IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

SEGMENTATION AND CLASSIFICATION OF BLACKBERRIES BASED ON MATURITY LEVEL

Shaistha Anjum F [1], Dhanesha R[2] Pooja Nagaraja Barki [3], Mehar Taj S M[4].

[1,2,3,4]DOS in Computer Science, Davangere University, Davangere, Karnataka, India

Abstract: Indian Blackberries (Syzygium Cumini), commonly known as jamun or black plum, are a popular fruit in India, it is small, dark purple to black fruit, valued for its distinct taste and medicinal properties. The determination of fruit maturity is a critical aspect of agricultural practices, impacting fruit quality, harvest timing, and market value. Proposed study explores an innovative approach to non-destructive maturity detection using image processing techniques. High-resolution images of Indian blackberries at various ripeness stages are analyzed for color and texture features. Here, Segmentation plays an important role and it is performed by using Threshold method. Using the Randomizable Filtered classifier it classifies the fruit into distinct maturity levels with accuracy level of 96.61%. Proposed research offers a promising solution to enhance the precision of blackberry harvesting, minimize post-harvest losses, and support informed decision-making for growers, contributing to the sustainable cultivation of Indian blackberries and improved fruit quality for consumers.

Keywords: Blackberries, correlation, entropy, homogeneity, energy, maturity, immature, classifier, Randomizable Filtered.

1. Introduction

The Jamun is an evergreen, fruit plant in the Myrtaceae family with oblong opposite leaves that are smooth, glossy and have a turpentine smell. Jamun has fragrant white flowers in branched clusters at stem tips and purplish-black oval edible berries [1]. A slow-growing species, it can reach heights of up to 30 m and can live more than 100 years. Its dense foliage provides shade and is grown just for its ornamental value. At the base of the tree, the bark is rough and dark grey, becoming lighter grey and smoother higher up. The wood is water-resistant. Because of this it is used in railway sleepers and to install motors in wells. It is sometimes used to make cheap furniture and village dwellings though it is relatively hard to work on.

Indian Blackberry trees start flowering from March to April. The flowers are fragrant and small, about 5 mm in diameter. The fruits develop by May or June and resemble large berries. The fruit is oblong, ovoid. Unripe fruit looks green. As it matures, its colour changes to pink, then to shining crimson red and finally to black colour, a variant of the tree produces white coloured fruit. The fruit has a combination of sweet, mildly sour and astringent flavour and tends to colour the tongue purple. The ripe fruits are used to produce wine, squash, jellies, and health beverages. All components of the tree, but most significantly the seeds, are used to manage diabetes mellitus in connection with its nutritional usage. Jamun has antioxidant, anti-inflammatory, anti-HIV, an-tileishmanial and antifugal, nitric oxide scavenging, free radical scavenging, anorexigenic, gastroprotective, anti-ulcerogenic, and redio-protective effects [2] [3].

Traditionally, assessing the maturity of Indian blackberries has relied on subjective visual inspection or manual sampling techniques. However, these methods can be time-consuming, labour-intensive, and prone to human error. In recent years, advancements in image processing technology have provided a promising alternative for non-destructive and efficient maturity detection in agricultural produce. This study explores the application of image processing techniques, with a focus on color and texture analysis, to detect the maturity levels of Indian blackberries. By harnessing the power of digital imaging and computational analysis, this approach aims to offer a more accurate, consistent, and objective means of assessing fruit ripeness. The objectives of this research include the development of image-based algorithms that can differentiate between various stages of Indian blackberry maturity, from unripe to fully ripe. These algorithms will rely on the extraction of critical features, such as colour and texture patterns, from high-resolution images of the fruit. Machine learning models, trained on labelled datasets, will play a pivotal role in classifying the maturity levels based on the extracted features.

The potential benefits of this image processing-based maturity detection system for Indian blackberries are manifold. It promises to enhance the precision and efficiency of harvesting operations, reduce post-harvest losses, and ensure that consumers consistently receive high-quality fruit. Additionally, it can provide valuable insights to growers and farmers, enabling them to make data-driven decisions to optimize crop management practices. This research represents a step forward in the intersection of agriculture and technology, demonstrating how image processing can revolutionize the way Indian blackberries is cultivated and harvested, ultimately benefiting both the agricultural industry and consumers. Through this study, we aim to contribute to the advancement of precision agriculture in India, offering a more sustainable and economically viable approach to fruit production. India is well known for agricultural country wherein about 70% of the population depends on agriculture. Image processing is a technique that can be used to extract information from images. This information can then be used to classify the images or to perform other tasks. Image processing can be used to detect the maturity of Indian blackberry by extracting features from the leaves' images, such as their colour, texture, and shape. These features can then be used to train a machine learning algorithm to classify the images as mature or immature.

A. Benefits of using Image processing to identify the maturity level of Black Barriers

- Non-Destructive Assessment: Image processing allows for non-destructive assessment of fruit maturity. Unlike traditional methods that require manual sampling or destructive testing, images can be captured without harming the fruit, preserving the overall crop.
- Objective and Consistent: Image processing provides an objective and consistent means of assessing maturity levels. It eliminates human subjectivity and the potential for errors associated with visual inspection.
- Efficiency: The process is efficient and can quickly analyze a large number of fruits in a relatively short amount of time. This is especially valuable for commercial growers dealing with large harvests.
- High Precision: Image processing techniques can detect subtle changes in color, texture, and shape that are indicative of different maturity stages, offering high precision in classification.
- Market Competitiveness: Delivering consistently ripe and high-quality fruit to the market can enhance the competitiveness of growers and suppliers in the industry.

Overall, using image processing for Indian blackberry (jamun) maturity detection not only improves the efficiency and accuracy of the assessment process but also has the potential to positively impact crop management, reduce waste, and enhance the quality of the fruit delivered to consumers. The accuracy of the image processing system can be improved by using a better quality camera, by taking the images in a controlled environment, and by using a more sophisticated machine learning algorithm.

A. Related Work

Computer vision and image processing are rapidly becoming more common in precision agriculture. For fruit ripening classification, a variety of approaches based on shape, size, colors, and texture have been devised. Different ways for segmenting fruits to retrieve the features needed for the classification process were also available. Wan, Peng, et al.[4], proposed a method for detecting the maturity levels (green, orange, and red) of fresh market tomatoes (Roma and Pear varieties) by combining the feature color value with the backpropagation neural network (BPNN) classification technique.

Surya Prabha, D., et.al. [5], study attempted to use image processing technique to detect the maturity stage of fresh banana fruit by its color and size value of their images precisely. The mean color intensity from histogram; area, perimeter, major axis length and minor axis length from the size values, were extracted from the calibration images. The mean color intensity algorithm showed 99.1 % accuracy in classifying the banana fruit maturity. The area algorithm classified the under-mature fruit at 85 % accuracy.

Van Huy Pham and Byung Ryong Lee [6], Proposed a method for detecting deficiency in orange fruits was presented. Defects in orange fruits were detected using a segmentation technique. Graph-based and k-means clustering approaches were employed to classify defects in orange fruits.

X.E. Pantazi, et.al [7] proposed an automated approach for detecting leaf disease in numerous crop species by evaluating image features and deploying a classifier. The feature extraction was done with Local Binary Pattern (LBP), and the illness leaf classification was conducted with a support vector machine (SVM) classifier. Yanan Li, Zhiguo Cao, et.al [8], presented a technique for detecting in-field cotton using image segmentation. Unsupervised region creation and supervised semantic labeling prediction were utilized in conjunction with region-based segmentation. Each region was segmented using histogram-based color and texture features.

NST Saia, Ravindra Patil, et.al [9], developed a content-based image retrieval system which used two distinct feature vector methods. For gray scale, RGB, and YCbCr color images, the SVD (singular value decomposition) feature of increasingly truncated DCT (Discrete cosine transform) images and the DWT (Discrete wavelet transform) decomposed image were computed. The relevant images from the dataset were extracted based on color feature. Yaqoob Majeed, Jing Zhang, et.al [10], proposed a segmentation method to segment Apple trunk and branch. The images were stored in cloud from Kinect V2 sensor and using deep learning-based semantic method the images are segmented. To remove the background trees Depth and RGB features are extracted from cloud data.

By using the fundamental K nearest neighbour (KNN) model, Daneshwari Ashok Noola et.al. [11] concentrated on designing and developing the enhanced-K nearest neighbour (EKNN) model. EKNN is used to differentiate between illness classes. High-quality fine and coarse features are produced to gather discriminative, boundary, pattern, and structurally linked information, which is then utilized in the classification process. Gradient-based characteristics of excellent quality are provided by the classification procedure. Uoc Quang Ngo et.al. [12], devised a technique for precisely calculating the leaf area of cucumber plants using digital image processing. The suggested ways extract the cucumber plant's skeleton from RGB images and correctly estimate the leaf area of cucumber plants.

Megha.P.Arakeria, Lakshmana [13], To assess tomato quality, a computer vision method with 2 phases was developed. The hardware was created in the initial phase to collect tomato images and transport tomato fruit to suitable containers. The software was developed in the second phase to detect defects and ripeness in tomato fruit using image processing techniques. The RGB color model was used to determine ripeness. Yang Yu and Sergio A. Velastin, et.al [14], proposed a quick and efficient approach for achieving automated apple grading . For each apple sample, four images were taken (top, bottom and two sides). To differentiate the apple deformities, stem, and calyx, the grey value of each apple was acquired. K-means clustering was utilized to find the defective zone in an apple. Santi Kumari Behera, et.al [15], applied KNN, SVM and Naive Bayes classifier to classify the papaya fruits based on maturity status. Features such as LBP, Histogram of Oriented Gradients (HOG) and Gray Level Co-occurrence Matrix (GLCM) were extracted from papaya fruit image to fed as input to classifier.

Suresha M, et.al [16], presented a method to classify the diseased arecanut using texture features. The LBP , Haar Wavelets, GLCM and Gabor features of texture were used to determine the diseased arecanut. The HSI(Hue Saturation Intensity) and YCbCr color models were used. The LBP, Haar Wavelets, GLCM and Gabor methods were applied on HSI and YCbCr color model to extract the texture features. The KNN(K-nearest neighbor) classifier was used to classify the diseased and un-diseased arecanut.

Ajit Danti, et.al [17], presented a method for separating raw arecanuts into two categories. Red and green colors were used to determine upper and lower limits for the classification of raw arecanuts. Dhanesha R., et.al [18], introduced a brand-new method for segmenting arecanut bunches using the active contour method. The segmentation methodology was evaluated using the segmentation performance techniques VOE and DSC. Umesha D.K., et.al [19], proposed a study of different color models to segment the arecanut bunches. In order to separate the arecanut bunch from an input image, the HSV, YCbCr, YUV, YCgCr and YPbPr color models are applied to arecanut bunch images with manual threshold. Color models HSV and YCgCr were efficient in segmenting arecanut bunches from other color models used in the study.

According to the literature survey it is observed that the image processing and machine learning plays vital role in the agriculture field to classify the maturity level of vegetables and fruits. So, in this proposed work the image processing techniques are applied to classify the blackberries based on maturity level.

B. Feature Extraction for classification

A wide variety of classification processes employ color and texture feature analyses. The maturity levels of Blackberries were classified using color and texture features. To classify the matured and un-matured Blackberries the three color features (i.e., the Average intensity on the R, G, and B bands: RAvg, GAvg, and BAvg) were calculated by applying the Equations.

RAvg = sum of (red pixels) / the entire amount of pixels occupied by Blackberries Image (1)

GAvg = sum of (green pixels) / the entire amount of pixels occupied by Blackberries Image (2)

BAvg = sum of (blue pixels) / the entire amount of pixels occupied by Blackberries Image (3)

Texture Features : To classify the Blackberries along with color features contrast, entropy, correlation, homogeneity and energy features of texture were used.

C. Classifier

Randomizable Filtered classifier: There are a variety of classification problems that can be solved using he Randomizable filtered classification method. The Randomizable Filtered classifier is a type of classifier in Weka that combines a filter and a classifier to improve classification accuracy. It works by first applying a filter to the input data to remove irrelevant or redundant attributes. The filtered data is then passed to a classifier, which learns a model based on the remaining attributes and their corresponding class labels.

The Randomizable Filtered classifier is called "randomizable" because it allows users to randomize the order in which the filter and classifier are applied. This can help to reduce overfitting and improve generalization performance, as different combinations of filter and classifier may be more effective on different subsets of the data. The RandomizableFiltered classifier is also called "filtered" because it uses a filter to preprocess the input data. Weka provides a range of filters that can be used with the RandomizableFiltered classifier, including attribute selection filters (such as Principal Component Analysis or Chi-squared attribute evaluator), instance selection filters (such as Random subsampling or SpreadSubsample), and attribute transformation filters (such as Discretize or Normalize).

One of the main advantages of the RandomizableFiltered classifier is its ability to reduce the dimensionality of the input data while preserving or even improving classification accuracy. By removing irrelevant or redundant attributes, the classifier can focus on the most informative features of the data, which can lead to better generalization performance and faster training times.

D. Classification Performance metrics

The confusion matrix is used to evaluate classification performance based on elements from the matrix. To compare class labels, the terms TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative) are employed. The precision and accuracy of the classifier were estimated based on the values acquired for the confusion matrix in the proposed study to evaluate its performance.

- True Positive: These are cases in which predicted as matured (Yes), and they were actual Matured.

$$\text{TP Rate} = \text{TP} / \text{Actual Yes} \qquad (4)$$

- True Negative Rate: These are cases in which predicted as un-matured (No), and they were actually un-matured.

$$\text{TN Rate} = \text{TN} / \text{Actual No} \qquad (5)$$

- False positive Rate: Predicted matured, but they actually un-matured. (Also known as a "Type I error").

$$\text{FP Rate} = \text{FP} / \text{Actual No} \qquad (6)$$

- False positive Rate: Predicted un-matured, but they actually matured. (Also known as a "Type II error.").

$$\text{FN Rate} = \text{FN} / \text{Actual Yes} \qquad (7)$$

- Precision Rate: When it predicts matured, how often is it correct?

$$\text{Precision Rate} = \text{TP} / \text{Predicted Yes} \qquad (8)$$

- Accuracy: The most typical indicator to assess how well the categorization process is working. The ratio of properly identified samples to the total number of samples is used to determine accuracy.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \qquad (9)$$

2. Methodology

This section describes the Database used for experimentation. The segmentation techniques to segment Blackberries using manual thresholding were discussed. And also the tool and the classifier used to classify the Blackberries based on maturity level were mentioned.

A. Database

The Database contains 59 colored images which are captured with help of a smart phone fitted on selfie stick. These images were taken at different time of the day.

B. Segmentation and Classification: Segmentation divides an image into its constituent regions or objects. The level of detail to which the subdivision carried depends on the problem using being solved. i.e., segmentation should stop when the objects or regions of interest in an application have been detected. The input image is segmented using manual thresholding. In this proposed work the sobel mask with edge detection algorithm is used to segment the image. The output of edge segmentation is applied with morphological operations such as hole filling, than the median filter is applied to filter the image to get the region of interest. The image is transformed back to its RGB format. RGB components were recovered independently from the RGB segmented image to compute the average of each component. To extract contrast, entropy, correlation, homogeneity and energy texture features, an RGB segmented image is transformed to a gray scale image and using GLCM function available in Matlab these features were extracted. The computed feature values were saved in a.csv file.

The .csv file is used as an input by the Weka 3.8.4 utility. The black barriers are classified depending on their maturation stage using a RandomizableFiltered classifier. The values of the confusion matrix were used to calculate precision and accuracy metrics to evaluate the performance of classification.

3. Experiment and Results

A. Experiment

In order to categorize the maturity level of 59 images of immature and mature black barriers, experiments were conducted. To classify them, certain features were required.

(a) (b)

Fig 1: Segmentation Process of a) Immature Black Barry fruit b) Matured Black Barry Fruit

The Figure 1., shows the output of segmentation process. After segmentation color and texture features were taken from segmented images to execute the classification procedure. To extract the RGB components of an image, a segmented image is transformed to an RGB image. Color and texture features were extracted using GLCM method of MATLAB 14b.

The retrieved feature values were entered into the Weka 3.8.4 tool (.csv file). The experiment was carried out using aRandomizableFiltered classifier. The parameters set in RandomizableFiltered classifier are as follows:

- **Cross-Validation Fold:** is a method of the Evaluation class that is employed to carry out cross-validation using a single dataset and an untrained classifier. Cross-Validation Fold is given a value of 10 (default value).
- **BagSizePercent:** Size of each bag, as a percentage of the training set size. Value assigned is 100.
- **BatchSize:** The preferred number of instances to process if batch prediction is being performed. Value assigned is 100.
- **NumIterations:** The number of trees in the random forest. Value assigned is 100.

Other parameters were set to default value assigned by Weka 3.8.4. The findings were discussed in the section Results and Discussion.

4. Results and Discussions.

Blackberries are first segmented to extract required features before being classified according to maturity level. With dataset images provided with the proposed method, the segmentation results are compared to manually graded ground truth (GT). The value of the retrieved features is saved in .csv file with six columns labelled RAvg, GAvg, BAvg, Correlation, Entropy, and Class. There are two values in the Class column. Yes or No. Yes denotes matured Blackberries, whereas No denotes immature Blackberries.

Table 1: Performance of classifier evaluation using Confusion Matrix

N=1017	Predicted NO	Predicted YES	Correctly Classified	Precision	Accuracy(%)
Actual YES(42)	FN=0	TP=42	42	0.9661	**96.6%**
Actual No(17)	TN=15	FP=02	15		
Total(1017)	15	44	59		

The performance of classification is assessed using precision and accuracy metrics based on confusion matrix values, as shown in table 1. Table 1, shows that 57 samples out of 59 dataset samples were correctly classified, resulting in a **96.61%** of accuracy.

5. Conclusion

The detection of maturity level in Indian blackberry (jamun) using image processing is a promising and efficient method for assessing the ripeness of the fruit. By leveraging methods such as colour analysis, texture assessment, and machine learning algorithms, it can accurately and non-destructively assess the ripeness of jamun based on visual attributes. Used technology holds great promise in streamlining fruit quality control and grading processes, reducing waste, and ensuring consumers receive high-quality produce. Further research and refinement of these image processing methods can lead to the development of automated systems that enhance the efficiency and precision of fruit maturity detection, benefiting both growers and consumers in the agricultural sector. The results suggest that image processing can be used to automate the process of Indian blackberry maturity detection. A method for automatically detecting the maturity of jamun using image processing techniques is developed. By collecting a dataset of images of Indian blackberry at different stages of maturity, and used the dataset to train a MATLAB. It was able to achieve an accuracy of 96.61% in classifying the maturity of berries.

References

1. Aqil, Farrukh, et al. "The Indian blackberry (Jamun), antioxidant capacity, and cancer protection." Cancer. Academic Press, 2014. 101-113.
2. Ghosh, Payel, et al. "Quantification and concentration of anthocyanidin from Indian blackberry (Jamun) by combination of ultra-and nano-filtrations." Food and Bioprocess Technology 11 (2018): 2194-2203.
3. Ramteke, Vikas, Vivek Kurrey, and Sonali Kar. "Jamun: A traditional fruit and medicine." Popular Kheti 3.3 (2015): 188-190.
4. Wan, Peng, et al. "A methodology for fresh tomato maturity detection using computer vision." Computers and electronics in agriculture 146 (2018): 43-50.
5. Surya Prabha, D., and J. Satheesh Kumar. "Assessment of banana fruit maturity by image processing technique." Journal of food science and technology 52 (2015): 1316-1327.
6. Van Huy Pham and Byung Ryong Lee. An image segmentation approach for fruit defect detection using k-means clustering and graph-based algorithm. Computer Science, 2:25–33, 2015.
7. A.A. Tamouridou X.E. Pantazi, D. Moshou. Automated leaf disease detection in different crop speciesthrough image features analysis and one class classifiers. Elsevier Journal of Computers and Electronicsin Agriculture, 156:96–104, 2019.

8. Yanan Li, Zhiguo Cao et.al. In field cotton detection via region-based semantic image segmentation. Elsevier Journal of Computers and Electronics in Agriculture, 127:475–486, 2016.
9. NST Saia, Ravindra Patil et.al. Truncated dct and decomposed dwt svd features for image retrieval. In Proc. Elsevier Conf. 7th International Conference on Communication, Computing and Virtualization, volume 79, page 579 – 588, 2016.
10. Yaqoob Majeed, Jing Zhang et.al. Apple tree trunk and branch segmentation for automatic trellis training using convolutional neural network based semantic segmentation. In Proc. Elsevier Conf. IFAC(International Federation of Automatic Control) PapersOnLine, pages 51–17, 2018.
11. Daneshwari Ashok Noola et.al., "Corn leaf image classification based on machine learning techniques for accurate leaf disease detection", Indonesian Journal of Electrical Engineering and Computer Science, Vol. 12, No. 3, June 2022, pp. 2509-2516, ISSN: 2088-8708, DOI: http://doi.org/10.11591/ijece.v12i3.
12. Uoc Quang Ngo et.al., "Digital image processing methods for estimating leaf area of cucumber plants", Indonesian Journal of Electrical Engineering and Computer Science, Vol 25, No. 1, January 2022, pp. 317-328, ISSN: 2502-4752, DOI: http://doi.org/10.11591/ijeecs.v25.i1.
13. Lakshmana Megha.P.Arakeria. Computer vision based fruit grading system for quality evaluation of tomato in agriculture industry. In Proc. Elsevier Conf. 7th International Conference on Communication, Computing and Virtualization, volume 79, page 426 – 433, 2016.
14. Yang Yu and et.al Sergio A. Velastin. Automatic grading of apples based on multi-features and weighted k-means clustering algorithm. Information Processing in Agriculture, 2019.
15. Santi Kumari Behera et.al. Maturity status classification of papaya fruits based on machine learning and transfer learning approach. Information Processing in Agriculture, 2020.
16. S. K Narasimhamurthy Suresha M, Ajit Danti. Classification of diseased arecanut based on texture features. Computer Applications (0975 – 8887) Recent Advances in Information Technology, 2014.
17. Suresha M Ajit Danti. Segmentation and classification of raw arecanuts based on threesigma control limits. Elsevier Journal of C3IT-2012, 4:215 – 219, 2012.
18. Shrinivasa Naika C. L. Dhanesha R. A novel approach to segment an arecanut bunches using active contour method. In Proc. Springer,2nd International conference on Integrated Intelligent Computing, Communication And Security, pages 677–682, 2019.
19. R. Dhanesha, D. K. Umesha, C. L. Shrinivasa Naika and G. N. Girish, "Segmentation of Arecanut Bunches: a Comparative Study of Different Color Models," 2021 IEEE Mysore Sub Section International Conference (MysuruCon), 2021, pp. 752-758, doi: 10.1109/MysuruCon52639.2021.9641680.