Fruit Quality Detection Using Random Forest Algorithm

Ms. R.M Danushree[1], Ms. M. Aishwaryashree[2]
[1]Assist Professor, Dept of CSE
[2]Dept of AI&DS
[1, 2] P.A College of Engineering and Technology,Pollachi.

Abstract- *The expense of the fruit is thought to be nearly a third of what is lost due to fruit rotting, which has major economic repercussions. Additionally, fruit sales will suffer because consumers believe that spoiled fruits are harmful to their health. Recently, a technology-based method for evaluating fruit quality was launched. This technology is increasingly being used in the produce business and in agriculture. Here, we suggested using the random forest method to identify fruit grade. This system's primary objective is to supplant human inspection methods. This saves time while also accelerating the process and improving precision and productivity. Before beginning the random forest process, the input picture was first divided into parts, unwanted regions and regions with flaws were segregated, and then some of the characteristics were extracted. The suggested idea could be a wonderful method for retailers to generate high-quality sales of food products, particularly perishable fruits. For managing food products in markets, shops, and other locations, this is a great option.*

I. INTRODUCTION

Modern agriculture is becoming increasingly automated, making human system management a time- and resource-intensive task. Fruit must first be tested for quality before being used to create gourmet goods. This research explains how fruit grade and quality evaluations frequently use machine learning. In the past, selecting high-quality fruits and veggies for industrial manufacturing required more work. High-quality fruit can now be identified using a variety of automatic technologies that have recently been created.

Automation is important in daily life. In India, more than half of the population lives on agriculture. Their main source of income is agriculture. Fresh fruit exports from India are increasing day by day. People are very conscious of their health. They prefer only good quality fresh fruit. Today, several research studies focused on agriculture offer different methods for identification and classification of fruits and vegetables.

After gathering, produce needs to be processed for selling. This can be carried out on a farm or at the level of a network of supermarkets or a retailer. Regardless of the location, the four most crucial steps for the fresh market are the removal of unmarketable material, separating by maturity and/or size, grading, and packaging. We require a scoring system to expedite the procedure and meet the demands. Products that have been evaluated command greater prices on the market. Grading helps buyers and producers build greater confidence with one another.

Machine learning techniques are used by the majority of current agricultural tools. Unsupervised learning and supervised learning are the two main categories for these machine learning methods. In supervised learning, labelled datasets are used to teach algorithms that correctly categorize data or anticipate results. Examples of supervised learning include the K-Nearest Neighbor (KNN) method, Naive Bayes (NB), Discriminant Analysis (DA), Vector Machines (SVM), and Discriminant Analysis (DA). In unsupervised learning, unlabeled samples are analyzed and clustered using some methods. Examples of autonomous learning algorithms include principal component analysis, hierarchical clustering and K-means clustering.

The proposed system uses Random Forest algorithm to detect the defective fruit from the dataset. This system would detect the defect of fruits apple and banana. Form the input image unwanted region will be removed and separate the defected region of the fruit, Some of the features like shape, colour, texture and size are extracted from the image and undergoes Random Forest technique to detect the defect of the fruit. This system can give better result by identifying the fruit is rotten or not.Human grading and sorting is feasible, but it is time-consuming, labor-intensive, mistake-prone, and exhausting, so a complex fruit quality detecting system is required. Fruit quality detection system is an important step for consumers since it effects both the quality of the fruit and the international market.

II. LITERATURE SURVEY

Fruit classification and identification are considered to be challenging activities that still provide some challenges. To create the optimum fruit classification and recognition system, challenges must be overcome. This section of the article displays the efforts of academics who have tried to employ various strategies to get beyond some of the obstacles they have encountered. The research that follows enumerates previous efforts. Each release is also given a critical evaluation based on the question it seeks to address, the data it uses, the methodology it employs, and potential follow-up research.

[1]Aaron Don M used to determine a product's quality with its appearance, form, colour, and texture, all of which are subject to human error. The poor quality of crops, fruits, and vegetables is one consequence of insufficient technological development. Fruit quality and ripeness levels must be consistently determined, which can be challenging and tiresome for humans to do when it becomes repetitive labour. The seeks to show several methodologies and approaches on how machine learning and machine vision algorithms can be used to make ripe fruit detection and classification simpler and more practical. The also offers systems that may be used for pre- and post-harvest analysis. Some of the most popular and effective techniques, including deep learning, image illumination, and faster-CNN, can provide greater accuracy for detection.

[2] An Adaptive Neuro-Heuristic Hybrid Model for Detecting Fruit Skin Problems was put forth by Marcin Wozniak. The project employs the Adaptive Neural Network architecture to categories segments and find flaws in fruits like Apple, Orange, and Banana. The input image is segmented, and information processing is used to calculate the primary colors red, green, and blue. The developed neural network topology that adjusts to processed information cooperates with the heuristic technique used to find regions of defect. Heuristic approaches can effectively look for unique features in input images, provided that both the input and the fitness condition were chosen with care. As a research tool, three different fruit types apples, bananas, and oranges had peel damage. The established hybrid technique, its components, and the steps of processing for which we have assembled an adaptive neural network and a heuristic detector which makes the new suggested solution.

[3] Fouzia Risdin, explains to detect fruits using deep convolutional neural networks. The system uses a convolutional neural network (CNN) to recognise fruit in picture input. Fruit image identification is typically quite challenging due to the large variety of fruit kinds. The accuracy of CNN was higher than that of traditional support-vector machine-based techniques using handcrafted features. CNN also shown noticeably more accuracy for fruit picture detection than a traditional approach. Additionally, the major benefit where the method can be applied for new fruits much more quickly. A new dataset made up of 2403 data points from 4 different fruit classes is used to retrain the algorithm to detect 4 different fruits. The study's results were off the mark despite being based on real-life circumstances. This project aims to expand the system to include additional objects and practical uses.

[4] Sharad Hasan worked with a machine learning and computer vision system for automatic apple disease diagnosis and classification based on leaf symptoms. Feature extraction, sick area segmentation, and classification are the three stages that make up the technique. The infected region of the leaf was divided up using the L*a*b* space-based colour segmentation technique. Using the average colour markers in a*b* space and the closest neighbour method, each pixel has been split into healthy, ill, and background areas. The suggested DWT feature and L*a*b* space-based colour histogram features are two distinct types of features that have been extracted. Horizontal feature fusing is used to create the finished feature vector. The feature vectors have been categorised using a number of classifiers, with Random Forest serving as the base classifier. The Cedar Apple Rust, Black Rot, and Apple Scab diseases are depicted in pictures from the Plant Village collection, which is used in the exercise. A cutting-edge technique that increased accuracy for detecting and treating apple leaf disease is the combination of recommended DWT and colour histogram features.

[5] Behera S.K proposed this research is to provide a brand-new, non-destructive classification system for papaya fruit maturity stage. The machine learning method uses three classifiers with various kernel functions and three sets of features. Local Binary Pattern, Histogram of Oriented Gradients, Gray Level Cooccurrence Matrix , K-Nearest Neighbor, Support Vector Machine, and Naive Bayes are some of the features and classifiers used in machine learning methodologies. To identify the Papaya's various phases of development, the transfer learning approach uses seven pretrained models, including Residual Neural Network, Visual Geometry Group, GoogleNet, and AlexNet. According to their aesthetic qualities, papaya fruit images are divided into three categories: immature, half-ripe, and mature. Utilizing MATLAB 2019a, these experimental tests were carried out. The primary flaw in this project was the lack of a prototype's creation and integration into the fruit sector to facilitate quick sorting. Visual Geometry Group with 19 layers, which relies

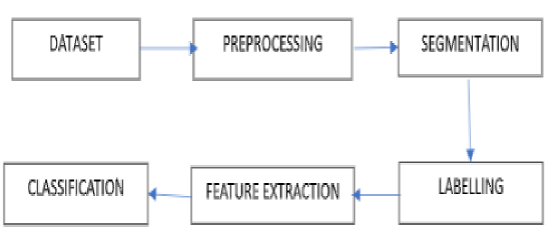
on a transfer learning approach, was able to accurately predict the results of the experiment.

[6] Linjuan Ma et al, developed a fruit detector based on the fruits-360 dataset, which contains 38409 images of fruits distributed across 60 labels and 75 classes of the fruits dataset and uses the state-of-the-art detection framework, Faster R-CNN. In agriculture, fruit detection is crucial since it can increase labour productivity and market price rivalry. There would have been more framework and tools employed. They would have enhanced our deep network's functionality to find more fruits and cater to our day-to-day needs.

III. METHODOLOGY

3.1MACHINEVISION TECHNOLOGY

It offers image-based automatic inspection and analysis for uses in robot navigation, process control, and autonomous inspection. The steps in the process involve outlining the requirements and project in detail before coming up with a solution. The procedure begins during a run-time with imaging, then automatically analyses the image and extracts the necessary information.A very wide range of applications have been examined, monitored, and controlled using machine vision. It involves using an optical, non-contact sensing device to automatically receive and interrupt the picture from a real sense in order to gather data, manage machinery, or process images. Machine vision techniques are now being used in a wide range of industries, including agriculture, food production, medical diagnostics, remote sensing, autonomous vehicles, automatic manufacturing and surveillance, and the inspection and grading of fruits and vegetables.

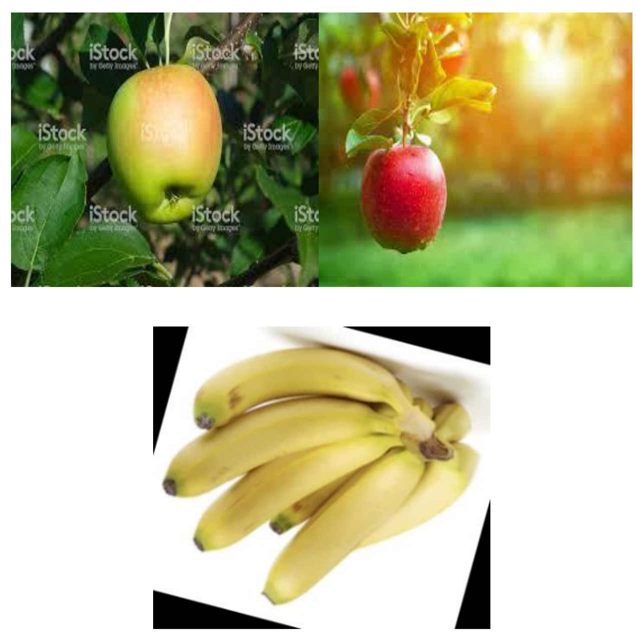


Steps Involved In Machine Vision Technology

3.2 DATASET

The fruits 360 dataset, which is available on Kaggle for free, was the source of all the pictures used in training and testing. There are 10,000 various fruit pictures in the collection, divided into categories. Different classes each stand in for a particular berry variety. The two dishes chosen are banana and apple. To prevent the research from extending too far, the data collection was constrained. All fruit types fall under the same division, which results in these limitations.This indicates that all types of apples are categorized as apples, and all other fruits are also categorized in this way. After the background was eliminated, all of the fruits were scaled to 100100 pixels of typical RGB pictures. The fruits-360 collection's 7500+ images were sorted with 1762 images. In which 870 fruits were affected and 892 were not affected.

Rotten Fruit Dataset Images



Good Fruit Dataset Images

3.3 IMAGE PREPROCESSING

There are numerous sizes and shapes for images. Additionally, they come from a range of locations.Any image data that needs pre-processing must take all of these variations into consideration. The most common decoding format is RGB, which is used to capture the majority of "natural pictures". Data pre-processing also includes a preliminary stage of uniformly sizing the images. Auto-resizing is used to automatically resize each image in the dataset to the same scale before training.Additionally, model pre-processing might make model training and inference go more quickly. In cases where the input images are quite big, shrinking the size of the images will considerably reduce training time without noticeably affecting model performance.

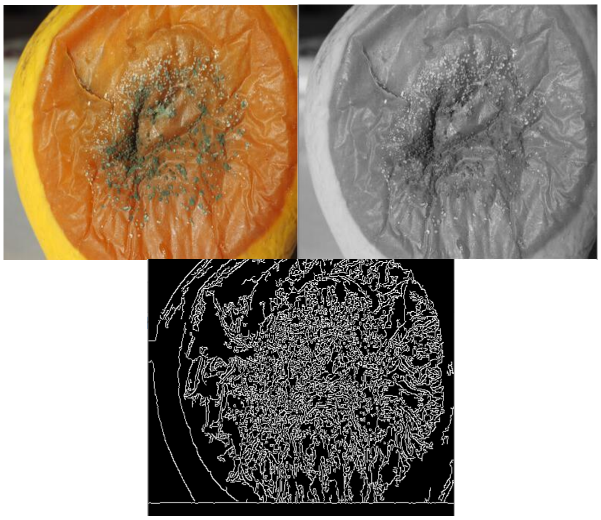


Image while pre-processing

3.4IMAGE SEGMENTATION

The act of segmenting an image into various sections is referred to as segmentation. By simplifying andaltering an image's depiction, picture segmentation aims to make it easier to comprehend and investigate.The characteristics of similarity and discontinuity are used to describe picture segmentation methods. Boundary-based strategies rely on gaps, whereas region-based strategies do so based on resemblance.

3.5 EXTRACTION OF FEATURES

The procedure of classifying fruits continues with feature extraction after segmentation. The fruit's size, form, colour, and texture are its most noticeable external features. A feature descriptor is a representation of a picture or a part of an image that highlights key details while omitting the rest. The two main applications are object identification and picture recognition.

The neural network's weights are optimised using the gradient descent algorithm and the backpropagation method, which may be stored as model file. The weights are used by the network to anticipate the output during the forward pass. The gradients are utilised to update the weight matrices as the "cost" or error value, or the difference between the actual and anticipated output, is back-propagated through the network.

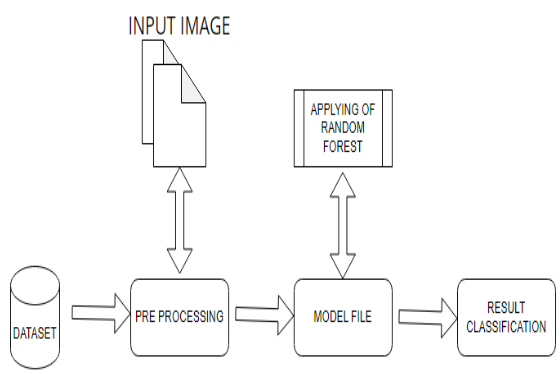
3.6 KNOWLEDGE-BASED COMPARISON AND DECISION MAKING

The extracted qualities from the image are contrasted with the predetermined categorization and ordering rules. Fruits are classified based on comparisons of their traits and the features that were recovered. Knowledge-based comparison and decision-making were conducted using Random Forest techniques. With this step, the categorization mechanism is finished.

3.6 WORKING OF RANDOM FOREST ALGORITHM

Random Forest is a well-known algorithm that is a part of the supervised learning technique, and it has been employed in this work under the heading of machine learning. It can be applied to classification and regression issues in machine learning. A random forest is a type of classifier that takes the average of a group of decision trees over various subsets of a given dataset to increase the dataset's predictive accuracy. Based on the majority vote of the projections, random forests use predictions from each tree to forecast the eventual result. It were chosen since it is a quick and effective way to find the flaw. Random Forest needs less time to train than other machine learning methods. It exhibits improved run and output forecast accuracy.

The Random Forest model is predicated on the notion that numerous uncorrelated models (the various decision trees) perform noticeably better when combined than when used independently. When utilising Random Forest to categorise, each tree offers a classification or a "vote". The forest chooses the category that receives the most "votes."



Proposed System

To aggregate the predictions of various decision trees into a single model, the main justification for using a random forest rather than a decision tree. According to this reasoning, a single model that is composed of numerous mediocre models will nevertheless be superior to a single excellent model. Given the general effectiveness of random forests, this is true. Because of this, random forests are less likely to overfit.More quickly than other machine learning techniques, Random Forest requires less training time. Even with bigger databases, it still performs well, and its output prediction accuracy has increased. Significant quantities of missing data are still kept with accuracy.

Start by choosing random examples from a pre-existing collection.A decision tree will then be built by this method for each data. The forecast outcome from each decision tree will then be obtained.Voting will be done in this stage for each anticipated outcome. Finally, choose the forecast result that received the most votes as the ultimate prediction result.

IV. RESULT AND DISCUSSION

The quality of fruits has been determined using a variety of quality detection techniques, including k-mean clustering, the Edge Detection algorithm, artificial neural networks, convolutional neural networks, and support vector machines. Their effectiveness and performances have been compared.Using the Random forest algorithm we have discovered a novel method for determining the quality of fruits by comparing it to the currently suggested model. This method achieves high accuracy while being highly efficient.

This project suggests a novel method for determining whether or not fruits are in good condition. The sampled image effectively extracts the features. The parameters used to extract the features, such as colour, texture, size, and shapes.

```
The actual data is:
The model is 95.62192816635161% accurate
[1 0 0 1 1 1 0 1 1 1 1 0 1 0 1 0 1 0 1 0 0 0 0 0 1 1 0 0 1 1 0 0 1 1 0 1 1
 1 1 0 1 1 0 0 1 0 0 0 0 1 1 0 0 0 0 0 0 1 0 0 0 1 0 1 1 0 0 0 0 1 1 0 0 1
 0 1 1 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 1 0 1 0 0 1 1 0 0 1 1 1
 0 0 1 1 1 1 0 1 1 0 1 1 0 1 0 1 1 0 1 0 0 1 0 0 1 0 1 1 0 0 1 1 1 0 0 1 1
 1 1 0 0 1 0 1 0 1 0 0 1 1 0 1 1 1 1 0 0 1 1 1 1 0 0 1 0 0 1 1 0 1 0 0 1 0
 1 1 0 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 1 1 0 0
 0 1 0 0 0 1 0 1 1 0 1 0 0 1 1 0 0 1 1 0 0 1 0 0 1 0 1 0 1 1 1 0 0 1 1 1 1
 1 0 1 0 0 1 1 0 0 0 1 1 1 0 0 1 1 1 1 1 1 0 1 0 0 1 0 0 0 0 0 0 0 1 1 0 0
 0 1 1 0 0 0 0 0 1 1 1 0 0 1 0 1 0 1 1 0 1 1 1 0 1 0 0 1 1 0 0 0 0 1 1 1 0
 1 0 0 0 0 0 1 0 1 1 0 0 0 0 1 1 1 1 0 1 1 1 1 1 1 0 0 1 0 1 1 0 0 1 1 1 1
 1 1 0 0 0 1 0 1 1 1 1 1 0 0 1 0 0 1 1 1 0 0 0 1 1 0 1 0 0 0 0 0 1 0 1 0 0
 0 1 0 1 0 0 0 0 0 1 1 0 1 1 0 0 0 0 0 1 0 1 0 0 1 0 0 1 0 1 1 1 1 1 0 0 1
 0 1 1 0 1 0 0 1 1 0 1 0 0 0 1 1 1 0 0 0 0 1 1 1 0 1 0 1 0 1 0 0 0 0 1 1 0
 0 0 1 1 0 1 0 1 0 0 0 1 0 1 0 1 1 0 1 0 1 1 1 0 0 0 1 0 0 0 0 0 0 1 0 0 0
 1 0 0 0 1 1 1 0 1 1 1]

Classification Report

              precision    recall  f1-score   support

           0       1.00      0.99      1.00       261
           1       0.99      1.00      1.00       268

    accuracy                           1.00       529
   macro avg       1.00      1.00      1.00       529
weighted avg       1.00      1.00      1.00       529
```

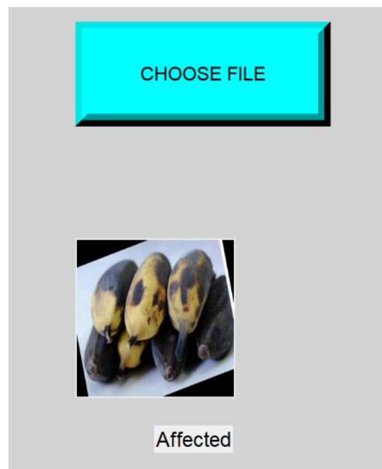
Training The Dataset

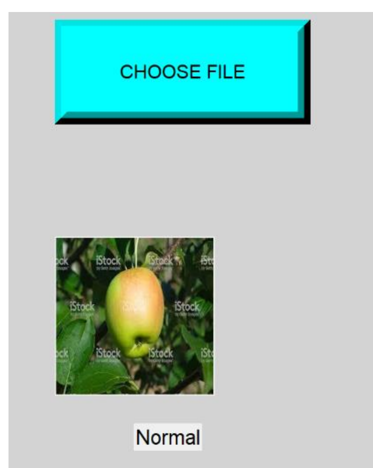
The random forest method is employed to assess the level of quality. The quality is assessed using the features of the fruits that have been extracted and the values that were given to the random forest algorithm during training. The suggested method effectively assesses fruit quality. The outcome will be either good or bad.

Juice manufacturing facilities, food safety businesses, fruit and vegetable farms, fruit and vegetable packaging businesses, etc. can all use this type of system.

V. CONCLUSION AND FUTURE WORK

The fruit quality detection is used to automate the work in such, when the image of the fruit is once scanned or uploaded the result will be shown weather the fruit is rotton or not. The results were obtained with an accuracy level of 95.6% using the random forest Algorithm and the sample Kaggle dataset for banana and apple.





```
Shell
Python 3.7.3 (C:\Users\ASUS\AppData\Local\Programs\
>>> %Run final.py
 0
 The predicted image is : Affected
 1
 The predicted image is : Normal
```

Future research should compare our Fruit quality detection to detect quality to that of other mechanical and automated methods, and new parameters or features may be added. The proposed model can be scaled up to detect the quality of multiple fruits of various types at once. It can detect the quality of a single fruit at a time. The accuracy level of fruit quality detection can be improved, and the time-consuming process can be shortened. The mobile application can be used to upgrade the smart fruit separator to the next level, where it is fully automated and controlled.

Fruits' exterior features, including their colour, size, shape and texture are crucial for grading and classification. With the advancement of technology, machine vision and the availability of affordable hardware and software are becoming more widespread. Due to software, automated machine vision systems have taken the place of manual fruit sorting and grading. Non-destructive automation may be used because it has the ability to produce more accurate, rapid, objective, and efficient results than manual work. Despite several obstacles that still need to be cleared, machine vision will be the non-destructive testing method of the future. In the future, we can focus on categorising the images of the locally grown produce. Also possible are machine learning methods for grading fruits and vegetables. Based on the aforementioned techniques, mobile applications for farmers and the general public to use for the identification, classification, and grading of horticulture products can be created.

REFERNCES

[1] Behera SK, Rath AK, Sethy PK. Maturity status classification of papaya fruits based on machine learning and transfer learning approach. Inf Process Agric 2020

[2] Sabzi S, Abbaspour-Gilandeh Y, Garcı Mateos G. A new approach for visual identification of orange varieties using neural networks and metaheuristic algorithms. Inf Process Agric 2018;5, pp:162–72

[3] Rojas-Aranda JL, Nunez-Varela JI, Cuevas-Tello JC, RangelRamirez G, Fruit Classification for retail stores using deep learning. In: Proc. proceedings of pattern recognition 12th mexican conference. Morelia, Mexico; 2020. pp: 3–13

[4] Aaron Don M. Africa, Anna Rovia V. Tabalan, Mharela Angela A. Tan, “Ripe Fruit Detection and Classification using Machine Learning”, International Journal of Emerging Trends in Engineering Research, Volume 8, May 2020

[5] Marcin Wozniak , Dawid Polap. Adaptive neuro-heuristic hybrid model for fruit peel defects detection.Journal of Neural Network,Volume 9, 2019, pp 16-33

[6] Fouzia Risdin, Pronab Kumar Mondal, Kazi Mahmudul Hassan,Convolutional Neural Networks (CNN) for Detecting Fruit Information Using Machine Learning Techniques,IOSR Journal of Computer Engineering,Volume 22, Issue 2, Ser. I (Mar - Apr 2020), pp: 01-13

[7] Nashwa El-Bendary , Esraa El Hariri , Aboul Ella Hassanien , Amr Badr, Using machine learning techniques for evaluating tomato ripeness, Expert Systems with Applications 42(2018),pp: 210-215

[8] Pallvi U, Patil, G.C Wakchaure. Grading and sorting technique of dragon fruits using machine learning algorithms:journal of agriculture and food research,volume 4, June 2021

[9] Sharad Hasan, Sarwar Jahan, Md. Imdadul Islam. Disease detection of apple leaf with combination of color segmentation and modified DWT, Journal of King Saud University.Volume34,Jun 2022

[10] R. S. Latha et al., "Automatic Fruit Detection System using Multilayer Deep Convolution Neural Network," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp: 1-5

[11] B. Rodrigues, R. Kansara, S. Singh, D. Save and S. Parihar, "Ripe-Unripe: Machine Learning based Ripeness Classification," 2021 5th International Conference on Intelligent Computing and controlling system, pp: 15-20

[12] Vishnu H S, Sindhushree B, Punith A, Aishwarya K, Praveen G, 2020, Fruit Recognition System for Calorie Management, International Journal Of Engineering Research & Technology (IJERT) IETE , Volume 8 - Issue 11, 2020

[13] Ghazal, Sumaira& Qureshi, Waqar & Khan, Umar & Iqbal, Javaid & Rashid, Nasir &Tiwana, MI, “Analysis of visual features and classifiers for Fruit classification problem”, Computers and Electronics in Agriculture,2021

[14] H. M. Zawbaa, M. Hazman, M. Abbass and A. E. Hassanien, "Automatic fruit classification using random forest algorithm”, 14th International Conference on Hybrid Intelligent Systems, 2014, pp: 164-168

[15] V. B. Kasyap and N. Jayapandian, "Food Calorie Estimation using Convolutional Neural Network”, 3rd

International Conference on Signal Processing and Communication (ICPSC), 2021, pp: 666-670

[16] A. Hutagalung, H. Nugroho, A. Suheryadi and P. E. Yunanto, "Detection and Counting of Mango Fruits in Occluded Condition Using Image Analysis," 2017 5th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME), 2017, pp: 190-195

[17] Raja Sekar L, Ambika N, Divya V, “Fruit Classification System Using Computer Vision”, International Journal of Trend in Research and Development, ISSN:2394-9333, pp:22-26, 2018

[18] Bhanu pratap, Navneeth agarwal, sunil joshi, “Development of ANN based efficient fruit recognition technique”, The Global Journal of Computer Science (IJTRO), pp:18-28

[19] Xiaochun Mai, Hong Zhang and Max Q.-H. MengJ, “Faster R-CNN with Classifier Fusion for Small Fruit Detection”, IEEE International Conference on Robotics and Automation (ICRA) May 2018,pp: 24-26

[20] Hasan Basri, Iwan Syarif, Sritrustra Sukaridhoto , “Faster R-CNN Implementation Method for MultiFruit Detection Using Tensorflow Platform”, Department of Information and Computer Engineering Graduate Program Of Engineering Technology (IEEE), 2020

[21] Linjuan Ma, Fuquan Zhang, Lin Xu, “Fruit Detection Using Faster R-CNN Based On Deep Network“,Computer Science and Technology (IEEE), 2020.

[22] Polder, G., Van der Heijden, G., & Young, I, “Spectral image analysis for measuring ripeness of tomatoes”, Transactions-American Society of Agricultural Engineers(Research gate), 45(4), pp: 1155–1162

[23] Fadilah, N., Mohamad-Saleh, “Color feature extraction of oil palm fresh fruit bunch image for ripeness classification”,13th International Conference on Applied Computer and Applied Computational Science 2014 (ACACOS’14) , pp: 51–55

[24] Dominique Albert-Wei, Egla Hajdini, Matthias Heinrich, Ahmad Osman, ” CNN for ripeness classfication of watermelon fruits based on acoustic testing”, 3rd International Symposium on Structural Health Monitoring and Nondestructive Testing ,Nov 2020, pp:24-26

[25] Zhang, L., & McCarthy, M. J. (2012). Measurement and evaluation of tomato maturity using magnetic resonance imaging. Journal of Postharvest Biology and Technology, 67, pp: 37–43