

# Supervised Methods Applied to the Construction of a Vision System for the Classification of Cocoa Beans in the Cut-Test

Felipe A. Santos<sup>1</sup>, Anne M. P. Canuto<sup>1</sup>, Benjamin R. C. Bedregal<sup>1</sup>,  
Eduardo S. Palmeira<sup>2</sup>, Iaslan N. P. Silva<sup>1</sup>

<sup>1</sup>Departamento de Informática e Matemática Aplicada  
Universidade Federal do Rio Grande do Norte (UFRN)

<sup>2</sup>Departamento de Ciências Exatas e Tecnológicas  
Universidade Estadual de Santa Cruz (UESC)

fante@ppgsc.ufrn.br, {anne,bedregal}@dimap.ufrn.br,  
espalmeira@uesc.br, iaslannascimento@gmail.com

**Abstract.** *Supervised machine learning methods have been widely used in the literature for many classification tasks. In this context, some aspects of these algorithms, as the attributes used and the form they were built, have a direct impact in the system performance. Therefore, in this paper, we evaluate the application of classification algorithms, along with attribute selection, to study an improved version of a computer vision system that performs the classification of cocoa beans. The main aim of this investigation is to improve the performance of a cocoa classification system that aims to help farmers to classify the different cocoa beans based on images of these beans.*

## 1. Introduction

Cocoa seed is a product widely used in chocolate fabrication. Regulative norms around the world dictates several quality tests which those beans need to be subjected to in order to be commercialized internationally. One of those tests is the Cut-Test technique, in which a fixed amount of seeds are opened and their interior is analyzed by specialists to characterize their coloration, compartmentalizing, defects and other external characteristics [Brasil 2008, Catsberg and Dommelen 1990].

The Cut-Test technique is made manually in most parts of the world and, specifically in Brazil, only the manually characterization of cocoa beans characteristics is regimented. In this context, since the Cut-Test technique is made by human eyes and it uses linguistic variables to describe the outcome (e.g.: violet and partially violets beans), the evaluation made by experts are subjected to imprecision and discordance, resulting in cases where experts working at the same environment and conditions may provide different classifications for the same bean, without considering human emotional factors, which may also bias the classification process [Brasil 2008, Catsberg and Dommelen 1990, Santos 2019, Savakar 2012, Majumdar et al. 1996].

In [Santos 2019], the authors, observing these imprecision and discordance, created seven Cut-Test taxonomy for cocoa beans based on image-based cocoa dataset. Additionally, they proposed a computational vision system that aimed to help experts in this

classification process. In this proposed model, a Multi-layer Percetron neural network (MLP) was evaluated, varying only the number of neurons and hidden layers. Finally, the authors suggested the investigation of other classification methods and a more refined analysis of the MLP performance in order to prevent over-fitting and to potentially provide a better performance in the classification process.

Following the suggestions made by the authors in [Santos 2019], in this paper we will perform an extensive investigation of the use of classification methods in the cocoa classification process. In order to do this, we selected four classification algorithms, which are:  $k$ -NN, MLP, Naive Bayes and Decision Three. In addition, we intend to provide a more refined analysis of the used MLP model. These algorithms were chosen because they are simple, efficient and apply different classification criteria in their hypothesis. Decision tree, for instance, builds a classification tree during its training and use it during the test procedure, while  $k$ -NN is an instance-based algorithm. Finally, a feature selection technique will be applied to all classification algorithms to evaluate the performance of the cocoa classification task using a more compact representation of the cocoa images.

To this end, this article is organized in the following form: in Section 2.3 we will explore the features firstly proposed in [Santos et al. 2018], and perform an attribute selection using Decision Tree and Person Correlation; in Sections 3.2.1, 3.2.2 and 3.2.3 we will analyze the use of  $k$ -NN, Decision tree and Naive Bayes, respectively, to perform the classification using the features selected; in Section 3.2.4 we will analyze the variation of other parameters of the MLP to the construction of a classification model that does not present over-fitting; in Section 4, we will select the best configuration attained in this study to improve the system proposed in [Santos 2019] and propose potential continuations.

## 2. Material and Methods

In this section, we will present the methods and materials of the empirical analysis conducted in this paper. First, the system architecture will be briefly described, followed by a description of the used image dataset. Finally, the used feature selection techniques is described.

### 2.1. System Architecture

The system architecture is presented in Figure 1. As it can be observed in this image, the proposed system is composed of three phases, being the first one the feature acquisition, in which some features are extracted from the cocoa bean images. The feature extraction process will be described in Section 2.3.



Figure 1. System Architecture

Once the cocoa features were extracted, the second phase is the feature selection, which will select the most influential image features for the classification. The main aim is to have a more compact representation of the cocoa beans. In this paper, two feature selection methods, Decision Tree and Pearson Correlation Coefficient, were used in a

combined way in a wrapper approach. Details of each selection method will be presented in Section 2.4.

The third phase comprises the classification one, in which the cocoa bean features are classified into one Cut-Test class. In this paper, four supervised classification models will be used, which are: Decision Tree, Naive Bayes, k-NN and MLP. Those models were chosen due to their wide literature application in classification problems [Samaniego et al. 2008, Im and Jensen 2005, Yang et al. 2003, Haralick et al. 1973].

## 2.2. Image Dataset

The image dataset used in this research is composed of 1400 beans images, along with their respective Cut-Test classification, divided into 14 different classes, which was made by experts. The Cut-Test classes considered are: Agglutinated, Brittle, Compartmentalized Brown, Compartmentalized Purple, Compartmentalized Partially Purple, Compartmentalized White, Compartmentalized Slatty, Flattened, Moldered, Plated Brown, Plated Purple, Plated Partially Purple, Plated White and Plated Slatty [Brasil 2008, Catsberg and Dommelen 1990, Santos 2019]. This dataset was originally proposed in [Santos 2019] and is hosted in <http://nbcgib.uesc.br/teais>, under the name of "Sem Plano de Fundo - Versão 1 - Método A". Two samples of images of this base, as well as their corresponding classifications, can be seen in Figure 2. In this figure, a sample of a Compartmentalized Purple and a Plated White bean are presented.



**Figure 2. Samples of the dataset: a Compartmentalized Purple (left) and a Plated White (right) bean**

It is important to enhance that each bean was firstly classified by two specialists and, if the two of them agreed in the classification, then it was registered in the original dataset. In addition, this dataset and the algorithms used to extract the image feature are the same as used in [Santos 2019], this was done to isolate variables in algorithms and to enable the comparison the results obtained in this paper and the ones obtained in [Santos 2019]. Finally, the dataset images are presented in JPG format, with 4320x3240 pixels [Santos et al. 2019].

## 2.3. Image Features

In order to create the original model, three sets of features were extracted, textural, color and structural features.

- Textural features: the 14 Haralick's Textural Features were chosen as textural features. Haralick's textural features are a set of characteristics from an image that can be used in classification models, including food-related ones [Savakar 2012, Patil et al. 2011, Haralick et al. 1973]. Those features represent image aspects such as homogeneity of color, contrast and linear dependency between pixels [Haralick et al. 1973];
- Color features: the means of each the RGB layer and of the grayscale-converted image were used as color features;
- Structural features: the area and the perimeter of the bean were selected as structural features.

Regarding the dataset used in this paper, it has a total of 20 features. In comparison with the dataset proposed in [Santos 2019], we included the perimeter of the bean images. In order to map the characteristics used in the cocoa dataset, the following description is used, H1 to H14 represent the 14 Haralick's features, in the same order of [Santos et al. 2018]. Additionally, R Mean, G Mean, B Mean, GS Mean are the average values of the RGB and Grayscale layers, respectively, and Area and Perimeter represent the area and the perimeter of the bean images. Those features were chosen by analyzing similar researches in the image processing literature. Once the image dataset is defined, two different feature selection techniques are combined and applied to the cocoa images and they are described in the next subsection.

## 2.4. Feature Selection Techniques

In order to perform the feature selection step, a combination of Decision Tree (DT) and Pearson Correlation Coefficient (PCC) is used. The DT is applied aiming at obtaining the most representative features (information gain) by ranking these features according to the different level of the obtained decision tree. In other words, in analyzing the features of the first levels, they potentially provide the features that are important for further classification and, through an analysis of the obtained results, it is also possible to discard features that will have a modest contribution in the classification step.

The procedure of providing a Decision Tree and using it as feature selector can be described as follows.

1. The set of features available forms the input to the DT algorithm;
2. The Decision Tree has leaf nodes, which represent class labels;
3. The branches of the tree represent each possible value of the feature node from which they originate;
4. The Decision Tree can be used to classify feature vectors by starting at the root of the tree and moving through it until a leaf node, which provides a classification of the instance, is identified;
5. At each decision node in the Decision Tree, one can select the most useful feature for classification using appropriate estimation criteria. The criterion used to identify the best feature of the original dataset, based on an estimation criteria. In this paper information gain is used as estimation criteria [Fayyad and Irani 1992, Hall 2000].

Pearson's correlation coefficient is a statistical measure that evaluates the power of a linear relationship between two data variables (attributes). Therefore, suppose we have

two variables  $X = \{x_1, \dots, x_n\}$  and  $Y = \{y_1, \dots, y_n\}$ , both containing  $n$  values. Then, PC is given by the equation 1:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

Where  $\bar{x}$  and  $\bar{y}$  are the average values for variables  $X$  and  $Y$ , respectively.

The PC coefficient ( $r$ ) may to assume the following values:  $-1 \leq r \leq 1$ , in which: 0 denotes no linear correlation; and The closer the value is to  $-1$  or  $1$ , the stronger the correlation is, where negative values represent inverse correlation and positive values represent direct correlation.

The main aim of this feature selection step is to apply both techniques, DT and PCC, in a wrapper manner, and to select the feature subset which is common on both technique.

### 3. Results and Discussion

In this section, the obtained results of the empirical analysis will be presented, starting with the feature selection step, followed by the results of the classification algorithms.

#### 3.1. The feature selection method

The first feature selection technique was a Decision Tree.

The features that appeared in the different levels the DT were mapped, in a cumulative way per level. Each level contained all features that appeared in the tree until said level [Hall and Holmes 2003, Quinlan 1986, Fayyad and Irani 1992, Hall 2000, Al-Marakeby et al. 2013]. The feature distribution is the following.

- 1st Level: H6;
- 2nd Level: H6, H8 and H13;
- 3rd Level: H6, H8, H12, H13, R Mean, GS Mean and Area;
- 4th Level: H6, H8, H12, H13, R Mean, GS Mean, Perimeter and Area;
- 5th Level: H1, H2, H5, H6, H8, H12, H13, R Mean, GS Mean, Perimeter and Area;
- 6th Level: H1, H2, H3, H4, H5, H6, H8, H12, H13, H14, R Mean, B Mean, GS Mean, Perimeter and Area.

In order to define the most representative features, a wrapper-based procedure is adopted. In this procedure, two classification algorithms, Naive Bayes (NB) and  $k$ -NN, with odds values for  $k$  varying from 1 to 7, all with 10-fold cross validation, were applied using the features obtained at each level of the decision tree. The obtained results are presented in Table 1. This table presented the average accuracy level over all 10 folds in the cross validation process. The bold numbers in this table represent the highest accuracy level for each classification algorithm (column).

In analyzing Table 1, we can observe that the accuracy levels did not increase as the number of features increases, since the highest accuracy levels were obtained between the third and sixth level. An important aspect to observe is that the use of a feature subset

**Table 1. Accuracy levels using Naive Bayes, k-NN and DT for the different levels of the decision tree feature selector**

Level	DT	NB	k-NN (1)	k-NN (3)	k-NN (5)	k-NN (7)
1st	12.6429%	19.000%	11.2857%	12.1429%	12.4286%	14.5714%
2nd	26.9286%	32.5714%	25.7143%	28.2143%	32.3571%	33.5000%
3rd	<b>50.0000%</b>	42.9286%	51.2857%	49.0000%	49.5000%	49.5714%
4th	49.5714%	46.4286%	50.5714%	50.0714%	52.0000%	52.1429%
5th	<b>50.0000%</b>	46.7857%	53.4286%	<b>53.2857%</b>	54.2143%	<b>54.1429%</b>
6th	48.8571%	<b>47.3571%</b>	<b>54.0714%</b>	53.0000%	<b>54.3571%</b>	53.7143%
All features	49.1429%	46.0000%	52.7857%	51.1429%	53.2857%	52.8571%

improved the performance of the classification system, for all classification algorithms, since the bold numbers are always achieved by a feature subset.

The PCC measure was also applied to rank the image features. For PCC, all features are analyzed, using Eq.(1), in comparison with the class label. The idea is to assess the correlation of the features over the class labels (classification decision). Then, the features are ranked, in a descending order, as presented in Table 2.

**Table 2. Pearson Correlation Ranking**

Ranking	Attribute	Ranking	Attribute
1°	H6	11°	B Mean
2°	H7	12°	Perimeter
3°	H4	13°	GS Mean
4°	H13	14°	H11
5°	H8	15°	H10
6°	H12	16°	H2
7°	G Mean	17°	H1
8°	H3	18°	H5
9°	R Mean	19°	H14
10°	Area	20°	H9

The results provided in Table 3 are very promising. It can be observed that the highest ranked attribute is the same of the first level at the DT. Additionally, when we observe the results obtained by PCC with the third level of the DT (Table 1), that contains seven attributes, more than half of them are present in the seven best-ranked attributes of PCC (Table 2). It shows that there is an agreement between the feature selection techniques used in this paper.

In order to validate the results obtained by PCC, a similar procedure to the one made with DT is applied to the ranking obtained in Table 2. All features that appeared until a specific level of PCC are used as feature subset and used as input for Naive Bayes and k-NN classification algorithms. The classification algorithms have the same parameter setting of the DT procedure, generating the results showed in Table 3. Once again, the bold numbers represent the highest accuracy level for each classification algorithm (column).

As it can be observed in Table 3, the highest accuracy levels were obtained using around 13 attributes, for the majority of classification algorithms. Once again, it shows

**Table 3. Pearson Correlation level and the Naive Bayes, k-NN and DT results for the classification of the attributes**

Level	DT	NB	k-NN (1)	k-NN (3)	k-NN (5)	k-NN (7)
1°	12.6429%	19.000%	11.2857%	12.1429%	12.4286%	14.5714%
2°	20.2857%	19.6429%	19.5000%	19.5714%	23.2857%	23.5000%
3°	21.0000%	19.6429%	20.5714%	19.3571%	23.4286%	22.7143%
4°	25.5714%	26.9286%	26.2143%	25.5714%	26.1429%	27.7143%
5°	28.9286%	31.8571%	27.9286%	28.5000%	31.5000%	33.2143%
6°	28.8571%	31.9286%	27.9286%	27.5000%	30.5000%	32.3571%
7°	33.6429%	36.0714%	34.9286%	36.1429%	37.2857%	38.4286%
8°	35.7857%	36.2143%	38.1429%	39.0714%	41.5000%	41.8571%
9°	39.6429%	37.2857%	39.5714%	39.5714%	42.5000%	42.7857%
10°	49.5714%	42.3571%	50.8571%	49.4286%	50.2857%	51.2857%
11°	52.0000%	43.0000%	52.2143%	50.7857%	51.5714%	51.6429%
12°	49.7857%	45.0000%	52.6429%	53.5000%	53.7143%	54.3571%
13°	53.1429%	45.7143%	<b>56.0714%</b>	<b>55.6429%</b>	<b>56.8571%</b>	<b>56.7143%</b>
14°	<b>53.5714%</b>	45.5000%	54.5714%	53.0000%	54.4286%	54.5000%
15°	51.0000%	<b>47.7143%</b>	54.4286%	54.7857%	56.4286%	56.1429%
16°	50.8571%	46.7857%	54.1429%	54.0000%	56.0000%	56.2143%
17°	49.2857%	47.5000%	55.3571%	55.1429%	56.1429%	56.5714%
18°	49.0000%	47.3571%	55.2143%	54.3571%	56.4286%	56.0714%
19°	49.0000%	47.1429%	53.7143%	52.2857%	53.2857%	53.2857%
20°	49.1429%	46.0000%	52.7857%	51.1429%	53.2857%	52.8571%

that the use of a more compact representation of the cocoa image can lead to more efficient classification systems.

Finally, we need to select the most representative features to compose the feature subset to be evaluated in the next subsection. In observing Tables 1 and 3, we can see that the first 13 features in Table 3 are usually selected in the first levels of the decision tree. Therefore, we decided to select 13 highest ranked features of PCC as the feature subset to be used in the cocoa classification system.

### 3.2. Classification Methods

As mentioned previously, four supervised learning classification methods are used in this empirical analysis, which are: k-NN, Naive Bayes, Decision Tree and Multi-Layer Perceptron Neural Network. Those methods were select due to their wide use in the literature for classification problems, including image-based ones, providing several other studies for comparison purposes [Warfield 1996, Samaniego et al. 2008, Im and Jensen 2005, Yang et al. 2003].

#### 3.2.1. k-NN

k-NN methods were applied to eight different numbers of neighbour, being those: k=1, 3, 5, 7, 9, 11, 13 and 15. Also, these classifications were also evaluated under two different weighted distance procedures, considering 1/distance and 1-distance as weights, as showed in Table 4.

**Table 4. k-NN non-normalized attributes**

k-NN	k values			
	1	3	5	7
Non-Weighted	56.0714%	55.6429%	56.8571%	56.7143%
Weighted (1/distance)	56.0714%	57.5714%	57.9286%	58.2857%
Weighted (1-distance)	56.0714%	57.3571%	57.7143%	58.0714%
k-NN	k values			
	9	11	13	15
Non-Weighted	56.0000%	56.8571%	57.2143%	55.7857%
Weighted (1/distance)	58.3571%	57.9286%	<b>58.9286%</b>	58.1429%
Weighted (1-distance)	57.9286%	57.7143%	58.8571%	57.7857%

When observing Table 4, we can see that, for all analyzed number of neighbours, the weighted distance procedures delivered similar or higher accuracy levels than the non-weighted versions. Additionally, the increase of  $k$  did not necessarily caused an improvement the accuracy level of  $k$ -NN, as observed with  $k = 1$  and  $k = 3$ , in the non-weighted version, that resulted in a decrease of the accuracy level from 56.0714% to 55.6429%. The same pattern of behavior can be seen in weighted distance procedures, with 1/distance for  $k = 9$  to  $k = 11$ , that decrease in the accuracy level from 58.3571% to 57.9286%. In general, the highest accuracy level was delivered by one weighted procedure (1/distance), with  $k = 13$ , resulting in a accuracy of 58.9286%.

Once the different weighted procedures were evaluated, the impact of normalization will be assessed and the obtained results are illustrated in Table 5.

**Table 5. k-NN normalized attributes**

k-NN	K values			
	1	3	5	7
Non-Weighted	49.2857%	48.6429%	50.1429%	49.5000%
Weighted (1/distance)	49.2857%	50.0714%	<b>50.9286%</b>	49.8571%
Weighted (1-distance)	49.2857%	49.4286%	50.4286%	49.7857%
k-NN	K values			
	9	11	13	15
Non-Weighted	49.5000%	48.2857%	47.4286%	46.2143%
Weighted (1/distance)	49.8571%	49.0000%	48.2143%	47.7143%
Weighted (1-distance)	49.7143%	49.0714%	48.2857%	47.5714%

The attribute normalization, as showed in the differences of the accuracy levels presented in Tables 4 and 5, did cause a decrease in the accuracy levels of the  $k$ -NN method. We believe that this occurred due to a numerical representation problem. In the cocoa dataset, many attribute values have very large numbers, with six or more decimal numbers, with values varying in the order of  $10^3$ . In this sense, when calculating distances of normalized numbers, it will result in a very low number, with many decimal digits, and the use of rounding operations that can cause a numerical representation problem.



### 3.2.2. Decision Three

For this classification method, four DT versions are evaluated, with and without pruning for the non-normalized and for the normalized dataset, as showed in Table 6.

**Table 6. Decision Tree Results**

	Non-normalized		Normalized	
	Non-Pruned	Pruned	Non-Pruned	Pruned
Accuracy	<b>53.2143%</b>	53.0714%	50.1429%	48.8571%
Tree Size	495	473	721	360
Laves Number	248	237	538	483

It can be observed that, for the non-normalized dataset, the pruning strategy has caused a relative decrease of accuracy level, less than one percentage point. On the other hand, the size of the provided tree and the number of leaves decreased approximately 4.5%, suggesting an improvement in the pruned tree. For the normalized attributes, the accuracy level decreased more steadily than for the non-normalized dataset, approximately 1.3 percentage points. However, the tree size decrease was almost 50.0% while the decrease of the leaves number was 10.3%, suggesting that this tree was over-fitted before the pruning strategy.

### 3.2.3. Naive Bayes

For the Naive Bayes methods, three different versions will be evaluated, using the normalized and the non-normalized values, one with the standard NB configuration, one with the use of Kernel Estimator instead of the normal distribution and one with the use of a supervised discretization procedure instead of using the normal distribution, generating the results showed in Table 7.

**Table 7. Naive Bayes results**

	Non-normalized	Normalized
Standard	45.6429%	<b>45.7857%</b>
Kernel Estimator	44.8571%	<b>45.7857%</b>
Supervised Discretization	43.5000%	<b>45.7857%</b>

When observing Table 7, we can see that, just as it happened with  $k$ -NN and DT, the attribute normalization did not cause an improvement of the accuracy level. Additionally, it is also shown that all three versions with the normalized attributes provided the same accuracy level. Finally, for the Non-normalized dataset, the use of kernel estimation as well as supervised discretization caused a decrease in the accuracy level, suggesting that the numeric data do follow a normal distribution.

### 3.2.4. Multi-Layer Perceptron

For MLP, seventy five different architectures of a single hidden-layered MLP are evaluated, varying the number of hidden neurons, the learning rate and the number of epochs,

as showed in Table 8. For all architectures, the momentum rate was set to 0.8, applying a 2-fold cross-validation (CV) technique. The 2-fold CV technique was selected due to limited computational resources but, for the selected architecture, five 10-fold cross-validation tests were made and the mean of each test will be presented along with the mean and standard deviation of the five tests combined.

As showed in Table 8, the highest accuracy level was with the following architecture: 28 hidden neurons, learning rate of 0.005 and 10000 epochs, resulting in 90.2143% accuracy. When analyzing the other results, we can see that, for the same configuration with 10 times lower epochs, the accuracy level was 77.8571%, leading to a decrease in performance approximately 12.3572 percentage points. This observance of decrease is important because, it would strongly suggest an under-fitting event in the training model.

Therefore, once the most accurate architecture is selected, it will be investigated in a 10-fold CV technique, executed five times. The obtained results are shown in Table 9.

**Table 8. ANN architecture tests**

Neuron N°	L. Rate	Epochs		
		100	1000	10000
7	0.1	74.0000%	82.2143%	82.6429%
	0.05	70.0714%	78.0000%	82.1429%
	0.01	24.2857%	79.3571%	86.5714%
	0.005	9.2857%	72.0000%	84.6429%
	0.001	7.2857%	24.1429%	79.4286%
14	0.1	80.0714%	85.5714%	88.4286%
	0.05	73.7143%	87.2857%	89.0714%
	0.01	31.4286%	83.9286%	88.7857%
	0.005	12.3571%	75.6429%	89.5714%
	0.001	7.1429%	31.0714%	84.2857%
21	0.1	78.5000%	87.0000%	89.7143%
	0.05	73.9286%	87.0714%	88.3571%
	0.01	34.5714%	84.2143%	89.9286%
	0.005	14.7143%	76.9286%	89.9286%
	0.001	7.1429%	34.9286%	85.2143%
28	0.1	79.2857%	87.6429%	87.9286%
	0.05	74.6429%	87.6429%	89.4286%
	0.01	36.9286%	84.0000%	89.5000%
	0.005	17.4286%	77.8571%	<b>90.2143%</b>
	0.001	7.5000%	37.5714%	85.3571%
35	0.1	78.3571%	86.5000%	88.1429%
	0.05	74.3571%	88.0714%	88.2857%
	0.01	36.4286%	84.5000%	87.9286%
	0.005	18.5000%	77.7857%	88.7143%
	0.001	10.6429%	37.2857%	85.2143%

As it can be observed from Table 9, the use of a 10-fold cross-validated technique

resulted in higher accuracy level than with the 2-fold version. Additionally, the average accuracy was 92.9714%, with a standard deviation of 0.2178.

**Table 9. Five executions of the most accurate MLP architecture**

	Executions				
	1°	2°	3°	4°	5°
Accuracies	92.7857%	93.2143%	93.0000%	93.1429%	92.7143%

In summary, we can conclude that the highest accuracy level was achieved by a MLP, just as the original work [Santos 2019]. However, in this paper, we performed a wider investigation, with different aspects, less complexity and with care of not losing generalization potential of a MLP and avoiding over-fitting. In this paper, the most accurate model achieved 92.9714% accuracy with a 0.2178 standard deviation. This result is lower than the one provided by the original work, which was 99.64%. However, we believe that the model achieved in this paper is better in terms of generalization and much lower in terms of complexity, given that in [Santos 2019] only neuron quantity were evaluated for the MLP construction. Therefore, we believe that the model obtained in this paper provide a better version to be used in a computer vision model, when compared to the one originally presented in [Santos 2019].

#### 4. Conclusions and Future Works

In this paper, we presented an extensive investigation of classification algorithms to an important image-based decision making problem, cocoa beans classification. In order to do this, an empirical analysis with four classification algorithm was conducted. Finally, a feature selection technique was applied to all classification algorithms to evaluate the performance of the cocoa classification task using a more compact representation of the cocoa images. As a result of this empirical analysis, we believe that the model obtained in this paper provided a better version to be used in a computer vision model, when compared to the one originally presented in [Santos 2019].

As future work, we suggest a deeper investigation of neural network architectures, since we firmly believe that it can generate an even better model. In addition, it is possible to assess the application and study of unsupervised classification methods for this problem.

#### 5. Acknowledgments

This paper was built with the help of the Departamento de Informática e Matemática Aplicada (DIMAP) of the Universidade Federal do Rio Grande do Norte (UFRN). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

#### References

Al-Marakeby, A., Aly, A. A., and Salem, F. A. (2013). Fast quality inspection of food products using computer vision. *International Journal of Advanced Research in Computer and Communication Engineering* [1], 2.

- Brasil, Ministério da Agricultura, P. e. A. (2008). Regulamento técnico da amêndoa de cacau. *Diário Oficial da União, Brasília*, 38. Instrução Normativa.
- Catsberg, C. and Dommelen, K.-V. G. (1990). Cocoa and chocolate. In *Food Handbook*, pages 330–334. Springer.
- Fayyad, U. M. and Irani, K. B. (1992). The attribute selection problem in decision tree generation. In *AAAI*, pages 104–110.
- Hall, M. A. (2000). Correlation-based feature selection of discrete and numeric class machine learning.
- Hall, M. A. and Holmes, G. (2003). Benchmarking attribute selection techniques for discrete class data mining. *IEEE Transactions on Knowledge and Data engineering*, 15(6):1437–1447.
- Haralick, R. M., Shanmugam, K., Dinstein, I., et al. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, 3(6):610–621.
- Im, J. and Jensen, J. R. (2005). A change detection model based on neighborhood correlation image analysis and decision tree classification. *Remote Sensing of Environment*, 99(3):326–340.
- Majumdar, S., Jayas, D., Hehn, J., and Bulley, N. (1996). Classification of various grains using optical properties. *Canadian Agricultural Engineering*, 38(2):139–144.
- Patil, N. K., Malemath, V. S., and Yadahalli, R. M. (2011). Color and texture based identification and classification of food grains using different color models and haralick features. *International Journal on Computer Science and Engineering*, 3(12):3669.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1):81–106.
- Samaniego, L., Bárdossy, A., and Schulz, K. (2008). Supervised classification of remotely sensed imagery using a modified  $k$ -nn technique. *IEEE Transactions on Geoscience and Remote Sensing*, 46(7):2112–2125.
- Santos, F., Palmeira, E., and Jesus, G. (2019). An image dataset of cut-test-classified cocoa beans. *Data in brief*, 24:103916.
- Santos, F. A. (2019). Modelagem de um sistema de visão computacional para a classificação de amêndoas de cacau na prova de corte. Master's thesis, Universidade Estadual de Santa Cruz, Ilhéus.
- Santos, F. A., Palmeira, E. S., and Jesus, G. Q. (2018). Color, structural and textural features for the classification of a cocoa beans image dataset using artificial neural network.
- Savakar, D. (2012). Identification and classification of bulk fruits images using artificial neural networks. *International Journal of Engineering and Innovative Technology (IJEIT)*, 1(3):35–40.
- Warfield, S. (1996). Fast  $k$ -nn classification for multichannel image data. *Pattern Recognition Letters*, 17(7):713–721.
- Yang, C.-C., Prasher, S. O., Enright, P., Madramootoo, C., Burgess, M., Goel, P. K., and Callum, I. (2003). Application of decision tree technology for image classification using remote sensing data. *Agricultural Systems*, 76(3):1101–1117.