Proceedings of the Third International Conference on Intelligent Sustainable Systems [ICISS 2020]
IEEE Xplore Part Number: CFP20M19-ART; ISBN: 978-1-7281-7089-3

Classification of Durian Fruits based on Ripening with Machine Learning Techniques

Muthulakshmi . A
Research Scholar,
Hindustan Institute of Technology and Science,Chennai,
Tamilnadu, India
muthulakshmi.samy0911@gmail.com

Dr.P N.Renjith
Associate Professor,
Hindustan Institute of Technology and Science,Chennai,
Tamilnadu, India
pnrenjith@hindustanuniv.ac.in

2020 3rd International Conference on Intelligent Sustainable Systems (ICISS) | 978-1-7281-7089-3/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICISS49785.2020.9316006

Abstract—Eating fruits is good for health and an excellent source of essential vitamins and minerals. But fruits are unfit for intake if they are not fully ripened. Ripening of fruits is an important phase in Pomology based on which the fruits are categorized. In this context, various machine learning techniques were proposed to detect the classification of fruits. This paper provides a special architecture to identify the features of different classes of Durian fruits. Durian, popularly given the name, "king of fruits", is one of the rarest and sweetest fruits grown in South Asia. The model of image processing helps to divide the classification into two parts of the feature extraction part and classification part of the fruit used. In this article, the system is evaluated using the datasets of the Durian Fruit. The datasets is divided into two parts as training and testing. By applying the Edge Detection and color extraction the features of the durian are properly measured. Finally, the performance is measured using Non-Destructive Machine Learning techniques such as SVM, GNB, Random Forest. The results obtained provide the best accuracy of 89.3 % using the SVM technique and 84.3% using Random Forest.

***Keywords*—Fruit Classification, Machine Learning, SVM, Decision Tree, KNN, Random Forest.**

I INTRODUCTION

Durian is one of the sweetest fruits grown in South-East Asia. It is popularly known as, "king of fruits", due to its rich sweet creamy delicious taste with a strong distinctive odor. he harvest area for durian in India has been T growing, as durian is one of the most profitable fruits for farmers and exporters. The farmer producers only harvest The .hat have fallen from the treesripe durians t %100 producers harvest durians before they are fully ripened to increase their shelf life. However, harvesting the durians early compromises their quality and the Brix level does not reach its maximum potential. As a result, the Musang King variety is .mier and sweeter than the Thai Monthong durian creaIn Malaysia, farmers don't pluck the durian from the tree, rather wait for them to fall off naturally so the fruits are ripe. Hence, the farmers experience huge losses while exporting out.

Determination of ripening of fruits manually is not an easy task which is high time consumes, inaccurate in prediction, and which also needs skills. Thus this work explains how different types of algorithms for the early detection of ripening of durian fruits are applied for the determination of the classification of the durian fruit. This proposal is done to help farmers with high production and minimal loss. ML techniques give an analysis of the high-quality fruit images of data. It generally gives the fastest output with better accuracy.

For the necessary determination of effective identification of the fruits, the process is divided into two phases of Pre-Processing and Classification once the fruit image data set is determined. In the first phase, the images are preprocessed based on its features of size, shape, texture, and color. Once the above features are identified the fruit can be identified separately from the other parts of the plants such as leaves, flowers, branches, bushes, etc. The second phase is the classification based on fruit ripeness is determined. If the Destructive technique is applied to the fruits then the used fruits become waste and they become unfit in the future. Thus the study goes for Non-Destructive model helps in predicting the level of classification based on its performance level, accuracy, and processing speed using machine learning techniques. Generally, the fruits are categorized as Un-ripe, Semi-ripe, and Fully riped.

Usually, the fruit images undergo certain pre-processing steps before the training and testing process in techniques. First, in this phase, a filter is applied to remove the noise and to smoothen the captured image. The three basic pre-processing steps applied in this works are Filtering the image, Color Enhancement, and using Contrast enhancement. Filtering is a technique used for enhancement of the image. Certain operations like smoothening, sharpening, and edge enhancements are used in this step. On the other hand, the contrast enhancement helps in the manipulation of colors to increase the contrast. Linear and Non-Linear transformation of the images can be done in this process. In the last step of pre-processing, the images are processed based on color where the RGB images are converted to HSV and L*a*b which converts binary to grayscale images for effective identification.

The main objective of the durian image classification is to detect, identify, and classify the features of an image. After the non-destructive identification of fruits, the performance is measured using machine

learning techniques. The performance accuracy is measured by applying the algorithm in Durian. The effective use of the Machine learning algorithm is to improve the classification rate and to minimize execution time. The techniques such as Support Vector Machine, K-Nearest Neighbour, Decision Tree, and Artificial Neural Network are applied in the extracted datasets of fruits. These techniques help in the identification of the performance using a graph.

The organization of this research paper is as follows: Section I deals with Introduction about the paper. Section II discusses literature review and the problem statement. Section III details the principle of the proposed method, based on feature extraction and classification methods. Section IV addresses the experimental results. The concluding remarks are provided in Section V.

II LITERATURE REVIEW

This Literature work explains in detail about the methods used for the classification of different types of fruits based on its size, shape, color, texture and ripening stages of fruits. There are a couple of image classification methods discussed henceforth. In this review the comparison is made with all the machine learning techniques used along with the performance accuracy obtained.

Using the combination of Hyper-spectral Imaging and Support Vector Machine, a novel approach was presented in [1] for the performance improvement in Strawberry fruit classification. HSI a spectral imaging technique was used to estimate the quality, ripeness, and fruit damages that were effectively applied on the strawberry. The spectral data, optimal wavelength, and texture features were extracted from HSI. The Texture features like color, roughness, intensity, firmness were extracted using the technique. The author used the SVM algorithm for classification, which by using the fruit datasets from hyperspectral images at an accuracy above 71.67%.

In this study presented in [2], the author used an automatic image processing technique for the identification of bunches of banana ripening. The ripening was done under two processes, in which a four-class dataset (Un-ripe, Semi-ripe, Mid-ripe, and Fully-ripe) of the fruit images were prepared and an artificial neural network-based framework was applied. This ANN helped in the classification of fruit based on color, identified brown spots on the fruit, and identified the Tamura statistical texture features. The optimal performance was compared with the Supervised techniques of SVM, KNN, DT, and Naive Bayes. Results reveal that the proposed ANN system has the highest overall recognition rate, which is 87.75%, among other techniques.

In this paper [3], the author used the digital imaging technique along with an effective random forest algorithm for the successful classification of papaya fruit based on its ripeness. First, the color features were computed using RGB, HSV, and L*a*b methods. By combining the image features with the machine learning algorithms a higher accuracy of the classification model was determined. In this Random forest algorithm, the two methods of cross-validation and prediction set were applied on the datasets, in which the accuracy of 94.3% and 94.7% were obtained respectively. This method gave a satisfying result based on the pulp firmness of the papaya fruit. The final normalized mean of the RGB color space achieved the performance of 78.1% with the machine learning techniques.

For detecting the ripeness of Cape Gooseberry fruit a combination of 4 machine learning techniques with 3 color spaces was proposed and evaluated in this paper [4]. Here author used a four-step pre-processing of the location of samples, image acquisition, image enhancement, and segmentation process of converting to grayscale. Then the mean value of the RGB color space was extracted and converted to HSV and Lab. Finally, the supervised algorithms of ANN, DT, SVM, and KNN were combined to classify the gooseberry fruit based on the degree of ripeness. After comparing the obtained outcome the SVM technique with the Lab color space yielded the performance based on accuracy and f-measure. The performance over the three color spaces should about 89.339% of variance and performance of above 85%.

In this [5] the author explains the determination of the non-destructive internal quality of the peer fruit. The two types of peer fruits Conference and Cepuna were used. To detect the internal disorders, the supervised algorithm of support vector machine along with X-ray Computer Tomography was proposed. Classifiers trained on 'Conference' data achieved high validation results when compared with 'Cepuna'. The classifier SVM achieved an accuracy above 85% depending on the cultivar and number of features used. This process helped to increase inspection speed and reduce the equipment costs effectively.

Author [citation]	Fruits Used	Performance	Methodology
Chu zhang,	Strawberry	85%	SVM
Fatma M. A.	Banana	87.7%	SVM, KNN, DT, ANN
Luiz Fernando	Papaya	78.1%	Random Forest
Wilson Castro	Cape Gooseberry	85%	ANN, DT, SVM, and KNN
Tim Van De Looverbosch	Peer	85%	SVM

Fig. 1. Comparison of Methods in previous study.

III PROPOSED METHOD

To overcome the problems faced by the farmers, a solution has been revised towards the introduction and advancements of innovative imaging technologies for fast,

non-invasive, and non-destructive monitoring of ripening and maturity stages of fresh fruits. The proposed methods include Dataset collection of Durian Images, Pre-processing, and Classification.

The novelty of this work is to compare all the classifying algorithms and to determine the best method of classification based on Accuracy, Precision call and F1 Score.

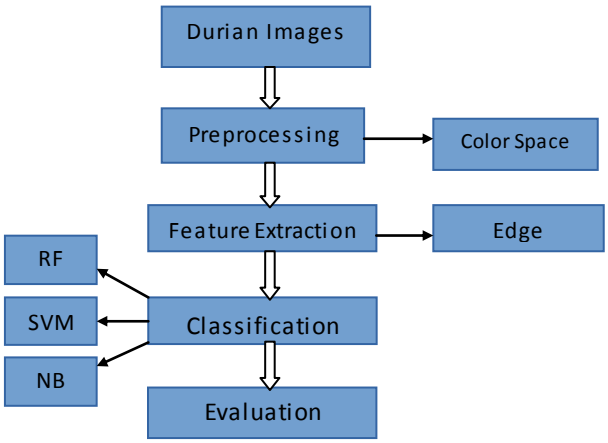

Fig. 2. System Design Model

A Image Collection :

For developing and testing the proposed algorithm, the durian fruit images were taken from the Kaggle website. From these available sets of datasets of different fruits, 300 images of durian were taken from that 180 images were taken for training and validation set, and 120 images for the test set. With the above set of durian images, the two processes of feature extraction and classification process were done. [6]

Fig. 3. Durian Images

B Pre-Processing based on color spaces:

This process is applied to perform smoothing of the durian image by resizing the image to 100*100 pixels. An image generally consists of RGB color components which represent three planes. This RGB is further converted to HSV color space to determine the mean and standard deviation. First, the RGB image is converted to a Grayscale and then to HSV. HSV color space determines the color in terms of Hue, Saturation, and Value. In which the Hue represents the color types, Saturation represents the vibrancy and Value represents the brightness calculated with the formula: [8].

The use of color space in image processing depends on two main factors of color and shades. In which color identifies the object and extracts the object from the image. The shades of an object are compared with the gray scale. The RGB channel is widely used in image processing for color monitoring and HSV is created to describe and interpret the colors. [9] HSV express the color more clearer than RGB and the calculation of determination of HSV is very simple and easy. These values help in stability of the dataset to give as the input for classification purpose. The images are resized and the HSV color analysis method is used to divide the background and durian fruit area.

Formula used for HSV :

Hue calculation:

$$H = \left(\{60^\circ * \frac{(G'-B')}{\Delta} mod 6\right) \ , C_{max} = R'$$

$$\left(\{60^\circ * \frac{(B'-R')}{\Delta} + 2\right) \ , C_{max} = G'$$

$$\left(\{60^\circ * \frac{(R'-G')}{\Delta} + 4\right) \ , C_{max} = B'$$

Saturation calculation:

$$S = \begin{cases} 0, C_{max} = 0 \\ \frac{(\Delta)}{C_{max}}, C_{max} \neq 0 \end{cases}$$

Value calculation: $V = C_{max}$ ----------(2)

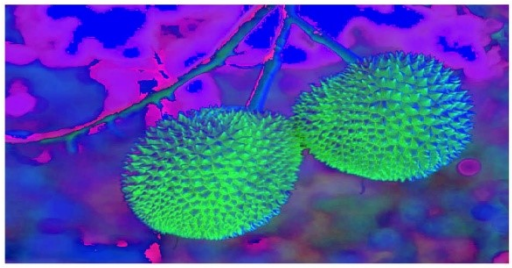

Fig. 4. Color Detection of Durian Images

C. Feature Extraction based on edge detection :

The purpose of this method is used to significantly determine the edge of the durian fruit with that of the background images. The Canny Edge Detector is a technique that uses multistage in algorithms to determine the edges of the images. The steps involved are:

(a) Smoothing of the durian image by applying Gradient estimation and to reduce the noise with the formula.

(b) Canny filter is used for edge detection. This process of edge detection is identified by works by calculating the gradient of image intensity at every pixel of an image.

(c) Finally double threshold is applied to determine the potential edges and the final edges are tracked by hysteresis.[7]

The Gradient of the image is calculated to determine the edge intensity and the direction of the image. To detect the edge the easiest way is to apply filters in two direction of horizontal (X) and for vertical (Y). When the image is smoothed, the $I_x I_y$ are calculated as in (1):

$$Gradient\ (G) = \sqrt{I_x^2 + I_y^2}$$

$$\Theta\ (x,y) = arctan\left(\frac{I_y}{I_x}\right) \qquad \text{----------(1)}$$

Fig. 5. Canny Edge Detection of Durian Images

D. *Methodology :*

After preprocessing, the durian images are trained and validated using the Non-Destructive Techniques. The Machine Learning algorithms namely SVM, Naive Bayes, and Random Forest algorithm are used to compare the classification accuracy of the ripening of Durian Fruit.

These supervised algorithms analyze the given training dataset. The algorithms like Linear Discriminant Analysis, Naive Bayes, Logistic Regression, and Simple Neural Networks are more simple, easy to understand, very fast to learn from the given data. While these are more constrained and have limited complexity of working with simpler problems with a minimum number of the dataset. On the other hand, algorithms like K-Nearest Neighbour, Decision tree with CART, and SVM are more flexible and give better accuracy in the model for prediction. But these algorithms are of risk in overfitting of the training data.

1) *Support Vector Machine*

SVM is a supervised learning algorithm that can be used to analyze the data, identify the pattern, and can be used for classification and regression process. This helps in multi-class classification purposes. SVM helps to determine the decision boundary based on hyper-plane, classified as Linear and Non-linear.[10] This helps to determine the optimal hyperplane that can classify data in 2-dimensional data space as by the formula:

$$g(x) = w^T x + b \qquad \text{---------(3)}$$

The classification decision formula is given as :

$$f(x) = sing(w^T x + b) \qquad \text{---------(4)}$$

Where the value for f(x) determines the classification of data samples, in which if f(x)=0 then the data is on the hyper plane and if the value is lesser than 0 then the data is in the negative hyper plane and if the value is greater than 0 then the data is in the positive hyper plane. These are illustrated in the figure (5).

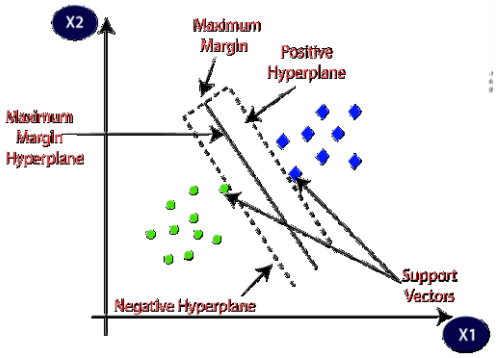

Fig. 6. Support Vector Machine Model

The performance measure with SVM shows that there is about 89.33% accuracy in the classification of the fruits. The Durian images are classified based on maturity level. With the training sample of 75 % and testing of 25% the classification showed a better result.

From the literature survey, it is found that 80-85% classification is obtained using a logistic support vector machine algorithm. In Figure 6, the classification accuracy model for the SVM model is shown based on sample data with 89.33%. From the survey, it is accepted that the prediction of data using SVM with a result above 80% is considered to be feasible and efficient.

```
[ ] from sklearn.svm import SVC
    from sklearn.pipeline import make_pipeline
    from sklearn.preprocessing import StandardScaler

    clf = make_pipeline(StandardScaler(), SVC(gamma='auto'))
    clf.fit(train_X, train_y)
    Pipeline(steps=[('standardscaler', StandardScaler()),('svc', SVC(gamma='auto'))])
```

```
    out = clf.predict(test_X)
    print( accuracy_score(test_y, out) * 100)

    89.33333333333333
```

Fig.7 : Classification Accuracy of SVM

2) *Gaussian Naive Bayes Algorithm*

GNB algorithm is based on Bayes theorem which has a common principle of every pair of features that is independent

of the other features which are being classified. This GNB stores all the available data and classifies a new data point based on the similarity and it is easy to build. This can be used for both the Classification and Regression process. GNB is easy to implement, robust to noisy training data also effective for a huge volume of data. The Naive Bayes model is useful for very huge data sets. This algorithm works well for highly sophisticated classification methods.[11] By applying this algorithm to the extracted images of Durian fruit, it has been determined a minimum accuracy of about 65.3% illustrated in the Fig(7).

$$P(A/B) = \frac{P(B/A) * P(A)}{P(B)} \qquad \text{-------(5)}$$

```
from sklearn.naive_bayes import GaussianNB
gnb = GaussianNB()
y_pred = gnb.fit(train_X, train_y)
out = gnb.predict(test_X)
print( accuracy_score(test_y, out) * 100)

c:\users\aksha\appdata\local\programs\python\python36\lib\site-packages\sklearn\naiv
  y = column_or_1d(y, warn=True)
65.33333333333333
```

Fig.8 : Classification Accuracy of GNB

3) Random Forest Algorithm

A random forest is a supervised model that helps to classify the unlabeled data with the trained label set of data. This fits for both classification and regression tree to the dataset and then combines the prediction from all the datasets.. This algorithm prevents overfitting of the data and it helps to create more accurate results. This algorithm also helps with the selection of many bootstrap samples from the dataset.[12]

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.datasets import make_classification
clf = RandomForestClassifier(max_depth=2, random_state=0)
clf.fit(train_X, train_y)
print(accuracy_score(clf.predict(test_X),out)*100)

c:\users\aksha\appdata\local\programs\python\python36\lib\site-packages\ipykernel_la
  after removing the cwd from sys.path.
84.33333333333334
```

Fig. 9. Classification Accuracy of Random Forest Algorithm

When the datasets are trained with this algorithm, it gives an accuracy of about 84.3% is obtained with the durian dataset of classification based on ripeness illustrated in fig(8).

IV RESULT ANALYSIS

The results of Non-Destructive Machine Learning techniques are evaluated based on the accuracy obtained. The ripening classification of the durian fruit is finally determined. The algorithms of SVM gives an accuracy of about 89.3%, the GNB gives an accuracy of about 65.3% and Random forest yields about 84.3% accuracy. This highest performance of support vector machine gives a better classification of the durian fruit into Unripen, Semi ripens, and Fully ripen based on color. The henceforth diagrams explains about the graphical representation of the classification results.

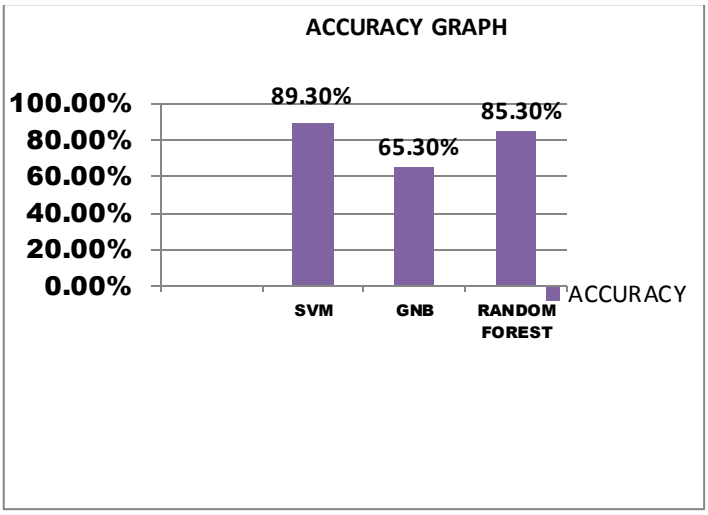

Fig.10. Accuracy Graph of Algorithm

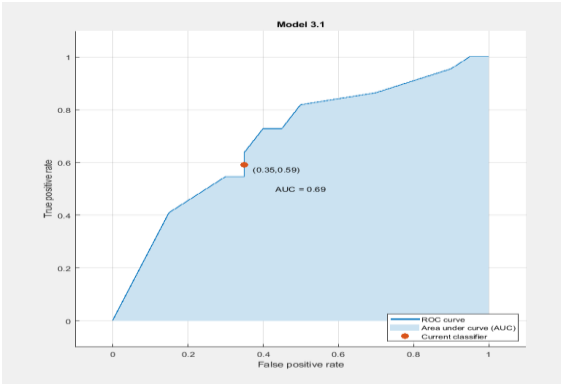

Fig.11. Graphical Representation of Algorithm

There are four performance measures computed after the classification degree of ripeness of Durian fruit to measure the color space on the classification based on machine learning techniques. They are determination of Accuracy, Precision, Recall and F Measure which was applied on the three algorithms of SVM, GNB and Random forest and the results are tabulated. The above parameters are calculated using the formula:

$$Accuracy = \frac{True\ Positive + True\ Negative}{Posistive + Negative} * 100\ \% \quad --(6)$$

$$Precision = \frac{True\ Positive}{True\ Posistive + False\ Positive} * 100\% \quad --(7)$$

$$Recall = \frac{True\ Positive}{True\ Posistive + False\ Negative} * 100\% \quad ----(8)$$

Algorithm	Accuracy	Precision	Recall
SVM	89.3%	95%	79.33%
GNB	65.3%	75%	83.67%
RF	84.3%	70.67%	90%

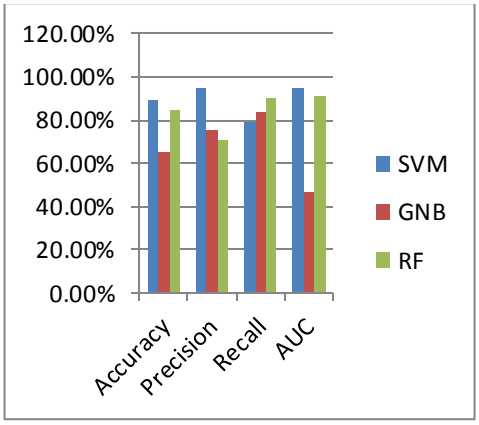

Fig: 12 Graph representation of Algorithms

V CONCLUSION

From the results obtained using state of art techniques of machine learning algorithms can be finalized that the fruits can be classified efficiently based on size, shape, color, and texture. From this work, the SVM gives a better result of accuracy. In this work, HSV was applied to evaluate the ripeness degree of the durian fruit. SVM model of the spectral analysis datas showed acceptable results. Color features and edge detection helped to determine the accuracy with the gray scale images at an optimal wavelength. SVM model using the 300 combined dataset samples of ripe and unripe showed best results of 89.3% of classification. However the ripe and unripe samples were rarely misclassified with each other.

In future study, more than 500 samples with wider variations of external and internal quality of durian fruits at different ripeness stage will be needed for better identification of strawberry ripeness. Further enhancement in this work would be classification based on ANN. ANN is used as a highly successful tool for dataset classification with more number of sample images. The dataset are the suitable combination of training, learning and testing which yields in better accuracy.

REFERENCES

[1] Chu zhang, Chentong Guo,(2016) "Hyperspectral imaging analysis for ripeness evaluation of strawberry with support vector machine"

[2] Fatma M. A. Mazen, Ahmed .A.Nasant (2018) "Ripeness Classification of Bananas Usingan Artificial Neural Network"

[3] Luiz Fernando Santos Pereiraa, Sylvio Barbon (2018) "Predicting the ripening of papaya fruit with digital imaging and random forests".

[4] Wilson Castro, Jimy Oblitas (2019) "Classification of Cape Gooseberry Fruit According to Its Level of Ripeness Using Machine Learning Techniques and Different Color Spaces ".

[5] Santi Kumari Behera, Amiya Kumar Rath (2020) "Maturity status classification of papaya fruits based on machine learning and transfer learning approach ".

[6] Fruits Dataset of images from https://www.kaggle.com/moltean/fruits.

[7] https://towardsdatascience.com/canny-edge-detection-step-by-step-in-python-computer-vision-b49c3a2d8123

[8] Bhanu Pratap , Navneet Agarwal,(2014) "Development of ANN based Efficient Fruit Recognition Technique".

[9] https://globalaihub.com/image-processing-color-spaces-rgb-hsv-and-cmy

[10] Hongxing Peng , Yuanyuan Shao, (2018), "Research on Multi-class Fruits Recognition Based on Machine Vision and SVM".

[11] Dayakar L.Naik[a] , RaviKiran,(2018), " Naïve Bayes classifier, multivariate linear regression and experimental testing for classification "

[12] Shinji Fukuda , Eriko Yasunaga (2014), "Modelling the relationship between peel colour and the quality of fresh mango fruit using Random Forests".

[13] Lanhui Fu, Jieli Duan (2019), "Banana detection based on color and texture features in the natural environment".

[14] Nidhi Goel ,Priti Sehga(2015), "Fuzzy classification of pre-harvest tomatoes for ripeness estimation – An approach based on automatic rule learning using decision tree."