Full Paper

A DECISION TREE CLASSIFICATION MODEL FOR COCOA BEANS QUALITY EVALUATION BASED ON DIGITAL IMAGING

A. O. Tanimowo
Department of Computer Science and Engineering, Obafemi Awlowo University, Ile-Ife, Nigeria.
aminatadedamola@yahoo.co.uk

S. I. Eludiora
Department of Computer Science and Engineering, Obafemi Awlowo University, Ile-Ife, Nigeria

A. O. Ajayi
Department of Computer Science and Engineering, Obafemi Awlowo University, Ile-Ife, Nigeria

ABSTRACT

Cocoa is an important export commodity and a major foreign exchange earner for producing Countries. Cocoa beans of good qualities have more economic value than defective ones. The chocolate obtained from the good cocoa gives an aroma that is sought after. Defects in cocoa beans are detected and revealed through a cut test procedure usually carried out by the Federal Produce Inspectors. In this research, the use of a supervised machine learning approach, a decision tree, for the classification of cocoa beans into good, slatty, mouldy and weevilly beans was proposed. A digital camera was used to capture images of 300 cut cocoa beans and the images were saved in the RGB colour format on a computer system. Colour and statistical features were extracted from these images by moment analysis of the colour channels (Red, Blue and Green). Fourteen features were totally extracted and saved in a feature database. Classification and regression tree (CART), a decision tree algorithm, was employed for the classification and three splitting criteria were tested: gini diversity index (gdi), twoing and deviance. The model was evaluated by its average prediction accuracy and the receiver operating characteristics (ROC) curve for each class was plotted. The maximum classification accuracy of 89.2% was recorded from the gdi splitting criteria. With this high classification accuracy, it was concluded that the decision tree machine learning algorithm is effective in the classification of dried farm produce like cocoa.

Keywords: Machine learning; CART; cocoa beans classification; gini diversity index; ROC

1. INTRODUCTION

Cocoa is one of the major cash crops that contribute significantly to Nigeria's foreign exchange. It was the main agricultural stake of Nigeria's economy until the 1970's when crude oil was discovered in the country in commercial quantities (Fadipe et al., 2012). Since then, the production of cocoa witnessed a downward trend after 1971 season. Its export declined to 216,000 metric tons in 1976, and 150,000 metric tons in 1986, therefore reducing the country's market share to about 6% and the country to being the fifth largest producer (Folayan, 2010).

Cocoa has a unique natural taste, colour and delicious aroma used in many food products for extra flavour and colour. It is an important export commodity in the non-oil sector, there is the need to standardize method of its assessment for proper grading before exportation. The assessment would ensure consistency in the production of high-quality cocoa beans. This quality assessment is the responsibility of Federal Produce Inspection Service. They sample the cocoa lot randomly and carry out the Cut test according to international standards. The test reveals the percentage of defects present in the consignment by calculating the percentage of each class of defect in the 300 cocoa beans randomly sampled. This test involves cutting the cocoa beans longitudinally. Each cut bean is visually examined in full daylight or equivalent artificial light to determine its class. The quality of cocoa beans is mainly determined by its flavour and physical characteristics (CAOBISCO et al., 2015), which are revealed through this procedure. Among the tests carried out on cocoa beans are sensory test, moisture test, and beans count, but only the cut test reveals defects in samples which makes it a very important test for cocoa beans quality evaluation. It also provides a numerical score for the beans' quality. Colour is a major distinguishing factor between good fully fermented cocoa beans and the defective ones (Folayan, 2010). Slatty beans are blackish in colour and bitter to taste, mouldy beans are yellowish or may exhibit other odd colours asides brown (Niemenaka et al., 2014) while weevilly beans are partly whitish in colour and live insects may be seen. Only good cocoa beans are fully brown in colour. Figure 1 shows the distinguishing colours of different classes of cocoa beans obtained from this work.

This research only addressed the problem of misclassifications among these classes by critically analysing the process involved in the cocoa beans quality

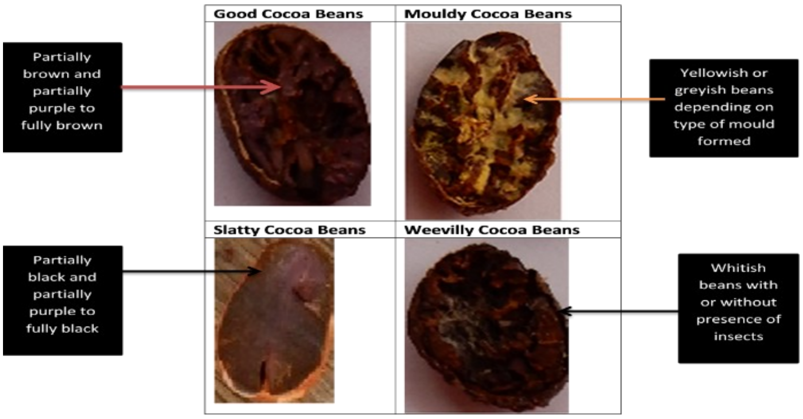

Figure 1. Cocoa Beans Internal Colours

evaluation. A decision tree algorithm was used to achieve the classification. There were two main steps strictly followed in building the tree: recursive partitioning of the feature (independent) variables, and pruning using the validation data. The accuracy of the algorithm is subject to the illumination conditions where the test is carried out. The objectives of this work were to analyse the process involved in cocoa beans quality evaluation; obtain images of cocoa beans and extract features therein; develop a model for the classification and evaluate the model.

2. RELATED WORKS

Although, an appreciable amount of work has been done in the area of fruit and produce classification, grading and quality assessment, research focus has not been on automatic classification of cocoa beans to identify defects using the cut test procedure. The design requirements for developing sorting or grading systems vary from fruit to fruit. Most of the research works focus on building dedicated systems that can sort a particular fruit or product type (Ohali, 2011). For instance, apple grading and sorting (Mizushima and Lu, 2013; Li and Zhu, 2011), Beans classification (Kilic et al.,2007), classification of corn tortillas (Mery et al., 2010), date grading system (Ohali, 2011), cocoa beans quality assessment (Olunloyo et al., 2011; Soto et al., 2015) and many others, each work on a particular kind of fruit or produce. Also, extensive research has reported the success and ground-breaking achievements of decision tree algorithms in fruit and produce classification and recognition. Unay et al. (2006) applied the decision tree method in apple stem and calyx recognition. Mercol et al. (2008) also used decision trees in the automatic classification of orange. Ground water quality classification by the decision tree method in Ardebil region of Iran was carried out by Saghebian et al. (2014), while Ahmed et al. (2015) published system availability enhancement using a computational intelligence-based decision tree predictive model.

Recently, Soto et al. (2015) proposed the use of hyper-spectral imaging in the analysis of cocoa beans. Thirty samples were examined. The images were acquired by a hyperspectral camera and segmentation by Principal Component Analysis (PCA). The internal colour and crack were extracted, and Fuzzy logic was used for classification as unfermented, partly fermented and well fermented. The classification was done prior to the drying of the beans. All the defects in the cocoa beans could not be represented because wet beans were used. Many defects show after drying. The quality of cocoa beans cannot be effectively obtained only through the examination and analysis of wet cocoa beans.

Olunloyo et al. (2011) developed a prototype electronic nose for monitoring the quality of cocoa beans. An artificial neural network was used for the classification. The sample was preheated at a certain temperature and humidity and the cocoa smell was acquired. The sample were classified into "Good (fruity flavoured)" or "Bad (off flavoured) based on sensory analysis". During the test, 38 out of 40 samples were correctly classified when compared with the results obtained from cut test. Defects present in the cocoa beans cannot be quantified with the electronic nose developed.

3. METHODOLOGY

3.1. Classification Stages

There are different stages involved to achieve a better classification model. The first is cocoa beans image capturing. The images of cut cocoa beans were manually captured with a digital camera. The second stage is processing the image and performing segmentation. The third involves extraction of features from the cocoa beans' image and the last stage is the classification.

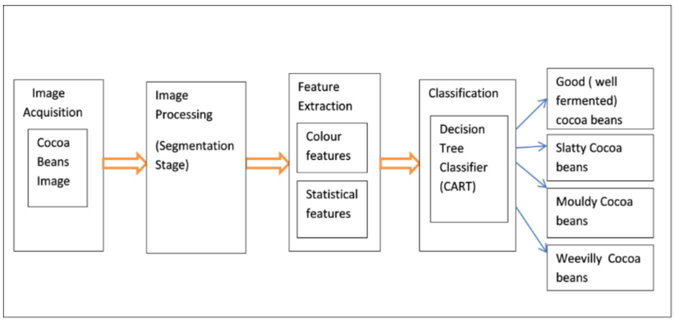

Figure 2. The Cocoa beans Classification stages

3.2. Image Acquisition

Samples of cut cocoa beans were captured with a Nikon Coolpix S9900 colour digital camera (CDC). The camera with 16 megapixels was placed vertically over a group of cocoa beans. The cocoa beans used in this research were obtained from Southwestern Nigeria during the main season of cocoa. Three hundred cocoa beans were analysed and divided into batches randomly for easy capture by the camera. Some batches contained 10 cocoa beans while some contained less. Each batch had 1152 x 864 pixels. The process of acquiring the images was done in a bright environment to prevent shadows around the images and to enhance the internal features of the seeds. These images were in the RGB colour space for easy processing. The shape of this colour space is cube with colours red, green and blue constituting the coordinate axis. Each component's values range from [0,255] though this is sometimes normalized to [0, 1]. The brightness and hue of the images are determined by the intensity of these different components. The images were saved in the .JPG format for further analysis. Samples of cocoa beans used in this work can be seen in Figure 3.

3.3. Image processing / segmentation

The acquired images were transferred to the MATLAB workspace which were then stored. The goal of the image processing was to improve the accuracy and speed of the feature extraction as well as to remove unwanted information such as background and artifacts. The region of interest (ROI) was the entire surface of the cut cocoa beans. Hence, there was need to separate the foreground (object) from the background. To achieve this, threshold-based segmentation was employed. This segmentation method was easy to implement. It was also fast because it was repeated on similar images. Two procedures were followed after the histogram of the grey scale image had been obtained:

i. A fixed value was estimated to filter the histogram and to the distinguish background from the object. A threshold value of 78 was used as determined during the preliminary work. This value is the Global threshold for the whole image. Any value above the set threshold was regarded as the background while values from 78 downward were assumed to be the object. In this process, it is admitted that some parts of the object may be eroded creating holes in them. Although the holes are minimal, they still need to be filled so as not to jeopardize the feature extraction process.

ii. A morphological operation was performed on the segmented binary image obtained in order to fill the possible holes in the image. The morphological operators employed for this operation were erosion and dilation. One of the factors that determines whether a background pixel has its values changed and becomes an object is the structuring elements. The element is an entity that determines which neighbourhood pixel around a given background is to be considered in order to change its value. The pixel constituting holes within connected components were identified and a flood-filling operation was performed. Their values were set below the fixed threshold value.

3.4. Feature extraction

The most common features in food applications include colour, texture, shape, and size. These features reveal the quality and defects of a produce or fruit. In this study, colour, texture, and statistical features were extracted from cocoa beans images in order to have a good description of the image and generate sufficient information to develop the model.

In the colour feature extraction, Mean analysis in the RGB colour space was performed. The mean, minimum and maximum values of each component of RGB were directly extracted from the RGB images. The Red, Green and Blue components were extracted into feature vectors. The Mean, Minimum and Maximum were then computed from the feature vectors for each component. Nine colour features were therefore extracted for each image.

Texture analysis was done using three statistical

Figure 3. Mouldy cocoa beans samples during cut test

descriptors, which use a co-occurrences grey level matrix as used by (Mercol et al., 2008). The three features are skewness, kurtosis, and entropy. In this study, all the descriptors were calculated in MATLAB using the appropriate functions available. Fourteen features were extracted from each image and saved in the database. The algorithm (Classification and Regression Tree) used in this classification automatically detects the most relevant and important features needed to perform the classification through its inbuilt feature selection capability. Table 1 shows part of the data obtained from the feature extraction.

3.5. Classification

A predictive model was formulated by associating the extracted features with the corresponding class, in this case Good, Mouldy, Slatty and Weevilly, as classified by the Federal Produce Inspectors by visual assessment. CART, which builds classification model in a tree structure using a divide-and-conquer strategy, was used in this research. This division is based on splitting criteria. The split is recursively performed until a bigger tree is created and the stopping criteria are satisfied. The Tree employed in this work is a binary tree and a univariate split. Being univariate implies, a node splits based on the value of a single attribute. MeanR was the first independent variable used to split the data into two parts as shown in Figure 5. The first part contained all the beans with MeanR < 91.296 and the other part contained those having MeanR >= 91.296. Each of these parts was divided in a similar way by choosing another variable or possibly the same variable used earlier with a different split value. This process was done recursively until the nodes became as pure or homogenous as possible. Being pure meant that the node contained cocoa beans that belonged to just one class. The next split in our tree was on MinR at value 11.5. Although, there are several splitting criteria available in the CART algorithm, in this research, the gini diversity index (gdi), twoing criterion for the twoing rule and deviance for maximum deviance reduction were tested. CART selects features that split the tree based on the values of their splitting criteria functions.

3.5.1. Gini Diversity Index (gdi):

The gini index of a node in this study is given by the equation below.

$$Gini\ Index = 1 - \sum_{i=1}^{c} P(i)^2 \qquad (1)$$

c is the number of classes which is 4 and Pi is the probability of an instance belonging to class i.

Gini index is based on the impurity measure. Hence, a feature which gives the minimum impurity on splitting the node is most considered. Splitting the dataset based on feature MeanR at value 91.296 gives the lowest gini index compared to splitting with other features.

$$Gini\ index\ for\ the\ dataset = 1 - [\left(\frac{83}{300}\right)^2 + \left(\frac{67}{300}\right)^2 + \left(\frac{79}{300}\right)^2 + (\frac{71}{300})^2]1 - 0.251$$

$$Gini\ index = 0.749$$

Here, 300 is the total number of instances in the sample

83 is the number of instances that belongs to class 1

67 is the number of instances that belongs to class 2

79 is the number of instances that belongs to class 3

71 is the number of instances that belongs to class

Splitting a node based on MeanR, we obtain the number of instances that split to the right and those that split to the left from each class to get the probability of an instance belonging to left or right.

This was repeated for all features and each node was split based on the features with the smallest gini values.

3.5.2. Twoing:

The twoing splitting function is given by the equation

$$Twoing = P(l)P(r)[\sum_{i=1}^{c} l(i) - r(i)]^2 \qquad (2)$$

P(l) is the probability of instances that split to the left while P(r) is the probability of instances that split to the right. This function was calculated for all features and the results were compared to select the features that best

Table 1: Part of the data obtained during feature extraction

minR	maxR	meanR	minG	maxG	meanG	minB	maxB	meanB	standard dev	entropy	skewness	variance	ku
5	255	84.66401	0	255	65.05724	0	255	255	77.89983596	6.5904031	0.926568058	6068.3844	2.
4	255	89.32933	0	255	63.82803	0	255	255	69.82509995	6.9887954	1.032719715	4875.5446	2.
10	255	81.49455	0	255	55.2617	0	255	255	74.37707141	6.65493	1.332363256	5531.9488	3.
68	255	209.0579	28	255	167.7983	2	255	255	61.8877097	5.6273446	-0.01272898	3830.0886	1.
5	255	68.40974	0	255	43.17528	0	255	255	62.22954157	6.5326071	1.488281037	3872.5158	4.
5	187	66.60118	0	158	37.27342	0	158	125.519	48.36029706	6.7920005	1.10995979	2338.7183	2.
5	184	69.10207	0	158	47.38828	0	161	135.3924	57.35123717	6.6236078	0.847376924	3289.1644	2.
3	255	83.988	0	255	51.79905	0	255	255	60.05160718	6.9885021	0.920048626	3606.1955	2.
3	198	67.7792	0	169	42.53526	0	168	135.6329	57.3070685	6.622282	0.988946169	3284.1001	2.
3	255	64.80402	0	255	46.32139	0	255	255	61.54003095	6.4711574	1.010456924	3787.1754	2.
6	211	88.21266	0	176	59.64946	0	177	165.3797	62.7122396	7.0687115	0.544725914	3932.825	1.
8	255	91.15786	0	255	59.17714	0	255	167.0125	62.74155084	7.1204439	1.052218505	3936.5022	3.
5	202	74.78436	0	175	49.98725	0	179	166.5949	60.1686664	6.8375852	0.772670214	3620.2684	2.
5	255	79.50188	0	255	53.62315	0	255	255	65.16184468	6.8073993	0.89177989	4246.066	2.
10	255	87.26345	0	255	56.2638	0	255	255	66.60922704	6.9122046	1.063151514	4436.7891	3.
8	255	75.58371	0	255	54.55634	0	255	255	64.24959326	6.7480642	0.946610888	4128.0102	2.
8	255	80.77057	0	255	59.97444	0	255	255	71.28062003	6.7277163	1.052072024	5080.9268	2.
13	255	82.92288	0	255	58.97145	0	255	255	63.71040257	6.988668	0.775909991	4059.0154	2.
5	255	85.47875	0	255	56.95441	0	255	255	64.23495257	7.0476315	0.873573003	4126.1291	2.

split the node.

3.5.3. Deviance:

Deviance, also called cross entropy, is given by

$$Deviance = -\sum_i P(i) log P(i) \quad (3)$$

P(i) is the fraction of class i with total instances that reached the node.

If the deviance obtained from the calculation is zero it shows that the node is pure. Otherwise, the calculation will be positive. A pure node does not need to be split any further as it is a leaf of the tree. The values obtained from these functions were used in building our decision tree for the classification. The values from each splitting criterion were ranked for the selection of features to split the decision tree. Since different values were obtained from the three criteria, different tree structures were also generated as depicted in the Figure 5, Figure 6 and Figure 7 below.

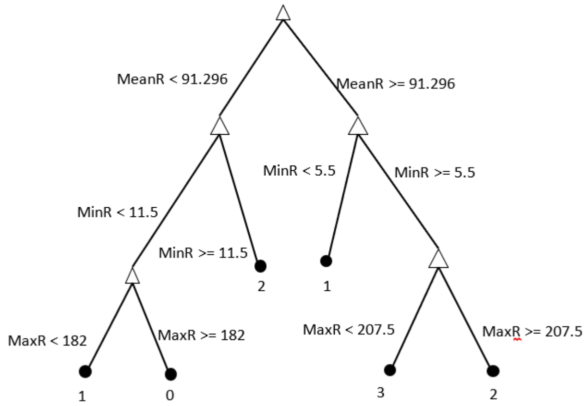

Figure 5. The cocoa beans decision tree using GDI splitting criteria

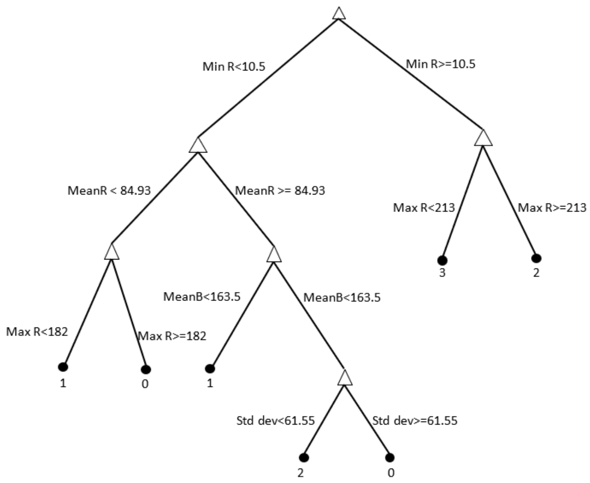

Figure 6. The cocoa beans decision tree using twoing

The 300 samples of cocoa beans were manually classified served as training and validation datasets. The Table 1 shows the number of cocoa beans for each class that makes up the sample.

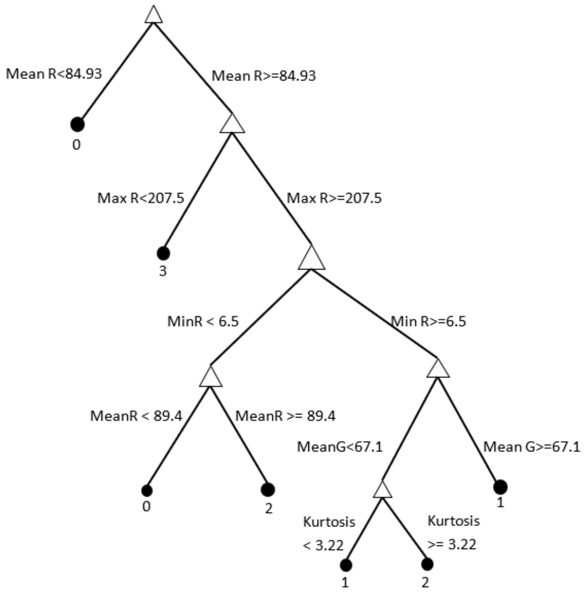

Figure 7. The cocoa beans decision tree using deviance

Table 1. Distribution of defects (classes) in the dataset

Class code	Class name	Number in sample
0	Good (well fermented) cocoa	83
1	Slatty cocoa beans	67
2	Mouldy cocoa beans	79
3	Weevilly cocoa beans	71

The MATLAB *classregtree* function was used to construct the tree using the three splitting criteria. The samples were randomized and divided into 10 folds each containing 30 samples. Nine folds were used for training and the remaining fold for validation. This process was repeated 10 times such that all folds passed through training and validation.

During training, the classifier was supplied with the feature database, which contain the file names of images to be classified, the features extracted earlier, and the class that each image belongs. The classes were determined by the federal produce inspectors responsible for the quality assessment of cocoa beans. The training of the classifier was based on the judgment of the inspectors.

After training, the model was tested with set of unknown cocoa images (another set of cocoa beans images were obtained for this purpose). Here, the database contained only the images to be classified described by their individual set of features extracted. The model was not supplied with the classes of these images.

4. RESULTS AND DISCUSSION

Different classification results were obtained from the three splitting criteria used. Table 2 shows the results obtained from the three criteria against the actual samples classified by the inspectors. From the evaluation of the three splitting criteria, it was observed that gdi slightly outperformed the other two and therefore were employed in building the proposed system.

Table 2. Classification Results obtained from the three splitting criteria

Class	GDI	Twoing	Deviance	Actual data Sample
Good	79	72	76	83
Slatty	70	75	77	67
Mouldy	75	80	73	79
Weevilly	76	73	74	71

It must be reported that five federal produce inspectors were involved in the manual classification and there were times of disagreement on the classes of some cocoa beans. This was as a result of the cocoa beans being in-between the boundary of two different classes. The most common of this instance was deciding whether a particular cocoa bean was slatty or good. The final class of such cocoa beans was determined by a simple voting majority. Hence, there may be some misclassification both from the inspectors and the newly developed system as a result of this especially in the class good and the class slatty.

4.1. Model Evaluation

The results of classification obtained from the three splitting criteria tested in this work were compared with that of the federal produce inspectors and the classification accuracy of each was calculated based on the equation below

$$Accuracy = \frac{Number\ of\ correctly\ classified\ Samples}{Total\ number\ of\ samples\ classified} \times 100 \quad (4)$$

Different average accuracies were obtained, as shown in Table 3, both in training and testing after running the code several times. It can be seen from the table that the average classification accuracies of the model using the three splitting criteria were not significantly different from one another but the maximum accuracy of 89.2% was recorded by the gdi which is a default splitting criteria for CART.

Table 3. Average accuracy of the classifier using different splitting criteria

Splitting criteria	Training (%)	Testing (%)
GDI	92.3	89.2
Twoing	93.1	86.7
Deviance	91.7	88.3

In order to closely evaluate the model, the ROC of each class was plotted and the area under the ROC curve (AUC) were calculated. The ROC is a graph of sensitivity against False alarm (1-specificity).

Sensitivity or recall is the true positive rate and is given by

$$Sensitivity = \frac{TP}{TP+FN} \quad (5)$$

where TP = True Positive, and FN = False Negative

Specificity is the true negative rate

$$Specificity = \frac{TN}{TN+FP} \quad (6)$$

where TN = True Negative, and FP = False Positive

False alarm or False Positive rate (1-specificity) is

$$False\ Positive\ Rate = \frac{FP}{Actual\ Negative} \quad (7)$$

The ROC curves were constructed from the values of sensitivity and false positive rate obtained for each of the group as depicted in the following tables and figures.

Table 4: Sensitivities and False positive rates for (a) Class Good, (b) Class Slatty, (c) Class Mouldy, and (d) Class Weevilly

(a)

Sensitivity	False Positive Rate
0	0
0.0357	0
0.8571	0.0238
0.9643	0.0952
1.0000	0.1667
1.0000	0.5952
1.0000	0.7143
1.0000	0.7857
1.0000	1.0000

(b)

Sensitivity	False Positive Rate
0	0
0.7500	0
0.8750	0
1.0000	0.0217
1.0000	0.5435
1.0000	0.6522
1.0000	0.7826
1.0000	0.8043
1.0000	1.0000

(c)

Sensitivity	False Positive Rate
0	0
0.7500	0
1.0000	0.0517
1.0000	0.4655
1.0000	0.7759
1.0000	0.8621
1.0000	0.9138
1.0000	0.9828
1.0000	1.0000

(d)

Sensitivity	False Positive Rate
0	0
0.8333	0
1.0000	0.3594
1.0000	0.6406
1.0000	0.6875
1.0000	0.7813
1.0000	0.8438
1.0000	0.8594
1.0000	1.0000

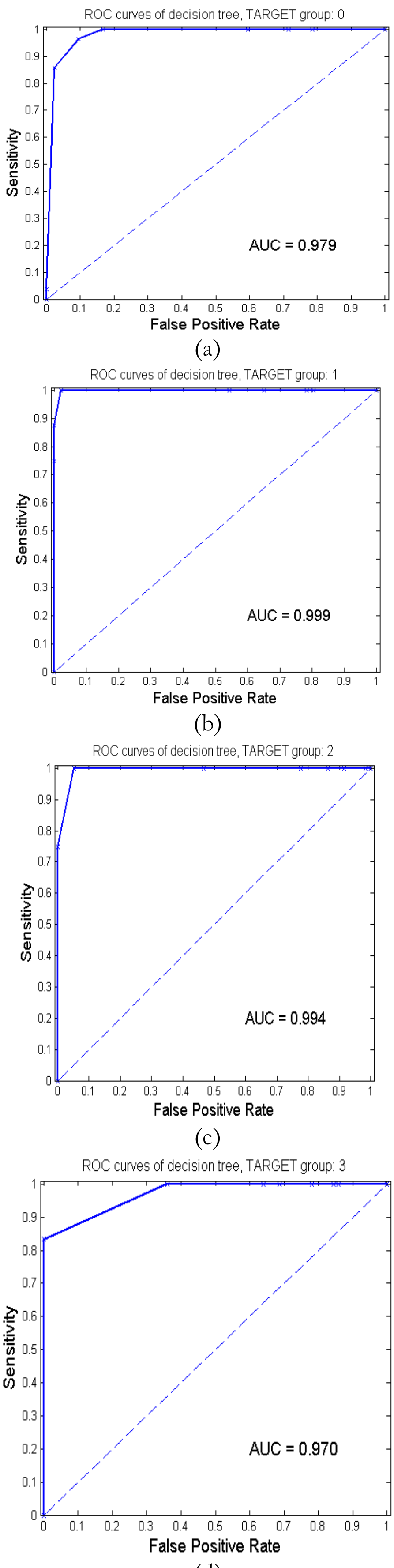

Figure 11. The ROC curves for (a) Class Good (b) Class Slatty (c) Class Mouldy (d) Class Weevilly.

The AUC measures the accuracy of the prediction for each class and was calculated by constructing trapezoids under the curve as an approximation of the area covered by the curve. The AUC values obtained for each class were not significantly different except for class weevilly with a little lower AUC. AUC values range from 0 to 1 with maximum classification accuracy being 1.0. A random guess for any model prediction is 0.5 AUC. The closer the curve is to the left-hand border and then the top border of the ROC space, the more accurate the prediction for that class is. The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the prediction is.

5. CONCLUSION

In this work, a sample of dried cocoa beans was collected during the peak season from local buying agents. The beans were cut and pictures were taken for analysis. A decision tree algorithm was used to formulate a model that classified cocoa beans into good, slatty, mouldy, and weevilly beans, and reveal the percentage of defects present in a sample. Gini's diversity index (gdi), twoing and deviance were the three splitting criteria tested in this research and the following accuracies were recorded respectively 89.2%; 86.7%; 88.3%. The model was developed in MATLAB for the classification and feature extraction. The results of this research can be utilized by cocoa exporters and the local buying agents to check for the percentage of defects present in the cocoa. This work can be improved upon by automating the process of cocoa image acquisition.

ACKNOWLEDGEMENT

We appreciate the effort of Federal Produce Inspectors, Osun State in the classification of the cocoa beans. The cocoa beans were supplied by the local buying agents.

REFERENCES

Ahmed, Q., Anifowose, F. A., and Khan, F. (2015). System availability enhancement using computational intelligence-based decision tree predictive model. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 229(6):612-626.

CAOBISCO, ECA, and FCC (2015). Cocoa Beans: Chocolate and Cocoa Industry Quality Requirements. ECA-Caobisco-FCC Cocoa Research Fund www.cocoaquality.eu ISBN: 978-2-9601817-0-8.

Fadipe A. E., Adenuga A. H. and Ilori, T. E. (2012) Economic Analysis of Cocoa Production In Oyo State, Nigeria. Nigerian Journal of Agriculture, Food and Environment. 8(4):58-63.

Folayan, J. (2010). Nigerian cocoa and cocoa by-products: quality parameters, specification and the roles of stakeholders in quality maintenance. Pakistan Journal of Nutrition, 9(9):915-919.

Kilic, K., Boyaci, I. H., Koksel, H., and Kusmenoglu, I. (2007). A classification system for beans using computer vision system and artificial neural networks. Journal of Food Engineering, 78:897-904.

Li, X. and Zhu, W. (2011). Apple grading method based on features fusion of size, shape and color. Procedia Engineering, 15:2885-2891.

Mery, D., Chanona-perez, J. J., Soto, A., Miguel, J., Cipriano, A., Vel_ez-rivera, N.,Arzate-v_azquez, I., and Guti_errez-l_opez, G. F. (2010). Quality classification of corn tortillas using computer vision. Journal of Food Engineering, 101(4):357-364.

Mercol, J. P., Gambini, M. J., and Santos, J. M. (2008). Automatic classification of oranges using image processing and data mining techniques. In XIV Congreso Argentino de Ciencias de la Computacion.

Mizushima, A. and Lu, R. (2013). An image segmentation method for apple sorting and grading using support vector machine and Otsu's method. Computers and Electronics in Agriculture, 94:29-37.

Niemenaka, N., Eyamoa, J. A., Onomo, P. E., and Youmbi, E. (2014). Physical and chemical assessment quality of cocoa beans in south and center regions of Cameroon. Syllabus Review Science Series, 5:27-33.

Ohali, Y. A. (2011). Computer vision based date fruit grading system: Design and implementation. Journal of King Saud University - Computer and Information Sciences, 23(1):29-36.

Olunloyo, V. O. S., Ibidapo, T. A., and Dinrifo, R. R. (2011). Neural network based electronic nose for cocoa beans quality assessment. Agricultural Engineering International: CIGR Journal, 13(4):1-12.

Saghebian, S. M., Sattari, M. T., Mirabbasi, R., and Pal, M. (2014). Ground water quality classification by decision tree method in Ardebil region, Iran. Arabian Journal of Geosciences, 7(11):4767-4777.

Soto, J., Granda, G., Prieto, F., Ipanaque, W., and Machacuay, J. (2015). Cocoa bean quality assessment by using hyperspectral images and fuzzy logic techniques. In Twelfth International Conference on Quality Control by Artificial Vision 2015, volume 9534, pp 95340M. International Society for Optics and Photonics.

Unay, D., Gosselin, B., and Debeir, O. (2006). Apple stem and calyx recognition by decision trees. In Proceedings of the 6th IASTED International Conference on Visualization, Imaging, and Image Processing, VIIP, pp. 549-552.