. A means for distinguishing apple varieties is needed and therefore, some reliable technique is needed to discriminate varieties rapidly and non-destructively. The main objective of this research was to investigate the applicability and performance of Naive Bayes algorithm in the classification of apple fruit varieties. The methodology involved image acquisition, pre-processing and segmentation, analysis and classification of apple varieties. Apple classification system prototype was built using MATLAB R2015a development platform environment. The results showed that the averaged values of the estimated accuracy, sensitivity, precision and specificity were 91%, 77%, 100% and 80% respectively. Through previous research works, the literature review identified MLP-Neural (Unay et al., 2006), fuzzy logic (Kavdir et al., 2003), principal components analysis (Bin et al., 2007) and neural networks (Ohali et al., 2011) as other technique which have been used previously to classify apple varieties. Comparison of their classification accuracy results with that of Naive Bayes technique showed that the accuracy of Naive Bayes was higher than the accuracy of principal components analysis, fuzzy logic and MLP-Neural with 91%, 90%, 89%, and 83% respectively. This study indicated that Naive Bayes has good potential for identification of apple varieties nondestructively and accurately.

Keywords: Apple fruit; Sorting and Grading; Image Processing Techniques; Naive Bayes Technique; Pattern Recognition; Classification.

1. INTRODUCTION

Apples (malus sp., Rosaceae) are one of the most commonly consumed fruits in the world. In 2011, world apple production was estimated at around 75 millions of tons according to Food and Agriculture Organization stats (15 July 2013). Apples are an important agricultural commodity in the global market for fresh products. The quality of an apple depends on its external characteristics, such as color, size, and surface texture, and internal parameters, such as sweetness, acidity, firmness, tissue texture, ascorbic acid, and polyphenolic compounds (Wojdylo et al. 2008). These characteristics, especially internal and external parameters, are similar to a variety. However, each variety has its special characteristics and flavor, which results in different prices and preferences by different people.

Apple produce dealers have warehouses that store different varieties of apple fruits. Therefore, different apple varieties easily get mixed up during harvesting, storage and marketing. Most apple produce dealers will sort the apples manually which results in high cost, subjectivity, tediousness and inconsistency associated with manual sorting. The main objective of this study was to investigate the applicability and performance of Naive Bayes algorithm in the classification of apple fruit varieties.

2. LITERATURE REVIEW

Digital image processing, as a computer-based technique, has been extremely used by scientists to solve problems in agriculture. Fernando et al (2010) built a system to diagnose six different types of surface defects in citrus fruits using a multivariate image analysis strategy. Images were unfolded and projected onto a reference eigenspace to arrive at a score matrix used to compute defective maps and 94.2% accuracy was reported. Cho et al. (2013) used hyperspectral fluorescence imaging for detecting cracking defects on cherry tomatoes while Omid et al. (2013) used shape, texture and color features to sort tomato fruits according to their circularity, size, maturity, and defects. They achieved 84.4% accuracy for defect detection using a probabilistic neural network (PNN) classifier.

Chowdhury et al. (2013) have recognized 10 different vegetables using the color histogram and statistical texture features. They have gained the classification accuracy up to 96.55% using a neural network as a classifier. Danti et al. (2012) classified 10 types of leafy vegetables using BPNN classifier with a success rate of 96.40%. They first cropped and resized the image and then extracted the mean and range of hue and saturation channel of HSV image to form the feature vector. Suresha et al. (2012) have achieved 95% classification accuracy over a dataset of containing 8 types of different vegetables using texture measures in RGB color space. They have used watershed segmentation to extract the region of interest as a pre-processing and decision tree classifier for training and classification purpose.

Omid et al. (2013) used shape, texture and color features to sort tomato fruits according to their circularity, size, maturity and defects. They achieved 84.4% accuracy for defect detection using a probabilistic neural network (PNN) classifier. Color, texture and shape features have been evaluated for fruit defect detection system, also in conjunctions with PNNs.

Dubey & Jalal (2012a cited in Dubey & Jalal 2013) proposed a framework for recognizing and classifying fruits and vegetables. They considered images of 15 different types of fruit and vegetable collected from a supermarket. Their approach was to first segment the image to extract the region of interest and then calculate image features from that segmented region which was further used in training and classification by a multi-class support vector machine. They also proposed an Improved Sum and Difference Histogram (ISADH) texture feature for this kind of problem. From their results, ISADH outperformed the other image color and texture features.

Haiguang et al. (2012) classified two kinds of wheat diseases based on color, shape and texture features to train a back propagation neural network. The resulting system achieved a classification accuracy of over 90%. Arefi et al. (2011) developed a segmentation algorithm for the guidance of a robot arm to pick the ripe tomato using image processing technique. To reach this aim, they prepared a machine vision system to acquire images from a tomato plant. Their algorithm works in two phases: (1) background subtraction in RGB color space and then extracting the ripe tomato considering a combination of RGB, HSI, and YIQ color spaces and (2) localizing the ripe tomato using morphological features of the image. They achieved accuracy up to 96.36% on 110 tomato images. Haidar et al. (2012) presented a method for classification of date fruits automatically based on pattern recognition and computer vision. They extracted appropriately crafted a mixture of 15 different visual features, and then, tried multiple classification methods. Their performance ranged between 89% and 99%.

Cho et al. (2013) used hyperspectral fluorescence imaging for detecting cracking defects on cherry tomatoes. Omid et al. (2013) used shape, texture and color features to sort tomato fruits according to their circularity, size, maturity and defects. They achieved 84.4% accuracy for defect detection using a probabilistic neural network (PNN) classifier. Danti et al. (2012) classified 10 types of leafy vegetables using BPNN classifier with a success rate of 96.40%. They first cropped and resized the image and then extracted the mean and range of hue and saturation channel of HSV image to form the feature vector. Suresha et al. (2012) have reached 95% classification accuracy over a dataset of containing 8 types of different vegetables using texture measures in RGB color space. They have used watershed segmentation to extract the region of interest as a pre-processing and decision tree classifier for training and classification purpose.

3. METHODOLOGY

This part describes the process of analysis and design, which describes the apple classification system structure chart. The details of each element are described below:

3.1 Materials and Apple Samples

The experimented apple varieties included: golden delicious, honey crisp and pink lady, (Fig 1) which were bought from the local market.

(a) (b) (c)

Fig. 1: Apple varieties: (a) Golden delicious (b) Honey crisp (c) Pink lady

In the presented method, color and size features were extracted from the apple images which were used as inputs for classification by being fed into Naive Bayes algorithm. Apple classification system prototype was built using MATLAB R2015a development platform environment. The hardware requirement was a processor- Intel® Core(TM) 2 duo CPU T6500 @ 2.10 GHz laptop with 4GB RAM.

3.2 Apple varieties classification system architecture

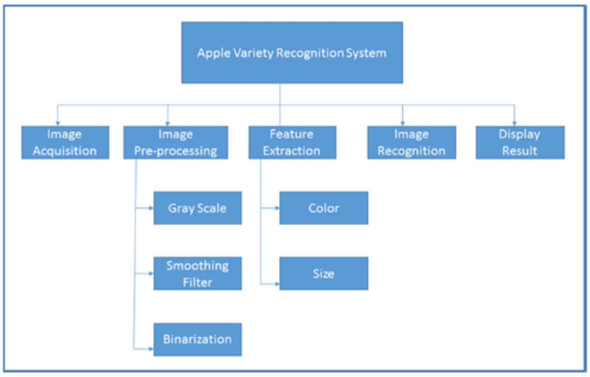

Fig. 2: Structure Chart of Apple Classification System

3.2.1 Image Acquisition

It is a preparation process to obtain apple varieties images. The 150 RGB color images of apple varieties were captured using a phone camera with a pixel resolution of 2048x1024 on a white background. These images were cropped into smaller images and stored in JPG format. The acquired apple varieties images are shown in Fig. 1.

3.2.2 Image Segmentation and Pre-processing

The raw data was subjected to several preliminary processing steps to make it functional in the descriptive stages of classification and grading. In order to get apple features accurately, apple fruits images were pre-processed through different pre-processing methods. These methods were converting RGB to grayscale images and filtering the images to remove noise as described below:

a. Converting RGB to Gray Scale Image

The segmentation and pre-processing task are the initial stages before the image is used for the next process. The main objective of this process is to obtain the binary image with Otsu method. The Otsu method is based on selecting the lowest point between two classes of the histogram by considering the between-class variance.

b. Filtering

The averaging filter was implemented in this process to remove noise. The average filter computes the mean (average) of the gray-scale values within a rectangular filter window surrounding each pixel. This has the effect of smoothing the image (eliminating noise). The filtered pixel was calculated using the equation (1) below:

$$r = \frac{a1 + a2 + \cdots \ldots . + a9}{9} \qquad (1)$$

3.3 Naive Bayes

Naive Bayes classifier is a probabilistic classifier based on the Bayes theorem, considering Naive (Strong) independence assumption. Naive Bayes classifiers assume that the effect of a variable value on a given class is independent of the values of another variable. This assumption is called class conditional independence. Naive Bayes can often perform more sophisticated classification methods. It is particularly suited when the dimensionality of the inputs is high. When we want more competent output, as compared to other methods output we can use Naive Bayes implementation. Naive Bayes is used to create models with predictive capabilities.

$$Bayes'Theorem: Probability\ (B\ given\ A) = \ Probability \frac{A\ and\ B}{Probability} \qquad (2)$$

3.3.1 Naive Bayes Training and Testing

In this stage, there was an apple varieties image database consisting of 50 samples of golden delicious apple images, 50 samples of pink lady apple image and 50 samples of honey crisp apple image which were used for training, validation, and testing purposes

3.4 Image Classification

After training the system, it then classified the apple varieties whether it is golden delicious apple, pink lady apple or honey crisp apple during testing and validation stages.

4. RESULT & DISCUSSIONS

The Naive Bayes classifier was tested for the given classification task. Random sampling was performed, as 60 images were used for training, 30 images for validation and the remaining 60 were used for testing purposes. We recorded true positive, true negative, false positive and false negative for each apple variety for the validation and testing set with the result shown in Table 1 & 2.

Table 1 Result for Validation Data Set

Apples Name	T	TP	FP	TN	FN
Pink Lady	10	6	0	4	1
Honey Crisp	10	8	0	4	1
Golden Delicious	10	9	0	0	1

Table 2.Result for Testing Data Set

Apples Name	T	TP	FP	TN	FN
Pink Lady	20	15	0	6	2
Honey Crisp	20	17	0	5	2
Golden Delicious	20	18	0	0	2

Where T = Number of Testing Data in Table 1 & Number of Validation data in Table 2, TP = True Positive, FP = False Negative, TN = True Negative, FN = False Negative

4.1 Evaluation of the System

To evaluate the performance of the system, statistical analysis of experimented results was done. Statistical results in terms of precision (6), sensitivity (3), specificity (4) and accuracy (5) was calculated.

$$Sensitivity = \frac{TP}{TP + FN} 100\% \qquad (3)$$

$$Specificity = \frac{TN}{TN + FP} 100\% \qquad (4)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} 100\% \qquad (5)$$

$$Precision = \frac{TP}{TP + FP} 100\% \qquad (6)$$

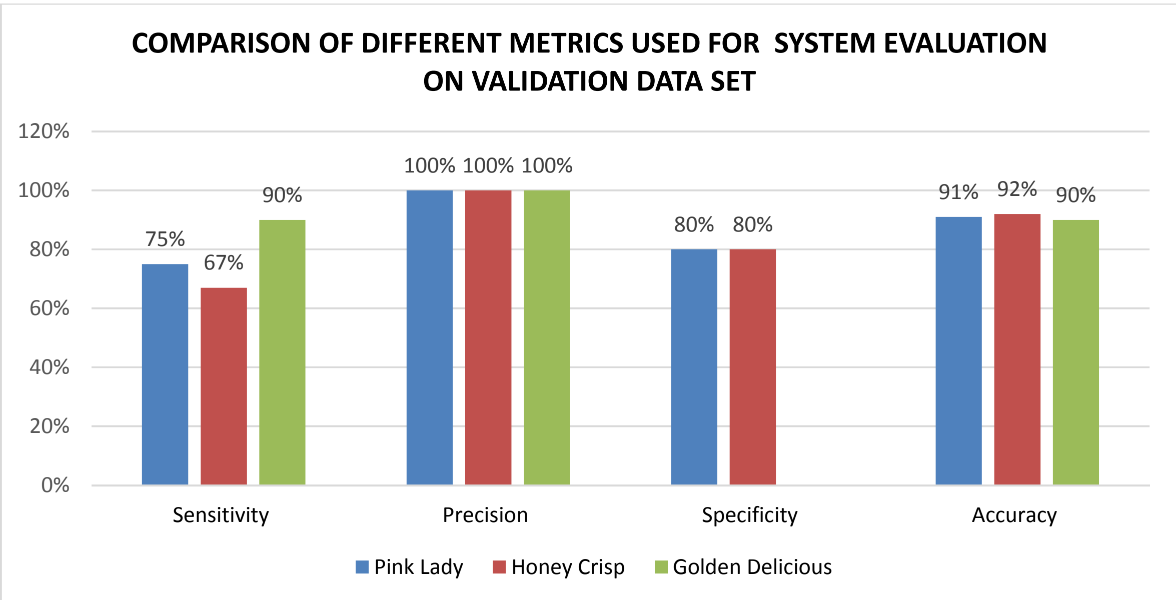

Fig. 3: Graph Showing Comparison of Different Metrics Used to Evaluate the System Using Validation Data Set

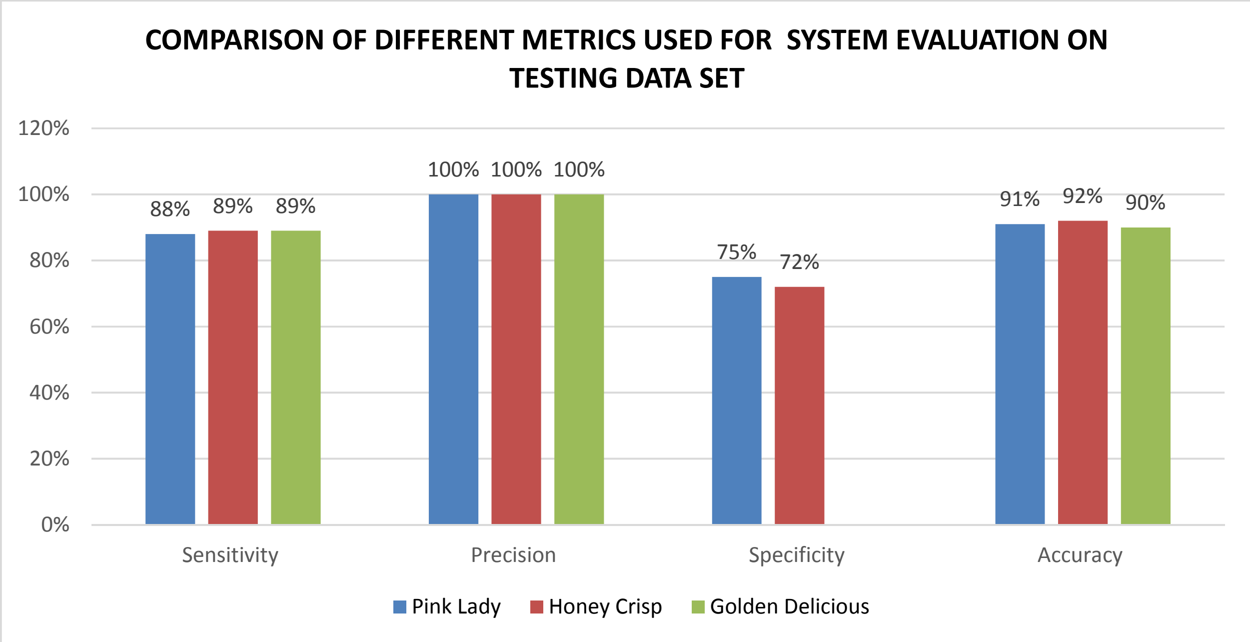

Fig. 4: Graph Showing Comparison of Different Metrics Used to Evaluate the System Using Test Data Set

4.2 Discussion

Comparison of the apple variety accuracy for the validation and testing data set showed that the highest accuracy in Naive Bayes was observed in Honey Crisp (92%), Pink lady (91%) and the last was Golden delicious (90%).The precision of the validation and testing data was 100%.The sensitivity rates of the validation data set were 75% for pink lady, 67% for honey crisp and 90% for golden delicious. For the testing data set, the sensitivity was 88% for pink lady, 89% for honey crisp and 90% for golden delicious. The average sensitivity of the system was 77% for the validation data set and 89% for the testing data set.

The specificity rates of the validation data set were 75% for pink lady, 72% for honey crisp and 0% for golden delicious. For the testing data set, the specificity was 80% for pink lady, 80% for honey crisp and 0% for golden delicious. Golden delicious had a specificity of 0% because during validation and testing we did not find its true negative and false positive values. This can be attributed to its unique color (green) which was easily distinguishable from the other two apple varieties whose colors were almost similar. The average specificity for the system was 80% for the validation data set and 74% for the testing data set.

4.3 Comparing the Results with Existing Methods

We compared the performance of the Naive Bayes classifier with other different approaches already used to solve apple classification and recognition problem. The comparisons were made on the basis of features extracted, classifiers used, and accuracy achieved.

The techniques used for validation were: MLP-Neural (Unay et al., 2006), fuzzy logic (Kavdir et al., 2003), principal components analysis (Bin et al., 2007) and neural networks (Ohali et al., 2011). The accuracy of Naive Bayes, principal components analysis, fuzzy logic and MLP-Neural were 91%, 90%, 89%, and 83% respectively.

5. CONCLUSION

Apple classification system prototype using image processing technique and Naive Bayes algorithm was built using MATLAB R2015a development platform environment. The results related to the three apple varieties: Honey crisp, golden delicious and pink lady showed that the averaged values of the estimated accuracy, sensitivity, precision and specificity were 91%, 77%, 100% and 80% respectively. Through previous research works, the literature review identified MLP-Neural (Unay et al., 2006), fuzzy logic (Kavdir et al., 2003), principal components analysis (Bin et al., 2007) and neural networks (Ohali et al., 2011) as other technique which have been used previously to classify apple varieties. Comparison of their classification accuracy with that of Naive Bayes technique showed that the accuracy of Naive Bayes was higher than the accuracy of principal components analysis, fuzzy logic and MLP-Neural with 91%, 90%, 89%, and 83% respectively. The study indicated that Naive Bayes has good potential for identifying apple varieties nondestructively and accurately. Though this system cannot match the precision and accuracy of the human eye and hand, but the speed and the cost at which they work can be easily be overcome.

Acknowledgments

The authors would like to thank Deputy Director, Dr. Christopher Muturi, University of Nairobi, School of Computing and Informatics for his continued guidance on the research.

References

[1] Arribas JI, Sánchez-Ferrero GV, Ruiz-Ruiz G and Gómez-Gil J. (2011). Leaf classification in sunflower crops by computer vision and neural networks, Computer Electron Agr 78: 9-18.
[2] Al Ohali Y. (2011). Computer vision based date fruit grading system: design and implementation. KSU_CIS 23: 29-36.
[3] Anami BS, Pujari JD and Yakkundimath R. (2011). Identification and classification of normal and affected agriculture/horticulture produce based on combined color and texture feature extraction. IJCAES I (III): 356-360.
[4] Dubey, S. R. (2012). Automatic Recognition of Fruits and Vegetables and Detection of Fruit Diseases. Master's theses, GLA University Mathura, India.
[5] Dubey, S. R., & Jalal, A. S. (2012a). Robust Approach for Fruit and Vegetable Classification. Procedia Engineering, 38, 3449 – 3453.
[6] Dubey, S. R., & Jalal, A. S. (2012b). Detection and Classification of Apple Fruit Diseases using Complete Local Binary Patterns. In Proceedings of the 3rd International Conference on Computer and Communication Technology (pp. 346-351), MNNIT Allahabad, India.
[7] Dubey, S. R., & Jalal, A. S. (2012c). Adapted Approach for Fruit Disease Identification using Images. International Journal of Computer Vision and Image Processing.
[8] Dubey, S. R., & Jalal, A. S. (2013). Species and Variety Detection of Fruits and Vegetables from Images. International Journal of Applied Pattern Recognition, 1(1), 108 – 126.
[9] Dubey, S. R., Dixit, P., Singh, N., & Gupta, J. P. (2013). Infected fruit part detection using K-means clustering segmentation technique. International Journal of Artificial Intelligence and Interactive Multimedia, 2(2).
[10] Ebrahimi E, Mollazade K, Arefi A. (2011). Detection of Greening in Potatoes using Image Processing Techniques. Journal of American Science. 7(3).
[11] Fernando, L. -G., Gabriela, A. G., Blasco, J., Aleixos, N. and Valiente, J. M. (2010). Automatic detection of skin defects in citrus fruits using a multivariate image analysis approach. Computers and Electronics in Agriculture. 71(2), 189-197.
[12] Gabriel, A. L. V. and Aguilera, J. M. (2013). Automatic detection of orientation and diseases in blueberries using image analysis to improve their postharvest storage quality. Food Control, 33(1), 166–173.
[13] Ouyang, C., Li, D., Wang, J., Wang, S., and Han, Y. (2013). The Research of the Strawberry Disease Identification Based on Image Processing and Pattern Recognition. Computer and Computing Technologies in Agriculture VI, 392, 69-77.