Fruit Ripeness Detection with Machine Learning using Raspberry Pi

Atal Tiwari[1], Anmol Sharma[2], Avinash Patil[3]
[1,2,3] *Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune*

***Abstract*- The term Machine Learning refers to the field of study that gives computer the ability to learn without being explicitly programmed. It is a boon to various fields that basically depends on the reliability of the product. When it comes to agriculture field, like quality of the fruit or we can say ripeness of the fruit, machine learning plays an important role in making it happen to identify the ripeness of the fruits based on the training datasets we fed. With the help of computer vision and digital image processing we can find the ripeness of a fruit.**

In this paper we are basically focusing on computer vision strategies used to recognize a fruit which rely on four basic features which characterize the object: intensity, color, shape and texture. The methodology bestowed is ready to acknowledge fruits in natural condition facing troublesome situations: shadows, bright areas, occlusions and overlapping fruits. Every technique uses coloured pictures of fruits from totally different positions as input file. In these techniques we have a tendency to set some threshold levels. By examination the input file image with these threshold levels we are able to realize the maturity level of fruits.

The fruit ripeness detection technique can play a vital role in the large scale industrial applications for detecting the quality of the fruits. Businesses like bigbasket, Grofers, Amazon Now, etc. which deals with the import and export of the fruits can adapt the technology in order to not only reduce the labour cost but also time. It is also useful for agriculture field.

Index terms-Machine Learning, Computer Vision, Clustering Algorithm, Fruit Ripeness, Digital Image Processing, Agriculture, etc

I. INTRODUCTION

Agriculture plays an important role and is the integral part of overall economic development of a nation. Since, it is known that the only way of living for two-third population in India depends on agriculture. With the rise in technology in automobile and electronics, it is observed that development in the field of agriculture is sluggish. Thus, it is necessary for us to innovate in the field of agriculture to improve the agriculture development.

Fruits market is subject of choice. Thus, it is important for the suppliers to label the good quality of fruits before selling. Nowadays, human experts rate the fruits based on its optical features. Notwithstanding with manual inspection which causes the inaccuracy, inconsistency, vagueness and inefficiency on grading the quality of fruits. So, we prefer to do it with automatic inspection with the help of trained model using machine learning algorithms based on predictive analysis. This will not only save time but also labors which will contribute to the economic development.

Machine Learning is an application of artificial intelligence (AI). Machine learning focuses on computer programs that can access data and use it learn for themselves. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly. Machine learning gives analysis of high quantities of data. While it generally delivers faster, more accurate results in order to identify profitable opportunities or dangerous risks, it may also require additional time and resources to train it properly. Since, it is important to note that accuracy of machine learning does not depend on the pieces of code but it depends upon the training datasets [1].

In past, human rely on their vision to differentiate between ripe and unripe fruits. But this method had high rate of illness, distraction and other factors during working hours [2]. This also may affect the

working speed of system. So, to decrease this failure rate human started to invent new methods.

These days, there are various methods to detect the ripeness of fruits. In some methods we apply chemicals on fruits and sometimes we use machines. It is quite obvious that chemicals may affect human health and cause severe so basically machines are used for this purpose. Machine use their visual-based classification system that provide reliability, high speed and repeatable operation. Hence, the production increases and reduces its dependency on manpower.

In computer vision system computer uses different method to analyze the given image of fruit. Previously, computer systems were not strong enough to large and real colors of images, so mostly gray scale images had been the main focus for researchers. But today, computer system has been developed enough to work on large and true color images [3].

II. LITERATURE REVIEW

To increase the efficiency of computer system different researchers perform various experiments on different fruits and vegetables to check their maturity levels. In 2004, F. Mendoza, et al converted the RGB image of bananas into CIELAB format and accuracy reaches to 98% [4]. In 2003 and 2008, author uses average values of RGB to evaluate maturity levels of peaches, apples and oranges [5]. In 2010, Zhi-yuan Wen, et all used machine vision to detect the maturity of citrus fruits. B. Ojeda-Magana, et al used different partitional clustering algorithms to detect ripeness level of bananas and tomatoes [1]. Fatma Susilawat mohamad and Azizah Abdul Manaf used histogram matching method to find ripeness of oil palm fruits. Scanlon proposed an approach to quantify colour of potato chips and Choi used Colour image analysis to detect tomatoes maturity rate [6]. In 2011, Chiunhsiun Lin, et al proposed a method to check the maturity rate of tomatoes [3].

III. SYSTEM OVERVIEW

The equipment used for this project is Raspberry Pi 3, Raspberry Pi Camera Module, Laptop, etc. Mainly, in our system, we will place a fruit in front of camera which will allow to click pictures from it. We will click a picture to extract maximum features from the clicked image. The clicked image will go under image processing for feature extraction, thus the extracted features will get compared with predefined datasets of images and will result accordingly. If an anomaly is detected in the featured image, then will not allow to pass the fruit further and will know us about the same.

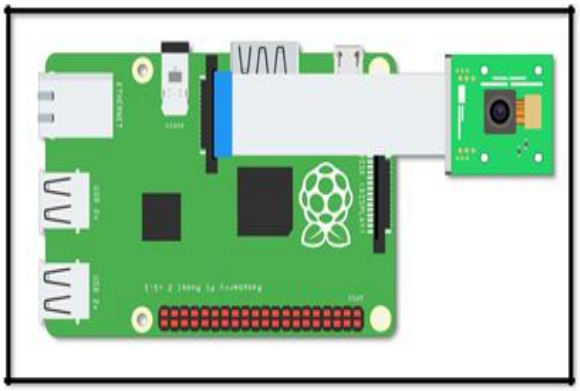

Fig 1. Raspberry Pi setup with camera module

III. (a) Implementation Plan

In implementation design, each component is sufficiently described to allow for its coding. This design provides an ongoing plan for assessing components of the program. It is an action plan to achieve program outcomes and program learning outcomes. The implementation plan breaks each strategy into identifiable steps, suggests the ways for the completion of each step individually and then combining them in order to have a fully developed Apple Classification system. This plan is developed during the design phase and updated during the development phase.

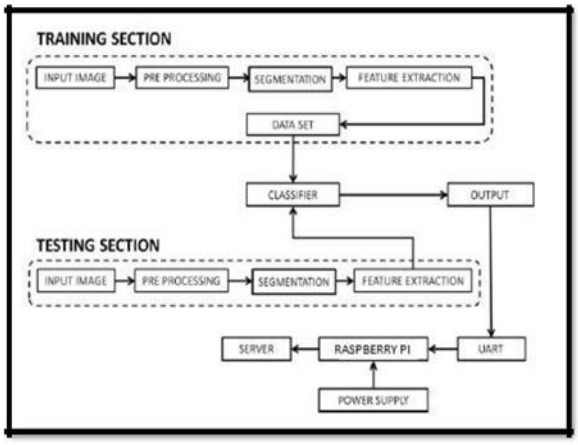

Fig 2. Methodology Implemented

III. (b) Data Collection

The data being collected should not be biased and collected in appropriate quantity that is used to predict the exactness of the system. The more data an algorithm has, the more accurate it becomes. In the proposed system, if the training data is limited, it may not be able to support the model complexity needed for the problem. Improper data collection will reduce the ability to build predictive, accurate machine learning models.

For Apple Classification, 200 images of each maturity stage are captured that makes total of 400 images. From these existing images, more number of images is generated in order to train the models well.

III. (c) Image Augmentation

Some additional images are constructed from the existing raw data to generate a large training dataset as more number of training set will give more accuracy. The new images are generated to provide better information that is not clearly defined by the sample dataset. This process is done by using the Pandas Python package. The Apple Classification data need to be loaded from the database into the Panda data frame and then it is processed further. The Pandas library in Python gives a huge set of data structures and analysis tools for data handling in Python programming.

From the given set of images more images are generated by using the Image Generator method of Keras which is a deep learning package for Python. Different techniques that are used to generate images are

- By rotation
- By flipping the image horizontally
- By zooming some random pixels of the image

III. (d) Feature Extraction

In Tomato Classification system, the most important characteristic to evaluate apple ripeness is its surface color. This system uses High, Variance and Average values of 100th and 500th pixel for every RGB channel in the image. For this purpose, Scikit-learn library is used for feature extraction in order to create a pickle file for each stage of apple. Fifteen color features, five for each channel (R, G and B) (100H, 100V, 500H, 500V and avg) are computed. Then pickle file is created for each maturity stage and combined into a single file for training and testing models.

III. (e) Training and Classification

After extracting relevant features from the dataset, now there is a need to train a model. To compare the performance of different models, seven classifiers have been used for classification of different ripeness stages of apple. The process of training a model contains provision of machine learning algorithm with training dataset to learn from. The data which is to be trained must have correct answer, which is a target attribute. The learning algorithm compares the input data to this target attribute and outputs the machine learning model that grabs these patterns in it. Then these models are used to make predictions on future data.

Once the training has been completed, next step is to understand, recognize and differentiate various stages of apple based on the training set of data. This process is called classification and the algorithm which implements classification is known as classifier. Classification generally refers to classifying images into distinct categories of apple ripeness in Apple Classification system.

III. (f) Testing and Validation

In order to estimate the wellness of the predictive model that has been trained that is dependent upon the data size that the system is fed, the testing has been performed on the test set (20% of the dataset) which is a set of examples that is used to evaluate the performance of fully trained classifiers. The models are adjusted accordingly to minimize the error rate of the classifier. Among all the classifiers used in Apple Classification system, the best one is chosen and applied to the real world problem to get the results.

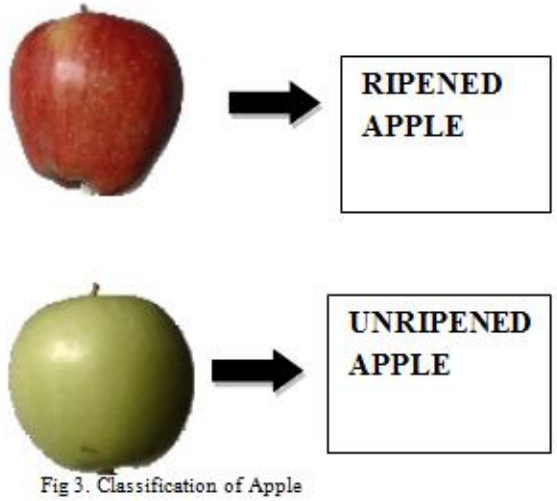

Fig 3. Classification of Apple

IV. RESULTS AND DISCUSSION

The performance of the learning model depends on the data set that has been provided to train it. The present section discusses the output of all the performance metrics that are used to evaluate the accuracy of the classifier. All the classifiers have shown different accuracy in the form of confusion matrix.

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 248, 248, 64)	640
activation_1 (Activation)	(None, 248, 248, 64)	0
max_pooling2d_1 (MaxPooling2	(None, 124, 124, 64)	0
conv2d_2 (Conv2D)	(None, 122, 122, 64)	36928
activation_2 (Activation)	(None, 122, 122, 64)	0
max_pooling2d_2 (MaxPooling2	(None, 61, 61, 64)	0
conv2d_3 (Conv2D)	(None, 59, 59, 64)	36928
activation_3 (Activation)	(None, 59, 59, 64)	0
max_pooling2d_3 (MaxPooling2	(None, 29, 29, 64)	0
flatten_1 (Flatten)	(None, 53824)	0
dense_1 (Dense)	(None, 1)	53825
activation_4 (Activation)	(None, 1)	0

Total params: 128,321
Trainable params: 128,321
Non-trainable params: 0

Fig. 4 Matrix representing various functions

Also, the graph between loss and accuracy has been plotted during training of our model for 4 epochs.

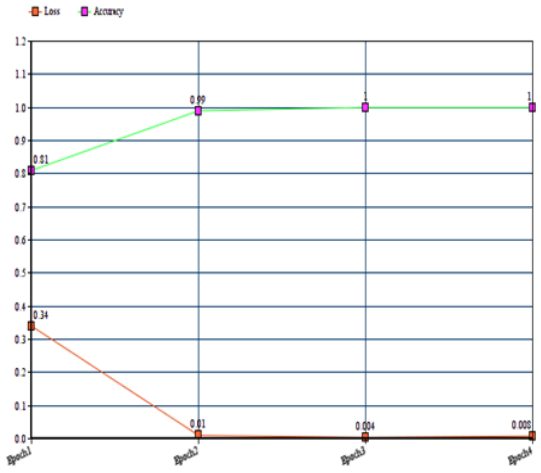

Fig 5. Graph between Loss and accuracy.

With the above results, we can say that above system accurate enough to predict good results with accuracy of 90-95%.

V. CONCLUSION AND FUTURE SCOPE

It is concluded from the results that the performance of the model depends on the following parameters:

Increasing the size of data will increase the accuracy of the classifier. Selection of appropriate machine learning algorithm will determine the success or failure. Features extracted from the dataset. Inaccurate data with noise will cause errors.

Understanding and improving upon features selecting. Proper handling of training and testing data.

Future work includes a drone for the real time measurement of the fruits on the trees for its ripeness so that we can take them down below the branches of trees before they get rotten. By using drone we can easily count the ripen fruits in a tree. The technology is developing day by day. So, it can easily carried out. Also, we can implement it in food industries to reduce the labor thereby reducing the cost. The system will automatically detect the ripe fruit and unripe fruit, and put these fruits in different containers based on their classification. Further, fruits will be used for its purpose.

REFERENCES

[1] Aggelopoulou, A.D.; Bochtis, D.; Fountas, S.; Swain, K.C.; Gemtos, T.A.; Nanos, G.D. Yield prediction in apple orchards based on image processing. Precis. Agric. 2011, 12, 448–456.

[2] B. Ojeda-Magana, R. Ruelas, J. Quintanilla-Domingvez and D.Andina, “Colour Imge Segmentation by Partitional Clustering Algorithns”, IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society, 7 10 Nov., Glendale, AZ, pp.2828 - 2833, 2010.

[3] Chiunheium Lin, ching- Hung Su,Hsuan Shu Huang and Kuo-Chin Fan, “Colour Image Segmentation Using Relative values of RGB in Various Illumination, Circumstances”, International Journal of computers, vol.5(2),pp.252-261, 2011.

[4] F. Mendoza, J .M. Aguilera, “Application of Image Analysis for Classification of Ripening Bananas”, Journal of Food Science. 69(9): pp.E471-E477, 2004.

[5] J. Blasco, N. Aleixos, E., "Molto. Machine Vision System for Automatic Quality Grading of Fruit", Biosystems Engineering, vol. 85 (4):pp.415-423, 2003

[6] Choi K. H., "Tomato maturity evaluation using colour image analysis", Transactions of the ASAE, Vol. 38(1), pp. 171-176, 1995.