*As per UGC guidelines an electronic bar code is provided to seure your paper*

*International Journal for Modern Trends in Science and Technology*, 8(06): 511-516, 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0806088
Available online at: http://www.ijmtst.com/vol8issue06.html

# Machine Learning based Prediction Model for Consumability Analysis of Mango Fruits

**Laxmi V[1] | Roopalakshmi. R[2]**

[1]Department of Information Science and Engineering, BNMIT, Bengaluru – 560085, Karnataka, India,
[2]Department of Computer Science and Engineering, MIT, Manipal, 576104, Karnataka,
Email: vlaxmiachar@gmail.com

**To Cite this Article**
Laxmi V and Roopalakshmi. R. Machine Learning based Prediction Model for Consumability Analysis of Mango Fruits. International Journal for Modern Trends in Science and Technology 2022, 8(06), pp. 511-516. https://doi.org/10.46501/IJMTST0806088

**Article Info**
Received: 15 May 2022; Accepted: 14 June 2022; Published: 22 June 2022.

## ABSTRACT

*India is one of the chief producers of mango fruit. Around 50% of world's mango production comes from India. Mango is a seasonal fruit and has an abundant source of nutrition, because of which it is highly consumed by the people. However, most of the greedy vendors will use chemical ripening agents like calcium carbide to increase the ripening rate of the fruit. Some studies show that around 75% of mango fruits are chemically ripened. Consumption of chemically ripened mangos will cause a lot of damage to the health. In the existing literature only image processing techniques are used which focuses primarily on image-based features for identification of chemically ripened mangoes. A major drawback with image processing techniques is that consumabilty of the fruit is neither predicted nor explored. To solve this issue a Machine Learning model is proposed in this article which makes use of most popular machine learning techniques to analyze consumability aspects of given mango fruit. Implementation of the proposed framework is done in 2 stages by employing four different machine learning techniques for consumability analysis of mango fruit. Four variety of mangoes are considered to demonstrate the significant performance of proposed frame work in terms of peer results and time complexity.*

*KEYWORDS:Machine Learning. AdaBoost, SVM, Decision Tree, Random Forest, MPEG-7, Mango fruit.*

## 1. INTRODUCTION

Mango is one of the most popular seasonal fruits all around the world. Indian Mangoes fulfill 50% of worlds Mango [1] demand. Mangoes are normally picked when they are unripe and allowed to ripen naturally. Due to rising client demand around the world, some greedy suppliers employ chemical ripening agents like calcium carbide to speed up the ripening process. Almost 50% of mangoes are ripened using such ripening agents [2]. Unfortunately, using such chemical ripening agents will cause a lot of damage to the health of consumers to a greater extent. Computer vision based techniques facilitates the consumers to identify such fruits at a faster rate so that consumers can avoid the consumption of such fruits.

Present literature uses image processing techniques which can only identify if a fruit is defective or diseased. Some computer vision based techniques focuses on quality assessment of the fruit. There are machine learning techniques to predict ripening stages of the fruit as well. Hence, there is a scope to conduct consumability analysis for mangoes which will ensure the health of cusumer.

However, in order to protect the health of consumers, a chemically ripened Mango fruit consumability prediction system based on machine learning model has to be para mounted. The proposed model will work in two stages. In stage 1 MPEG-7 based color features are used to identify chemically ripened fruits. In stage 2 machine learning model is used to do consumability analysis.

## STRUCTURE OF PAPER

The following is a breakdown of the paper's structure: Fruit classification, quality assessment, ripening stages of fruit, defect and disease detection in fruits are all covered in Section 2. The proposed mango fruit consumability analysis model is presented in Section 3. In Section 4, the model's performance and evaluation are presented. Section 5 concludes with closing remarks and a point of view.

## OBJECTIVES

This article proposes consumability analysis of mango fruits. To achieve this MPEG-7 based features are used to identify chemically ripened fruits. Then, machine learning model is used to perform consumability analysis of mango fruit. Based on the experimental results of proposed machine learning model, if the consumability analysis is greater than 60% then, the fruit is categorized as not good to consume. Else it is categorised as good to consume.

## 2. RELATED WORK

There are numerous works that have been done related to machine learning algorithms.

The usage of image processing and computer vision has risen dramatically in the agricultural industry. For fruit classification, excellence evaluation, flaw identification, and leveling issues, a variety of methodologies are being debated. The majority of the approaches rely solely on colour and shape attributes. Using GANFIC (Genetic Adaptive Neuro Fuzzy Inference System), Anurekha and Sankaran present a classification and grading system for mangoes [3]. Shape, texture, and colour are extracted from 2D Mangoimages using this method. The method uses a genomic algorithm for feature assortment and a ambiguous inference method for classification and grading. To perform classification, the approach also examines multi-feature class similarity metrics. This method has a sensitivity of 98.05 percent, a specificity of 97.39 percent, and a precision of 99.18 percent. Machine learning approaches, on the other hand, may help to minimise costs.

[4] presents an AI-based Mango classification for excellence valuation by Nguyen Truong Thinh et al. For classifying mangoes, the suggested system uses a mixture of image processing and AI. Mango quality is determined by colour, dimensions, magnitude, outline, and fruit thickness. Images are captured using CCD cameras, and split layer processing is used to measure mass and volume, which aids in determining mango maturity and sweetness. The system's primary flaw is the usage of CCD cameras to seizure imageries.

[5] offers an overview of various methodologies suited for fruit and vegetable excellence scrutiny by Anuja Bhargava and Atul Bansal. Various approaches for picture gaining, pre-processing, division, feature mining, and cataloguing are discussed in this study. The procedures for acquiring ultrasonic, infrared, and tomographic images are summarised under image acquisition. Homochromatic image preprocessing approaches have been debated for picture element preprocessing in the Color Space Transformation (CST), hue, saturation, and intensity (HSI) colour space. Some segmentation strategies have been discussed as well. RGB based colour mining with deep learning approaches for classification is not mentioned among the strategies discussed.

Recently, Dr. M. Renuka Devi and Archana Tamizharasan proposed ailment prediction in Mango tilling pre- and post-harvest period in [6]. Using a Hybrid Multiclass Neural SVM classifier and an RMBF feature mining approach, this method detects 7 diseases found in Mango fruit as well as the level of infection found in the fruit. Classifiers such as Fuzzy logic,KNN, and Naive Bayes are used in the evaluation. This system has a forecast precision of 98.4 percent, however one of the disadvantages is that RMBF feature extraction takes up a lot of computational storing space. Santi Kumari Behera proposes a algebraic feature and SVM-based instinctive cataloging of mangoes.

In [7], Yuanshen Zhao et al. created a simple machine vision system for detecting tomatoes in a green house. The uncovering method consists of two steps: Step 1 collecting Harr-like features from grayscale photos and (2) categorising them using the Ada-Boost classifier. Step 2 A colour analysis approach based on Average Pixel Value (APV) is utilised to remove incorrect undesirable classifications in the outcome. In a real-world scenario, the proposed method has a promising outcome of more than 96 percent for detecting ripped tomatoes. The probability of false negatives is around 10%, and only 3.5 percent of tomatoes are identified. The system's main flaw is that it use Harr-style feature extraction techniques.

Fatma M. A. Mazen recently developed an artificial neural network-based banana ripeness classification in [8]. An intuitive computer vision system is presented to detect banana ripening phases. To categorise banana fruit ripening stage, an ANN-based framework is utilised to extract textural parameters such as colour, brown spot development, and tamura statistics from a four-class handmade database. SVM, Naive Bayes, KNN, and decision tree analysis classifiers are compared to the proposed approach. The system achieves a result of 97.75 percent. The computational cost of ANN is one of its drawbacks.

In paper [9], Manjunath Jogin et al. emphasis on feature mining using CNN and deep learning. Current picture cataloguing processes are faster and more accurate than ever before. When properly trained, deep learning models can distinguish many phases of visual representation. Convolutional neural networks can absorb rudimentary forms in the 1st layer and picture topographies in subsequent layers, resulting in more exact image classification. Hubel and Wiesel's organised portrayal of neurones in 1962, as well as their investigation of branches of pictorial mantle in cats, inspired CNN's hypothesis.It was a significant breakthrough in the realm of computer vision, and it aided in the understanding of how the visual pallium functions in humans and animals. In this study, visual features are mined using CNN and the deep learning idea.

The current methods listed above focus on computer vision-based techniques used to categorize diverse fruits for excellence valuation, defect identification, ripeness measurement, and illness detection, according to the results of the survey. The majority of the methods discussed are standard classifiers such as SVM, Ada-boost K means, decision trees, and Naive bayes, which are commonly used for fruits other than mangoes. Despite the fact that mangoes are consumed all over the world, chemically ripened mangoes cause a slew of health problems. To overcome this problem, potential machine learning frameworks for mango fruit consumption analysis are required.

## 3. PROPOSED MACHINE LEARNING BASED PREDICTION MODEL FOR CONSUMABLE MANGO

Figure 1 shows the proposed consumability prediction model. There are 2 stages in the model. In stage 1 image acquisition and preprocessing takes place followed by feature extraction using MPEG-7. In order to acquire the images around 55 - 60 Mangoes belonging to for diverse types namely Badam, Neelam, Mallika and Alphonso are considered. These fruits are divided into 2 sets. One set of mangoes is left to ripen certainly, whereas the other is chemically ripened with a ripening agent (calcium carbide). The Mango photos are taken with a Canon power shot digital camera after the ripening phase is completed.

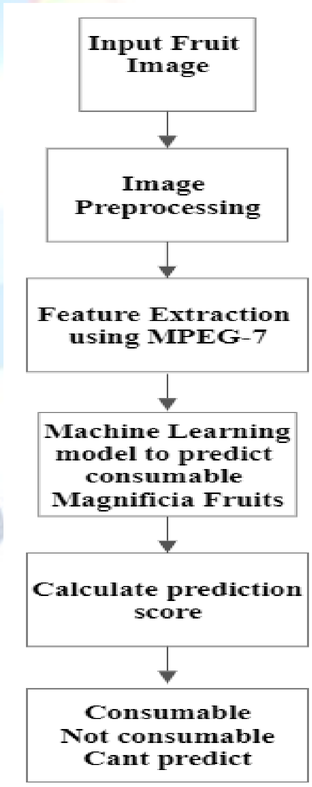


Figure 1: Proposed Model

Acquired images belonging to both the categories are to be preprocessed. In order to excerpt the parts of curiosity, the contouring approach is employed to crop the image. From the top left corner of the contour, the smallest horizontal and vertical coordinates are defined, which are then indexed, and these coordinates are clipped. As a result, the cropped image is subjected to additional processing in order to extract features. Figure 2 shows a contour cropped and masked image after preprocessing. The image is ready to extract features.

Figure 2: Contour cropping and masking sample output.

When it comes to recognising chemically ripened fruit, feature extraction is crucial. With the assistance of MPEG – 7 colour dominant characteristics [10], dominant colour values with respect to RGB are retrieved. For the picture element frequency of repetition in the original image, a Dominant Frequency Image (DFI) is created. In most cases, DFI structure is required to determine dominant colour. A 3D gamut of the image is created as described in [11] to determine colour frequencies for RGB values. DFI is abstracted to acquire accurate dominating colours. As a result, a better approach for DFI edifice is adopted. Table I lists the primary prominent colours' min, mean, max, average, and standard deviations for Naturally Ripened Mangoes (NRM) and Chemically Ripened Mangoes (CRM).

NRM fruits were found to be in the 140-205 range, whereas ARM fruits were in the 210-245 range. The values of secondary and tertiary dominant colours differ significantly. To move on to stage 2, these mined features are saved in a CSV data file and labelled appropriately.

Table I. The projected dominant colour mining algorithm yielded primary dominant colour values.

| | Prime Leading Colour | | | | |
|---|---|---|---|---|---|
| | Min | Mean | Max | AVG | SD |
| NRM | 66 | 183 | 248 | 0.24 | 0.06 |
| ARM | 123 | 222 | 255 | 0.18 | 0.04 |

In stage 2 Consumability analysis is carried out by using 4 machine learning algorithms namely Ada Boost, SVM, Decision tree and Random Forest. Each algorithm produces a threshold value which is further used to predict the consumability of the fruit. Consumability of the fruit is predicted by the percentage of accuracy obtained. If the prediction score is less than 50% then unpredictable, if >50% and < 80% then not fit to consume, >80% then consumable.

## 4. RESULTS AND DISCUSSIONS

**Experimental Setup**

A household dataset of Mango fruit photos is created for both training and testing tenacities to assess the performance of the projected Mango fruit consumability model. To create a dataset, approximately 1300 fruit photos are created. Figure 3 shows a few examples of mango pictures taken from various perspectives.

Figure 3: Sample snapshots of Mango images (vertical / horizontal axis, top / side / front views.)

A consumability analysis using machine learning algorithms namely Ada-Boost, SVM, Decision Tree and Random Forest is conducted for which the values extracted based on MPEG-7 colour dominant feature are used as input.

To evaluate the performance of proposed consumability model precision and confusion matrix metrics are considered. The performance results of different machine learning techniques are explained below. In fig 4 confusion matrix is constructed for prediction of fruit using AdaBoost. Here, confusion matrix specifies false positive, false negative, true positive and true negative which are indicated in bottom left, top left, bottom right and top right boxes of the confusion matrix respectively.

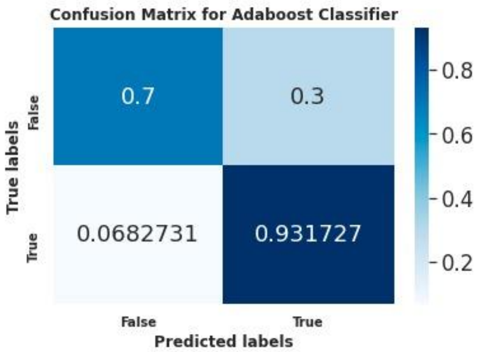


Figure. 4: Confusion matrix for Ada-boost

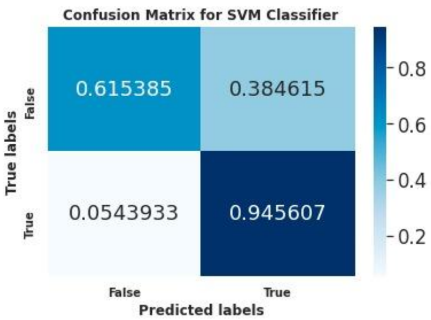


Figure. 5: Confusion matrix for SVM

For the same experimental dataset a detailed analysis of the individual machine learning algorithms namely SVM, Decision tree and Random forest are conducted in order to tabulate the entire prediction model. Confusion matrix for SVM, Decision Tree and Random Forest are shown in fig 5, 6 and 7 respectively. From table III one can make out that AdaBoost performance is slightly better compared to the other techniques because of the better decision-making capability of AdaBoost by focusing on areaswhere it predicts incorrectly.

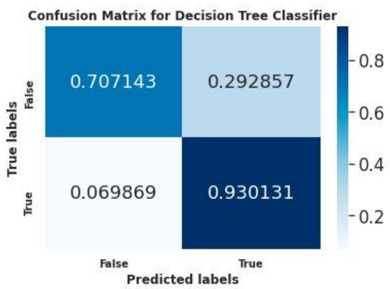


Figure. 6: Confusion matrix forDecision Tree

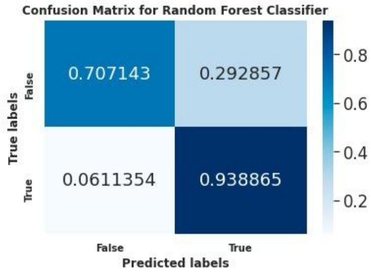


Figure. 7: Confusion matrix for Random Forest

Table III: Comparative analysis

| Sl. No. | Algorithm | Precision |
|---|---|---|
| 1 | Ada-Boost | 86.44%. |
| 2 | SVM | 83.46%. |
| 3 | Decision Tree | 85.55%. |
| 4 | Random Forest | 85.09%. |

To compare the performance of the proposed framework an existing technique described by Angelyn [13] which uses basic RGB features. A comparative table for the precision score of proposed model and Angelyn's model are tabulated in table IV. It is seen that the results of proposed methods are much promising when compared to Angelyn's method due to a collective combination of machine learning algorithms which gives a higher result.

Table IV : Comparison of scores

| | |
|---|---|
| Proposed Model Score | 86.44% |
| Angelyn's Model Score | 77.56% |

Based on the given fruit precision score, classification of the consumability of the fruit is done in three categories. If the prediction score of a fruit is greater than 80% it is good to consume. If the prediction score is less than 50% unpredictable. Prediction score greater than 50% and less than 80% bad to consume. Generally, effectiveness is tested based on time complexity or space complexity. Hence, proposed methods time complexity for consumability analysis for a given mango fruit image is 200sec.

## 5. FUTURE SCOPE AND CONCLUSION

This article discusses a machine learning-based model for predicting the consumability of chemically ripened mango fruit. As explained in section 3, the suggested model makes use of colour, texture, and shape features. The confusion matrix, prediction score metric, and time complexity of the model are used to evaluate the model's performance. A comparison of the existing system is also carried out. Consumability classification of the given fruit image is also done based on the model's prediction score. Using deep learning or transfer learning approaches, more emphasis can be placed on lowering time complexity for mango fruit classification.

## Conflict of interest statement

Authors declare that they do not have any conflict of interest.

## REFERENCES


[1] https://economictimes.indiatimes.com/news/economy/agriculture/india-needs-to-focus-on-quality-of-mangoes-to-expand-to-new-markets/articleshow/11758370.cms?from=mdr,

[2] https://timesofindia.indiatimes.com/city/chennai/78-of-consumers-unaware-of-how-to-identify-artificially-ripened-mangoes-survey/articleshow/69348156.cms

[3] Anurekha, D., Sankaran, R.A. Efficient classification and grading of MANGOES with GANFIS for improved performance. Multimed Tools Appl 79, 4169–4184 (2020).

[4] N. T. Thinh, N. Duc Thong, H. T. Cong and N. T. Thanh Phong, "MangoClassification System Based on Machine Vision and Artificial Intelligence," 2019 7th International Conference on Control, Mechatronics and Automation (ICCMA), 2019, pp. 475-482, doi: 10.1109/ICCMA46720.2019.8988603.

[5] Anuja Bhargava, Atul Bansal, Fruits and vegetables quality evaluation using computer vision: A review, Journal of King Saud University - Computer and Information Sciences, Volume 33, Issue 3, 2021, Pages 243-257, ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2018.06.002.

[6] Dr. M. Renuka Devi, Archana Tamizharasan. PREDICTION AND CLASSIFICATION OF DISEASE IN MANGOFRUIT USING HYBRID MULTICLASS NEURAL SVM. JCR. 2020; 7(10): 1770-1778. doi:10.31838/jcr.07.10.318

[7] Yuanshen Zhao, Liang Gong, Bin Zhou, Yixiang Huang, Chengliang Liu, Detecting tomatoes in greenhouse scenes by combining AdaBoost classifier and colour analysis, Biosystems Engineering, Volume 148, 2016, Pages 127-137, ISSN 1537-5110,

[8] Mazen, F.M.A., Nashat, A.A. Ripeness Classification of Bananas Using an Artificial Neural Network. Arab J Sci Eng 44, 6901–6910 (2019). https://doi.org/10.1007/s13369-018-03695-5

[9] Jogin, Manjunath & Mohana, Mohana &Madhulika, M &Divya, G & Meghana, R & Apoorva, S. (2018). Feature Extraction using Convolution Neural Networks (CNN) and Deep Learning. 2319-2323. 10.1109/RTEICT42901.2018.9012507.

[10] V. Laxmi, R. Roopalakshmi, A Novel Frame Work for Detection of Chemically Ripened MangoFruits Using Dominant Colour Descriptors, DOI10.3233/APC200114, Advances in Parallel Computing, volume 37, Intelligent Systems and Computer Technology (2020), pages 18 – 27, ISSN 0927-5452

[11] Mark Hayworth (2019). Color Frequency Image. https://www.mathworks.com/matlabcentral/fileexchange/28164-color-frequency-image. MATLAB Central File Exchange.

[12] V. Laxmi and R. Roopalakshmi, Detection of Chemically Matured MangoFruits Using Laplacian Descriptors and Scale Determinants (November 21, 2020). Proceedings of the 2nd International Conference on IoT, Social, Mobile, Analytics & Cloud in Computational Vision & Bio-Engineering (ISMAC)

[13] Angelyn T. Lacap, Emma Ruth V. Bayogan, Leizel B. Secretaria, Christine Diana S. Lubaton, and Daryl C. Joyce, Responses of 'Carabao' Mango to Various Ripening Agents, Philippine Journal of Science, 148 (3): 513-523, September 2019