



MACHINE LEARNING CLASSIFIER FOR *Dalium Guineense* FRUIT USING ITS PHYSICAL PROPERTIES

¹ George, U. D., ² Akpan, M. G. and ² Onwe, D. N.

¹Department of Computer Science, University of Uyo

²Department of Agricultural and Food Engineering, University of Uyo

Corresponding author email: mfrekemfongakpan@uniuyo.edu.ng

George, U. D., Akpan, M. G. and Onwe, D. N. (2023): Machine Learning Classifier for *Dalium Guineense* Fruit Using its Physical Properties *Journal of Engineering And Engineering Technology*/17(1), 91-101

Received Date: 15-03-23

Acceptance Date: 22-06-23

Abstract

Dalium guineense (DG) is a wild fruit with a brittle epicarp that may be broken accidentally or intentionally while processing during any of the unit operations thereby creating a binomial mixture. Having a binomial mixture of similar items that need to be separated for processing or storage purposes presents a common challenge. This research aims at selecting an appropriate machine learning classifier for the classification of DG fruits. Fifteen measured physical characteristics of randomly selected 200 DG fruits were obtained. Fifty percent of the fruits were deshelled, while the remaining 50% were whole fruits. Different machine learning classifiers were chosen from Decision Tree (DT), Support Vector Machines (SVM), and K-Nearest Neighbor (KNN) classification models, using the classification learner toolkit of MATLAB. The results revealed that Coarse Gaussian SVM and Cosine KNN presented an outstanding classification accuracy of 98.5% compared to other classifiers under investigation. The two classifiers also attained precision, sensitivity, specificity, and F-scores of 99.0%, 98.0%, 99.0%, and 98.5% respectively. The method deployed in this study demonstrates superiority to those reported in some literature. This research recommends the adoption of either the coarse Gaussian SVM or the cosine KNN as the most appropriate classifier for the DG fruits classification.

Keywords: machine learning, classifier, classification, deshelled fruit, whole fruit, *Dalium Guineense*, classification learner.

Introduction

Dalium guineense is a wild fruit found in the tropical parts of Africa: Nigeria, Ghana, Cameroun, Benin Republic, Burkina Faso, Central African Republic, Chad, Cote d'Ivoire, Equatorial Guinea and Senegal. It belongs to the Fabaceae family. It is very nutritious and especially rich in Vitamin C and beta carotene as evidenced in both its taste and colour (Lasekan and See, 2015). There has been an increase in interest in wild edible fruits due to so many reasons, one of which is their antioxidants. *Dalium guineense* has also received some attention recently. Lasekan and See, (2015) identified and quantified its volatile aroma. Samarkradhamrongi and Jannu, (2021) used this wild fruit as an ingredient to produce chewy candy. Abiodun *et al.* (2017) used an ingredient in the fruit to sweeten a local beverage called kunu-zakki. Okudu *et al.* (2017) exploited its functional properties in the production of jam. Afolabi *et al.* (2018) exploited its potential in the treatment of diabetes. Asoiro *et al.* (2017) studied the chemical and mechanical properties of this wild fruits. *Dalium guineense* has a brittle epicarp (Lasekan and See, 2015) that could be broken either accidentally or intentionally during/post-harvest, and sorting

manually could be time and energy-consuming with attendant low efficiency.

Advancements in science and technology have wielded much influence in the field of agriculture so much that production is enhanced. Information Technology has also contributed immensely to the current trends towards improved agricultural practices to guarantee food sufficiency. Recently, exploits in the field of machine learning have been reported to make tremendous impacts on food production. In Olamide *et al.* (2020), it was reported that apart from Agritech (a coinage from Agriculture and Technology) that allows all forms of technology, there is also Precision Agriculture (PA) which is a farm management approach that utilizes Information Technology (IT) to ensure that optimum need of plants and animals are met efficiently and effectively while guaranteeing maximum productivity and optimum health of farm produce. Several machine learning models have been employed in the classification and prediction of agricultural produce. Among such models include Decision Tree (DT), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN).

The type of features under measurement determines the classification models that could be utilized. Kaur *et al.* (2021) proposed an optimization of sunflower flower recognition through machine learning. The research used recognition and classification of sunflower using KNN classifier and the result recorded 88.52% classification accuracy. Huang *et al.* (2022) proposed a deep learning classification of soybean seeds. Two-step procedure of segmentation-classification was adopted, where image segmentation was done using the deep learning method, the Mask R-CNN, and classification performed using a novel network technique, SNet. The proposed model achieved an accuracy of 96.2%. Prediction of various crops in the agricultural field using decision tree and Naive Bayes algorithm in machine learning was carried out by Bhavani *et al.* (2021). The research used CART and Naive Bayes decision tree methods to predict the yields of various crops such as rice, wheat, jowar, mustard and sugar cane based on the type of soil. The results helped the researchers to make decision on the crop that is economically good for each location. Kalichkin *et al.* (2021) applied the decision tree method for predicting the yield of spring wheat. CI and CART methods of decision tree algorithm were used to predict spring wheat yield using qualitative features and a combination of qualitative and quantitative features. The coefficient of determination with the CI method was 0.783 (training sample) and 0.744 (test sample) when using CART method, coefficient of determination was 0.895 (training sample), and 0.811 (test sample).

To demonstrate the efficacy of Information Technology (IT) in precision agriculture, decision tree algorithms were evaluated in precision agriculture by Olamide *et al.* (2020). The research used three decision tree classifiers namely C5.0, ID3 and CART. Analysis of the classifiers were done by R 3.6.2. Results showed that CART marginally outperformed the C5.0 and ID3 classifiers in terms of the accuracy of 99.57%, precision of 97.83%, recall of 98.84%, and F-score of 98.33%. Kramar *et al.* (2018) demonstrated the application of support vector machines for prediction and classification. The outcome of the experiments indicated that SVM is a good technique for classification and prediction purposes. Jye *et al.* (2018) performed automated plant identification using an artificial neural network (ANN) and support vector machine (SVM). Result indicated the same accuracy of 83.3% for both ANN and SVM. However, comparing the performance of the two models using AUC measure shows a slightly better performance of ANN over SVM. A comparative study of four supervised machine-learning techniques for classification was reported by Mohamed (2017). The experiments were conducted with four common classifiers namely, support vector machine, artificial neural network, decision tree, and K-nearest neighbor. Results showed SVM performing more than others with the accuracy of 76.3% followed by decision tree with 73.9%. A K-nearest neighbor algorithm was used for the prediction of agricultural

crops by (Karthikeya *et al.*, 2020). The experiment showed that KNN predicted cocoa and coconut for the selected location with an accuracy of 63.63%. Analysis of paddy crop production was carried out using SVM and KNN by (Bhambri *et al.*, 2020). SVM and KNN were used to improve the accuracy of paddy production prediction and their performance compared. The result indicated the performance accuracy of about 81% and 89% for SVM and KNN respectively. Syahminan *et al.* (2019) used KNN to identify agricultural land and the result assisted in making decision on the type of food crops to cultivate based on the condition of the land. A data mining approach was proposed for dry bean seeds classification in Macuacua *et al.* (2023). The research developed a computer vision system to automatically classify several seed varieties based on machine learning (ML) models combined with data mining techniques. Three ML techniques of Random Forest (RF), SVM, and KNN were used, including Principal Component Analysis (PCA), hyperparameter tuning and dataset balancing for Data Mining (DM). The results showed that incorporation of DM techniques helped to improve the classification by about 2.6% accuracy, whereas KNN classifier showed a better performance of about 95% and 96% accuracy and recall respectively, while SVM performed with the accuracy of 94.4%.

Support Vector Machine (SVM) is known for its robustness, good generalization ability, and unique global optimum solutions with very simple principles. It is used for binary classification problems, and can also be used for regression analysis. When there are only two classes of objects, it is necessary to determine whether the object belongs to a particular class according to a given set of parameters. Optimality can be achieved with SVM by minimizing the upper bounds for the probability of a generalization error (Kramar *et al.*, 2018). It is one of the best-known methods in pattern classification and image classification (Thai *et al.*, 2012). K-Nearest neighbour is a lazy learner technique which depends on learning by analogy. The classifier is used extensively for supervised classification purposes. It waits till the last minute to build some model on a specified tuple as compared to other classifiers (song *et al.*, 2007). The K-nearest neighbour (K-NN) method is a data mining technique which ranks among the top five techniques for data mining (Karthikeya *et al.*, 2020) that can be used for classification purposes. The Decision Tree (DT) is a member of the supervised learning algorithm family. It aims at constructing a training model that can be used to predict the class through the process of learning the decision rules inferred from the training data. The DT algorithm can be used to solve regression and classification problems (Jijo and Abdulazeez, 2021).

Physical properties of agricultural products are part of their engineering properties that include the length, width, thickness, colour, geometric mean diameter, square mean diameter, equivalent mean diameter, sphericity etc. they are important in the effective design of equipment necessary for the handling, processing

and even the packaging of these product. The deshelled and whole fruit of DG are similar in these physical characteristics except in colour and since the epicarp is brittle which can break either accidentally or intentionally, therefore the binomial mixture is inevitable.

Therefore, this research aims to select a machine learning (ML) classifier from support vector machine, K Nearest Neighbor and Decision Tree for the efficient classification of Dalium Guineense using its physical properties such as length, width, thickness, geometric mean diameter, arithmetic mean diameter, square mean diameter, equivalent mean diameter, surface area, sphericity and aspect ratio.

Materials and Method

Dataset Description

A large number of high-quality DG fruits were obtained from a farmer in Uyo, Akwa Ibom State, Nigeria (5.0377 °N, 7.9128 °N) and were processed within 1 (one) hour of arrival at the Food Engineering Laboratory of University of Uyo by air-cleaning them using compressed air to remove any dirt and debris that might have been stocked to the fruits. 200 samples of the fruits were randomly selected for the experiment. 50 % were manually deshelled to expose the mesocarp (deshelled) while the remaining 50 % were left whole (whole fruits). Figures 1a and 1b show the whole and deshelled Dalium guineense fruits respectively.



Figure 1a: Dalium Guineense Fruits



Figure 1b: Deshelled Dalium Guineense Fruits

Length (L), width (W) and Thickness (T) as the axial dimensions of the whole and deshelled fruits were measured using an electronic vernier caliper of 0.001 mm. LW, WT and TL were the product of length and width, width and thickness; and thickness and length respectively. Geometric mean diameter (D_g), Arithmetic mean diameter (D_a), square mean diameter (SMD) and equivalent mean diameter (D_e) were calculated using Equations (1-4) according to (Kilican and Guner, 2008), (Asoegwu *et al.*, 2006) and (Onwe *et al.*, 2020).

$$D_g = (LWT)^{1/3} \tag{1}$$

$$D_a = \frac{L+W+T}{3} \tag{2}$$

$$SMD = \left(\frac{LW+WT+TL}{3} \right)^{1/2} \tag{3}$$

$$D_e = \frac{(D_g + D_a + SMD)}{3} \tag{4}$$

Surface Area (A_s) was evaluated based on Equation (5) according to (Onwe *et al.*, 2020)

$$A_s = \pi(D_g)^2 \tag{5}$$

Sphericity (S) was determined using Equation (6) as reported by (Asoegwu *et al.*, 2006)

$$S = \frac{D_g}{L} \tag{6}$$

Aspect Ratio (R_a) was evaluated using Equation (7) as reported by (Obi and Oforha, 2015)

$$R_a = \frac{W}{L} \tag{7}$$

These physical properties are described as used in this study as shown on Table 1. For this study, the categorical data was encoded as 1 for deshelled fruit and 0 for whole fruit. A sample of the dataset are also shown on Table 2.

Table 1: DG Dataset Features

Data Attribute	Data Type	Description
Length	Real	The physical length of the fruit
Width	Real	The physical width of the fruit
Thickness	Real	The physical thickness of the fruit
LW	Real	Product of length and width of the fruit
WT	Real	Product of width and thickness of the fruit
TL	Real	Product of thickness and length of the fruit
GeometricMeanDiameter	Real	The geometric mean diameter of the fruit
ArithmeticMeanDiameter	Real	The arithmetic mean diameter of the fruit
SquareMeanDiameter	Real	The square mean diameter of the fruit
EquivalentMeanDiameter	Real	The equivalent mean diameter of the fruit
SurfaceArea	Real	The surface area of the fruit
SP	Real	Specificity measurement of the fruit
AspectRatio	Real	The aspect ratio of the fruit
ActualClass	Categorical	Class label of the fruit (deshelled fruit, whole fruit)

Table 2: Sample Dataset of DG

Length	Width	Thickn ess	LW	WT	TL	B	Geometric Mean Diameter, Dg [m]	Arithmetic Mean Diameter, Da [m]	Specific Mean Diameter, Ds [m]	Equivale nt Mean Diamete r, De [m]	Surface Area
0.0116	0.0151	0.0119	0.000176	0.000180	0.000138	0.0132	0.0128	0.0129	0.0128	0.0128	0.000579
0.0178	0.0173	0.0097	0.000308	0.000168	0.000173	0.0175	0.0144	0.0149	0.0147	0.0147	0.000569
0.0183	0.0108	0.0121	0.000198	0.000131	0.000221	0.0141	0.0134	0.0137	0.0135	0.0135	0.000478
0.0221	0.0177	0.0052	0.000391	0.000092	0.000115	0.0198	0.0127	0.015	0.0141	0.0139	0.000425
0.0205	0.0161	0.0115	0.000330	0.000185	0.000236	0.0182	0.0156	0.016	0.0158	0.0158	0.000656
0.0147	0.0112	0.0078	0.000165	0.000087	0.000115	0.0128	0.0109	0.0112	0.0111	0.0111	0.000316
0.0222	0.0122	0.0057	0.000271	0.000070	0.000127	0.0165	0.0116	0.0134	0.0125	0.0125	0.000358
0.0158	0.0165	0.0069	0.000261	0.000114	0.000109	0.0161	0.0122	0.0131	0.0127	0.0126	0.000400
0.0208	0.0155	0.0084	0.000322	0.000130	0.000175	0.018	0.0139	0.0149	0.0145	0.0144	0.000514
0.0114	0.0115	0.0064	0.000131	0.000074	0.000073	0.0114	0.0094	0.0098	0.0096	0.0096	0.000246
0.0173	0.0149	0.0126	0.000258	0.000188	0.000218	0.0161	0.0148	0.0149	0.0149	0.0149	0.000616
0.0174	0.0112	0.0056	0.000195	0.000063	0.000097	0.014	0.0103	0.0114	0.0109	0.0109	0.000280
0.0191	0.0131	0.0079	0.000250	0.000103	0.000151	0.0158	0.0126	0.0134	0.013	0.013	0.000416
0.0126	0.0147	0.0056	0.000185	0.000082	0.000071	0.0136	0.0101	0.011	0.0106	0.0106	0.000281
0.0205	0.0157	0.0121	0.000322	0.000190	0.000248	0.0179	0.0157	0.0161	0.0159	0.0159	0.000669

Machine Learning Techniques Used

Three Machine Learning (ML) techniques, namely Decision Tree (DT), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) were used for the study based on the peculiarity of the data, where all the

data samples were numeric types. Table 3 shows the characteristic features of the selected machine learning tool. Matlab software was used for the classification (Mathlab.com).

Table 3: Classifiers Selection and Data Types

Classifier	All Predictors are Numeric	All Predictors are Categorical	Some are Categorical, Some are Numeric
Decision Tree	Yes	Yes	Yes
SVM	Yes	Yes	Yes
Nearest Neighbor	Euclidean distance only	Hamming distance only	No

The objective of designing the SVM was to transform binary class data into feature space or higher dimensional space and find an optimum separating hyperplane among the set of hyperplanes which provides maximum margin between two different classes. It is a binary classifier with the ability to classify both linear and non-linear separable data. The decision boundary is located by a very small subset of training samples known as support vectors. Linear SVM, Cubic SVM, Quadratic SVM, Fine Gaussian, Medium Gaussian SVM, and Coarse Gaussian SVM are the classifiers employed in the SVM classification model (Sing et al., 2021; Battineni et al., 2019). The class, y , is defined in Equation (8):

$$y^{(i)} = \begin{cases} -1, & \text{if } w^T x^{(i)} + b < 0 \\ 1, & \text{if } w^T x^{(i)} + b \geq 0 \end{cases} \quad (8)$$

w is the weight coefficient of both the independent and dependent variables of the line, and b is the shift factor or the intercept on the x -axis.

Two requirements are fulfilled by the objective of SVM:

a. There should be a maximum possible distance between the two decision boundaries. The equation of hyperplane is defined as

$$w^T x + b = 0$$

and the distance between hyperplanes H_1 and H_2 close to the two different classes is $\frac{2}{\|w\|}$. If the distance is to be maximized, then $\frac{2}{\|w\|}$ should be equivalent to minimized.

b. Support Vector Machine should correctly classify all which is the i th sample in the dataset. Without prior

knowledge, the KNN classifier usually applies Euclidean distances as the distance metric. It can however still yield competitive results compared to the most sophisticated machine learning methods (Song et al., 2007). The Euclidean distance is given as shown on Equation (9) (Syhminan et al., 2019):

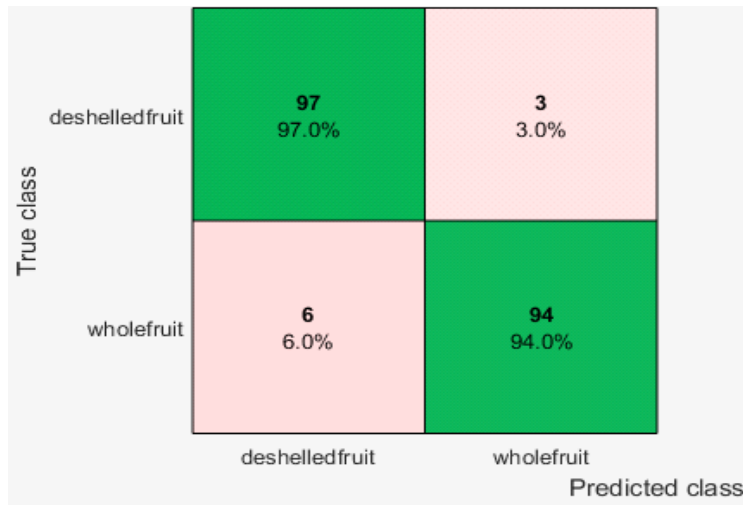
$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_1)^2} \quad (9)$$

The algorithm used was elaborated in a study by Karthikeya et al. (2020).

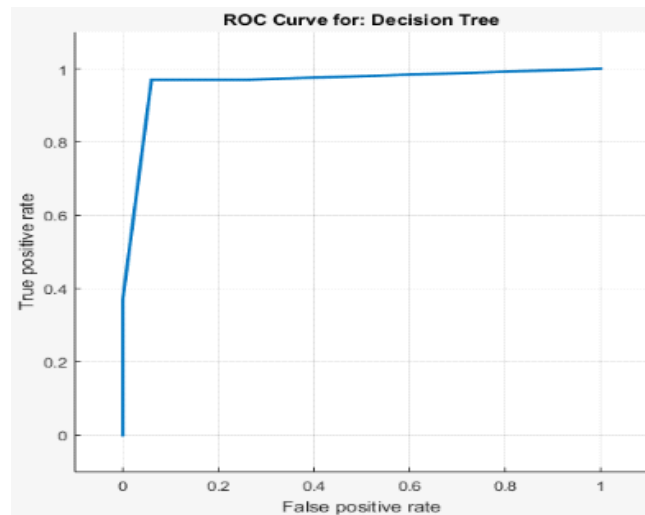
A decision tree model used for this work has nodes and branches as its main components with splitting, stopping, and pruning as the most important steps in building a model (Song and Lu, 2015).

Results and Discussion

The results obtained in the experiments using several classifiers for the classification of DG fruits are presented in this section. The classification learner toolbox of MATLAB was used in the experiment. A total of 15 machine learning classifiers shown in Table 4 were deployed in experimenting using the preprocessed dataset. A ten-fold cross-validation method was used for the validation of different classifiers. Performance metrics such as accuracy, precision, sensitivity, specificity and F-score were used to determine the overall functionality of each classifier. Tenfold cross-validation was used for the validation of different classifiers. Figure 2 shows the confusion matrix, Receiver Operating Characteristic (ROC) curve, for DT in 20- (a)–(c) respectively.

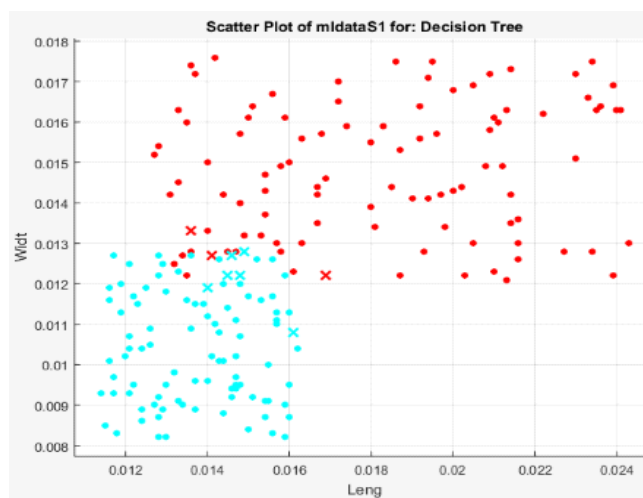


(2a)



(2b)

Figure 2 (a-b): The Confusion Matrix and Receiver Operating Characteristic (ROC) curve for DT



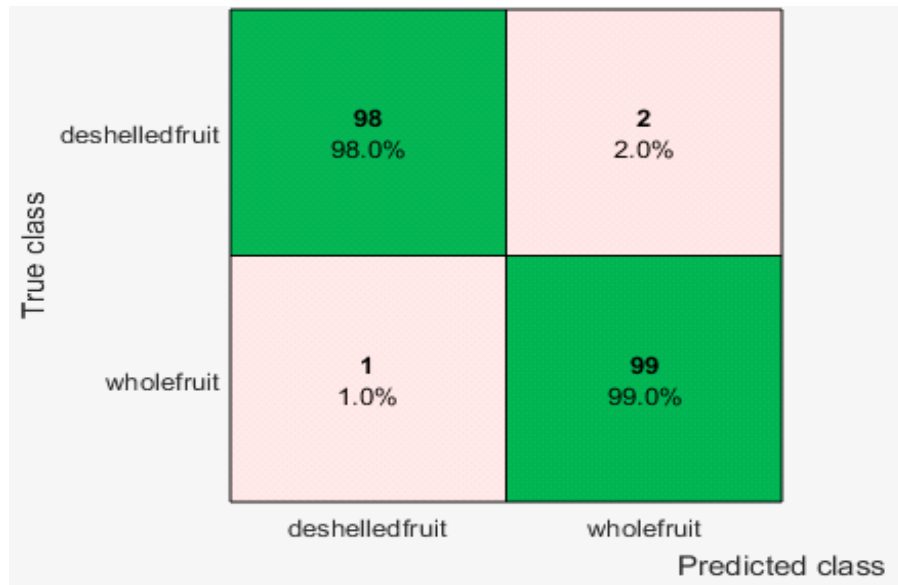
(2c)

Figure 2: (a). Confusion Matrix, (b). ROC and (c). Scatter Plot for DT

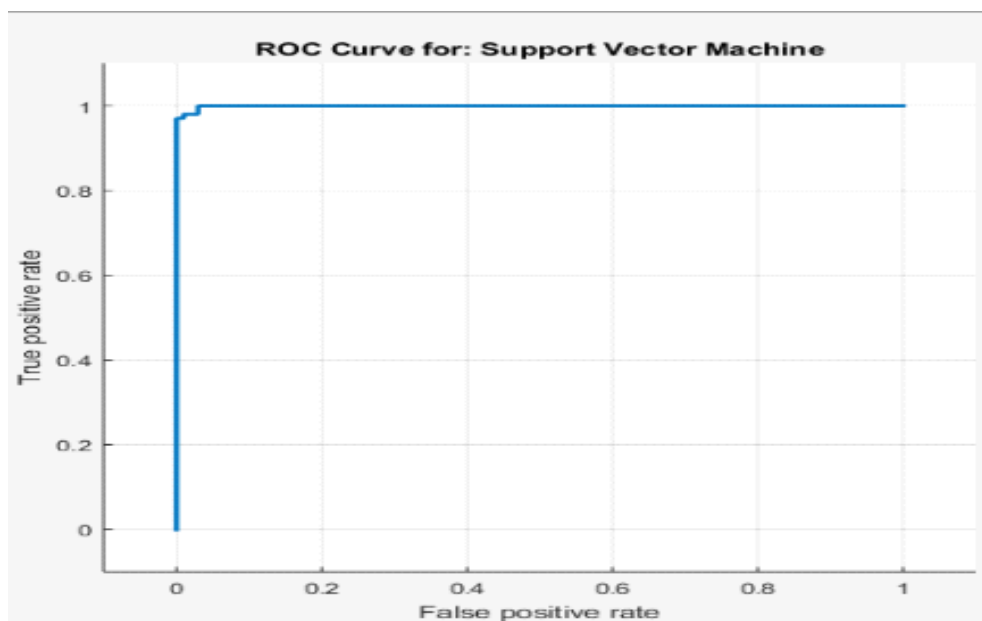
The confusion matrix shows that DT method misclassified 3 deshelled fruits and 6 whole fruits. The ROC curve does not indicate a very good classifier as the area under the curve is not large enough, and the

scatter plot of width against length of the fruits clearly indicates an interwoven data points at the boundaries between the two measured features.

Figure 3 shows the confusion matrix, Receiver Operating Characteristic (ROC) curve, and scatter plot for SVM in 2(a)–2© respectively.



(3a)



(3b)

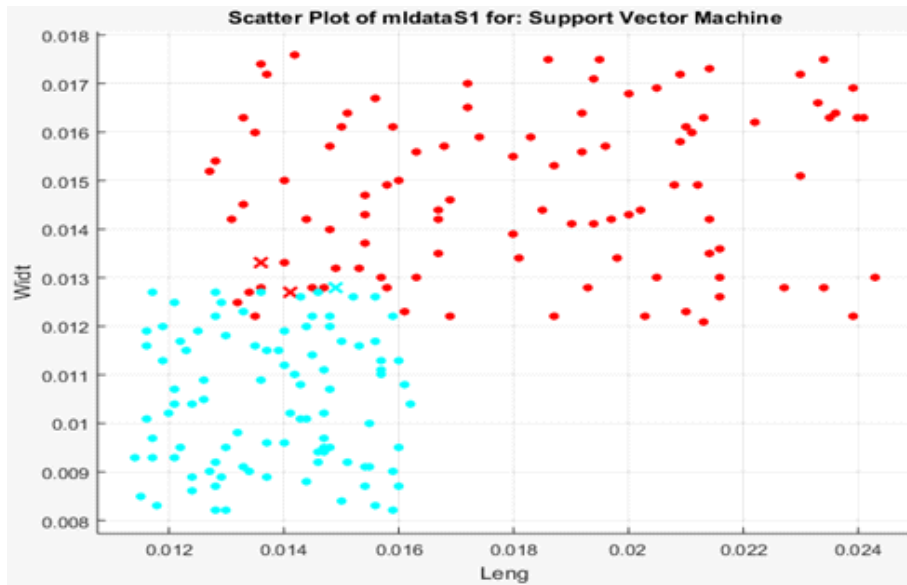
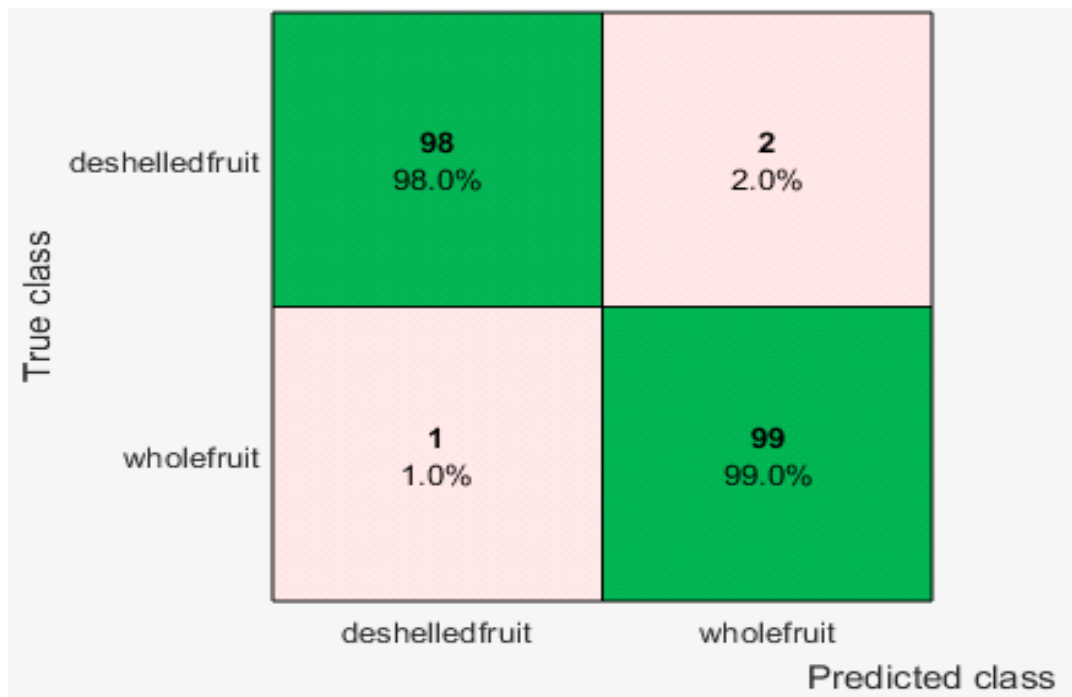


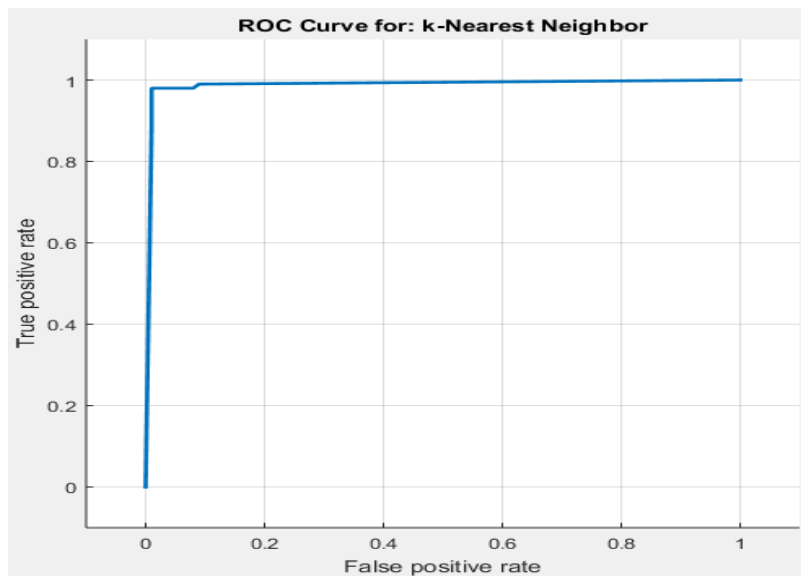
Figure 3: (a). Confusion Matrix, (b). ROC and (c). Scatter Plot for Coarse Gaussian SVM

The confusion matrix shows SVM method misclassified only two deshelled fruits and one whole fruits, indicating a better classification than that of the DT. The ROC curve shows a better classifier as the area under the curve is significantly large enough, and the scatter plot of width against length of the fruits clearly indicates a better separation of data points at the boundaries between the two measured features.

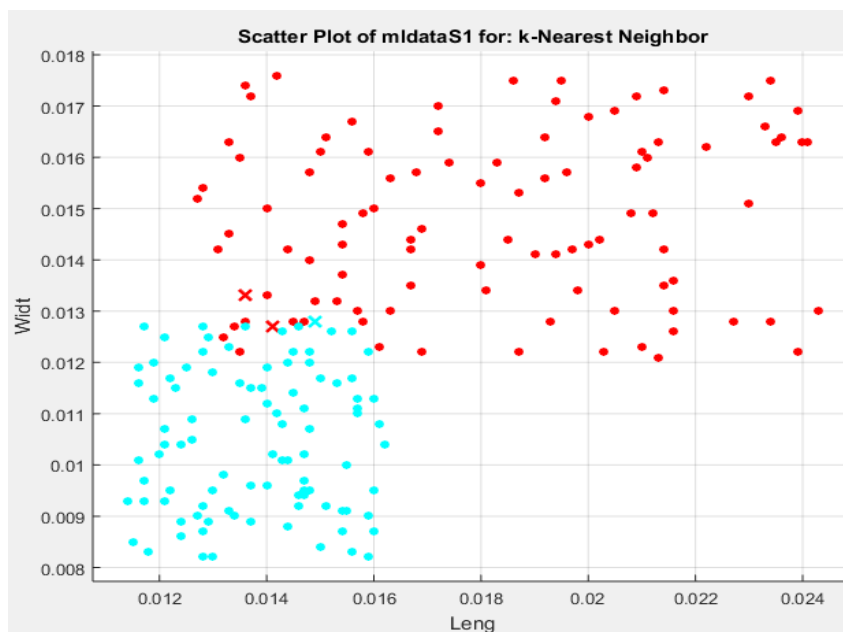
Figure 4 shows the confusion matrix, Receiver Operating Characteristic (ROC) curve, and scatter plot for KNN in 4(a)–3© respectively.



(4a)



(4b)



(4c)

Figure 4: (a). Confusion Matrix, (b). ROC and (c). Scatter Plot for Cosine KNN

The confusion matrix shows SVM method misclassified only two deshelled fruits and one whole fruits, also indicating a better classification than that of the DT, and similar to that of SVM above. Evaluation of the classifiers' performance using five metrics is presented in Table 4.

Table 4: Classifiers Performance Metrics

Model Type	Classifiers	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F-Score (%)
Decision Tree (DT)	Simple DT	95.5	94.2	97.0	94.0	95.6
	Medium DT					
	Complex DT					
Support Vector Machine (SVM)	Linear SVM	97.5	99.0	96.0	99.0	97.5
	Quadratic SVM	98.0	99.0	97.0	99.0	98.0
	Cubic SVM					
	Fine SVM					
	Medium SVM	98.5	99.0	98.0	99.0	98.5
	Coarse Gaussian SVM					
K-Nearest Neighbor (KNN)	Medium KNN	97.0	98.0	96.0	98.0	97.0
	Coarse KNN	96.0	100.0	92.0	100.0	95.8
	Cosine KNN	98.5	99.0	98.0	99.0	98.5
	Cubic KNN	97.5	99.0	96.0	99.0	97.5
	Weighted KNN					

All the DT classifiers (simple, medium, and complex) generated results above 95% accuracy, above 94% precision, 97% sensitivity, 94% specificity, and above 95% F-score. This means DT classifiers misclassified about 4 samples to be whole fruit whereas they were deshelled fruit. The work in Olamide et al. (2020) which carried out evaluation of decision tree algorithms in precision agriculture recorded higher accuracy of 99.57%. Obviously, the lower performance of decision trees used in this study is an indication that the classifiers are not very appropriate for the type of data in the research. It would be a good idea to experiment with C5.0, ID3 and CART decision tree classifiers for classification of *Dalium guineense*, to observe their performance. Quadratic, cubic, fine, and medium SVM classifiers generated the similar accuracy, precision, sensitivity, specificity, and F-score of 98.0%, 99.0%, 97.0%, 99.0%, and 98.0% respectively. Linear SVM on the other hand, produced an accuracy and F-score above 97%, while precision, sensitivity, and specificity were 99%, 96%, and 99% respectively. Coarse Gaussian SVM generated accuracy and F-score above 98%, as well as 99%, 98% and 99% for precision, sensitivity, and specificity respectively. The accuracy obtained in the literature (Jye et al., 2018 was 83.3%, compared to what is obtained in the current study which is above 98%. This demonstrates the effectiveness of SVM in the classification process. For the K-nearest neighbour, the medium, coarse, and cosine KNN classifiers recorded 97.0%, 96.0%, and 98.5% accurate classifications. However, two KNN classifiers, cubic, and weighted, generated similar results above 97% accuracy. The results this research, with over 95% accuracy performs better than the accuracy of 63.63% obtained in Karthikeya et al. (2020). (2020) where KNN was used in the prediction of agricultural crops, 88.52% obtained in Kuar et al. (2021) where it was used as one component of machine learning in classification and recognition of sunflower flower. The current system also recorded a higher accuracy measure than the 94.4% accuracy in Macuacua et al. (2023) where KNN was

used in conjunction with data mining techniques to classify dry bean seeds.

Two classifiers, coarse Gaussian SVM and cosine KNN were observed to have performed at the highest accuracy above 98% better than the other classifiers under investigation, meaning that generally they correctly classified more than 98 cases out of 100 of the two classes of deshelled and whole fruit of the fruits. Precision, sensitivity, specificity, and F-score of the two classifiers were 99.0%, 98.0%, 99.0%, and 98.5% respectively. A precision of 99% means that the two classifiers were able to classify the desired group correctly. A sensitivity value of 98.0% implies that 98 out of 100 cases were correctly classified as deshelled while 2 were misclassified. 99.0% specificity value indicates that 99 out of 100 cases of whole fruits were correctly classified while only 1 missed the classification. An F-score of 98.5% indicates more than 98 % false positive and false negative values. Therefore, either the Coarse Gaussian SVM or the cosine KNN could be adopted as the best classifier for DG fruits.

Conclusion

Dalium guineense fruits were randomly selected and 50% were deshelled and 50% were whole fruits. Fifteen supervised machine learning classifiers were used to train the dataset using the classification learner toolkit of MATLAB software. Two classifiers namely, Coarse Gaussian SVM and cosine KNN performed with higher accuracy of over 98% better than several other classifiers. Therefore, for high-accuracy classification of the DG fruits, either the Coarse Gaussian support vector machine or the cosine K-nearest neighbour classifier could be adopted. As a suggestion for further research, a deep learning platform for classification is recommended, where the performance evaluation can be compared with that of other implemented models.

References

- Abiodun, O. A., Dauda, A. O., Adebisi, T. T., and Alonge, C. D. (2017). Physico-chemical, microbial and sensory properties of kunu zaki beverage sweetened with black velvet tamarind (*Dialium guineense*). *Croatian Journal of Food Science and Technology*, 5(5), 1422–1424.
- Afolabi, O. B., Oloyede, O. I., Ojo, A. A., Onasanya, A. A., Agunbiade, S. O., Ajiboye, B. O., Jonathan, J., and Peters, O. A. (2018). In vitro antioxidant potential and inhibitory effect of hydroethanolic extract from African black velvet tamarind (*Dialium indium*) pulp on type 2 diabetes linked enzymes. *Potravinarstvo*, 12(1).
- Asoegwu, S. N., Ohanyere, S. O., Kanu, O. P., and Iwueke, C. N. (2006). Physical properties of African oil bean seed (*Pentaclethra macrophylla*). *Agricultural Engineering International: CIGR Journal E.Journal. Manuscript FP05 006. Vol. VIII*.
- Asoiro, F. U., Ezeoha, S. L., Ezenne, G. I., and Ugwu, C. B. (2017). Chemical and mechanical properties of velvet tamarind fruit (*Dialium guineense*). *Nigerian Journal of Technology*, 36 (1), 252–260.
- Battineni, G., Chintalapudi, N., and Amenta, F. (2019). Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM). *Informatics in Medicine Unlocked*, 16, 100200.
- Bhambri, P., Dhanoa, I. S., Sinha, V. K., and Kaur, J. (2020). Paddy Crop Production Analysis Based on SVM and KNN Classifier. *International Journal of Recent Technology and Engineering*, 8(5), 2790–2793.
- Bhavani, B. G., Kumar, G. L. N. V. S., Moram Lakshim, Rekha, M. L., K N V P S, and Ramesh, K. N. V. P. S. B. (2021). Prediction of Various Crops in Agricultural Field Using Decision Tree and Naviebayes Algorithm in Machine Learning. *International Journal of Engineering Research & Technology*, 9(5), 79–83.
- Jijo, B. T., and Abdulazeez, A. M. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 02(01), 20–28.
- Jye, K. S., Manickam, S., Malek, S., Mosleh, M., and Dhillon, S. K. (2018). Automated plant identification using artificial neural network and support vector machine. *Frontiers in Life Science*, 10(1), 98–107.
- Kalichkin, V. K., Alsova, O. K., and Maksimovich, K. Y. (2021). Application of the decision tree method for predicting the yield of spring wheat. *IOP. Conference Series: Earth and Environmental Science*. (Vol. 839. No 3) IOP Publishing.
- Karthikeya, H. K., Sudarshan, K., and Shetty, D. S. (2020). Prediction of Agricultural Crops using KNN Algorithm. *International Journal of Innovative Science and Research Technology*, 5(5), 1422–1424.
- Kılıçkan, A., and Güner, M. (2008). Physical properties and mechanical behavior of olive fruits (*Olea europaea* L.) under compression loading. *Journal of Food Engineering*, 87(2), 222–228.
- Kramar, V. A., Alchakov, V. V., Dushko, V. R., and Kramar, T. V. (2018). Application of support vector machine for prediction and classification. *Journal of Physics Conference series* (Vol. 1015, No. 3) IOP Publishing
- Lasekan, O., and See, N. S. (2015). Key volatile aroma compounds of three black velvet tamarind (*Dialium*) fruit species. *Food Chemistry*, 168, 561–565.
- Macuacua, J. C., Centeno, J. A.S. and Amisse, C. (2023). Data mining approach for dry bean seeds Classification. *Smart Agricultural Tech*. 5, 100240.
- Mohamed, A. E. (2017). Comparative Study of Four Supervised Machine Learning Techniques for Classification. *International Journal of Applied Science and Technology*, 7(2), 5–18.
- Obi, O. F., and Offorha, L. C. (2015). Moisture-dependent physical properties of melon (*Citrullus colocynthis lanatus*) seed and kernel relevant in bulk handling. *Cogent Food & Agriculture*, 1(1), 1020743.
- Okudu, H. O., Umoh, E. J., Ojinnaka, M. C., and Chianakwalam, O. F. (2017). nutritional functional and sensory attributes of jam from velvet tamarind pulp. *African Journal of Food Science*, 11(2), 44–49.
- Olamide, K., Oludele, A., Monday, E., K., and Chigozirim, A. (2020). Evaluation of Decision Tree Algorithms in Precision Agriculture. *International Journal of Computing and Technology*, 7(3), 25–33.
- Onwe, D. N., Umani K.C., Olusunde, W. A and Ossom, I. S. (2020). Comparative analysis of moisture-dependent physical and mechanical properties of African star apple (*Chrysophyllum albidum*) seeds relevant in engineering design. *Scientific African*, 8, e00303.
- Samakradhamrongthai, R.S., and Jannu, T. (2021). Effect of Stevia, xylitol, and corn syrup in the development of velvet tamarind (*Dalium indium* L.) chewy candy. *Food Chemistry*, 352, 129353
- Singh, L., Janghel, R. R., and Sahu, S. P. (2021). Classification of Hepatic Disease Using

- Machine Learning Algorithms. In *Advances in Biomedical Engineering and Technology. Select Proceedings of ICBEST 2018* (pp. 161-173) Springer Singapore.
- Song, Y., Huang, J., Zhou, D., Zha, H., & Giles, C. L. (2007). IKNN: Informative k-nearest neighbor pattern classification. In *Knowledge Discovery in Databases: PKDD 2007: 11th European Conference on Principles and Practice of Knowledge Discovery in Databases, Warsaw, Poland, September 17-21, 2007. Proceedings 11* (pp. 248-264). Springer Berlin Heidelberg.
- Song, Y., and Lu, Y. (2015). Decision tree methods: Applications for classification and prediction. *Shanghai Archives of Psychiatry*, 27(2), 130–135.
- Syahminan, S., Maknunah, J., Dijaya, R., and Hindarto, H. (2019). KNN (K-Nearby Neighbor) for identifying agricultural land. In *Journal of Physics Conference Series* (Vol. 1402, No 6, p 066059). IOP publishing
- Thai, L. H., Hai, T. S., and Thuy, N. T. (2012). Image Classification using Support Vector Machine and Artificial Neural Network. *International Journal of Information Technology and Computer Science*, 5, 32–38.<https://www.mathworks.com/help/stats/choose-a-classifier.html> Retrieved 10 February, 2023