**VISVESVARAYA TECHNOLOGICAL UNIVERSITY BELGAUM-590014**



MINI-PROJECT ENTITLED

“Fruits Classification Using k-Nearest Neighbors”

For the academic year 2019-2020 Submitted by:

**Name (USN)**

**Pradyumna K R 1MV16CS067**

**Rishabh Mishra 1MV16CS076**

**Rishabh Sharma 1MV16CS077**

**Rohit Mogra 1MV16CS079**

Project carried out at

**Sir M. Visvesvaraya Institute of Technology Bangalore-562157.**

Under the Guidance of

Ms. Mayuri K. P.,

Assistant Professor,

Dept. of CSE,

Sir MVIT, Bangalore.

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**Sir M. Visvesvaraya Institute of Technology HUNASAMARANAHALLI, Bangalore-562157.**

**CERTIFICATE**

**Certified that the work contained in the project titled “*Fruit Classification Using k-Nearest neighbor and Comparative Analysis via Foursquare*”, by Pradyumna K R, Rishabh Mishra, Rishabh Sharma and Rohit Mogra, has been carried out under my supervision and that this work has done for Machine learning subject.**

Prof. Supervisor Name - Mrs. Mayuri K. P. HOD- Prof. G. C. Bhanuprakash

Dept. Name – Computer Science and Engg. Dept. Name- Computer Science and Engg.

College – Sir MVIT College–Sir MVIT

**Abstract**

One of the important quality features of fruits is its appearance. Appearance not only influences their market value, the preferences and the choice of the consumer, but also their internal quality to a certain extent. Color, texture, size, shape, as well the visual flaws are generally examined to assess the outside quality of food. Manually controlling external quality control of fruit is time consuming and labor intensive. Thus for automatic external quality control of food and agricultural products, computer vision systems have been widely used in the food industry and have proved to be a scientific and powerful tool for by intensive work over decades. The use of machine and computer vision technology in the field of external quality inspection of fruit has been published based on studies carried on spatial image and / or spectral image processing and analysis. An overview of the process of fruit classification and grading has been presented in this project. Some common features of fruits like color, size and shape are used for fruit classification . Machine learning algorithm K-nearest neighbor (KNN) is applied. Process, advantages, disadvantages, challenges occurring in food-classification and grading is discussed in this project, which can give direction to researchers.

**Acknowledgements**

We take this opportunity to acknowledge the many people whose invaluable support have made it possible for us to come this far.

We are deeply indebted to our mentor, Mrs. Mayuri K. P. for guiding us through this project.

We would like to thank our HOD, Prof. G. C. Bhanuprakash for his support.

We thank Sir MVIT, and VTU for providing us this opportunity.

Finally, we would like to wind up by paying our heartfelt thanks to all our friends and classmates.

Pradyumna K R

Rishabh Mishra

Rishabh Sharma

Rohit Mogra

**Contents**

**Abstract**

##### Acknowledgements iii

##### 1 Introduction 1

1.1 Motivation....................…………..……………………………………. 1

**2 Related work 2**

**3. Implementation 3**

**4. Results 4**

**5 Conclusions 5**

##### References 6

**Chapter 1**

**Introduction**

India is an agricultural country. India exported $39 billion worth of agricultural products in 2013, making it 7th largest agricultural exporter worldwide. According to 2010 FAO, India is world’s largest producer of many fresh fruits and vegetables, milk, major species, jute, millet and castor oil seeds. India is world’s second largest producer of wheat and rice. India is world’s second or third largest producer of several dry fruits, agriculture based textile raw materials, roots and tuber crops, pulses, farmed fish, eggs, coconut, sugarcane and numerous vegetables

International comparisons reveal the average yield in India is generally 30%-50% of the highest average yield in the world. Agriculture has comprised of 16.5% GDP by sector (2016 est.) with approximately 50% labor force (2014 est.) and 10% total export. We can help to reduce the price of products by the classifying the fruits in the categories without the intervention of labor.

* 1. **Motivation**

Among the fruit classification techniques, fruit grading is one of the most important and difficult task as in the supermarket the wholesaler need to know the different categories of a fruit element to determine its price.

In order to reduce the manual work of classification and sorting to improve the quality of the fruit grading, we can use the image processing and machine learning algorithms. This helps to reduce the price of labor from wholesaler’s perspective and helps him to grade the fruits and price them accordingly.

**Chapter 2**

**Related Work**

Fruits Shape Variation Analyzer for tomato and other

plant species has been developed to analyze shape and

size of tomato fruit and other plant species[13], such as

Butternut squash, yellow squash, large jalapeno, banana

pepper, chili pepper, grape, strawberry, and Bartlett pear.

It can accurately determine boundaries of fruit with

different color. The Analyzer can perform detection for

tomato fruit shape variation and describe any other two-

dimensional fruit shape. It provides intuitive descriptors

and output that facilitates the analysis of fruit morphology.

Fruits Shape Variation Analyzer for tomato and other

plant species has been developed to analyze shape and

size of tomato fruit and other plant species[13], such as

Butternut squash, yellow squash, large jalapeno, banana

pepper, chili pepper, grape, strawberry, and Bartlett pear.

It can accurately determine boundaries of fruit with

different color. The Analyzer can perform detection for

tomato fruit shape variation and describe any other two-

dimensional fruit shape. It provides intuitive descriptors

and output that facilitates the analysis of fruit morphology.

Fruits Shape Variation Analyzer for tomato and other

plant species has been developed to analyze shape and

size of tomato fruit and other plant species[13], such as

Butternut squash, yellow squash, large jalapeno, banana

pepper, chili pepper, grape, strawberry, and Bartlett pear.

It can accurately determine boundaries of fruit with

different color. The Analyzer can perform detection for

tomato fruit shape variation and describe any other two-

dimensional fruit shape. It provides intuitive descriptors

and output that facilitates the analysis of fruit morphology.

Fruits Shape Variation Analyzer for tomato and other plant species has been developed to analyze shape and size of tomato fruit and other plant species, such as Butternut squash, yellow squash, large jalapeno, banana pepper, chili pepper, grape, strawberry, and Bartlett pear. It can accurately determine boundaries of fruit with different color. The Analyzer can perform detection for tomato fruit shape variation and describe any other two-dimensional fruit shape. It provides intuitive descriptors and output that facilitates the analysis of fruit morphology.

The mathematical descriptors can calculate the fruit traits with a single equation for each trait, include fruit shape, comprises fruit shape index, fruit shape triangle, fruit shape eccentric, fruit end shape, fruit shape heart, circular, ellipsoid, rectangular; and fruit size comprises fruit height, width, mass, area, and perimeter. The analyzer contains function for manually adjustment the distal and proximal ends of the fruit. Therefore, no matter fruit are positioned at an angle, the analyzer also can correctly identify the distal and proximal ends of the fruit. This method accuracy is different based on the fruit image but the overall is around 80% to 85%.

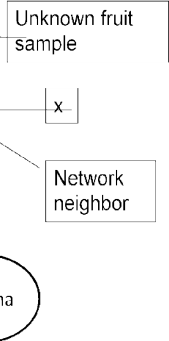
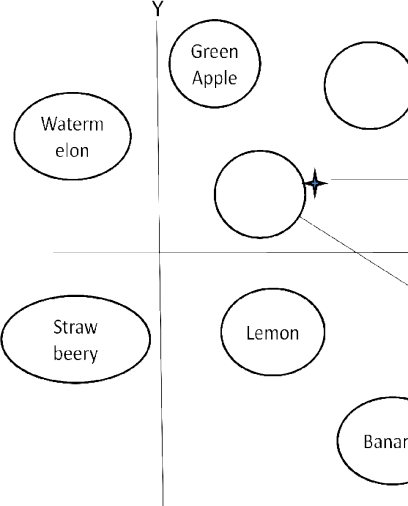
An AccuRange4000-IR point laser-range finder and a phase shift laser technique have been applied for the fruits recognition and localization to perform automatic selective harvesting. The system uses both of the range and amplitude images provide by the laser rangefinder scanner in each scanned scene, and exploits the shape-based methods analysis strategies for fruits recognition and position detection. The main stages of this shape-based methods recognition strategy are adaptive image smoothing, primitive generation, parameter and evidence estimation, hypothesis generation & verification. Adaptive image smoothing is necessary to filter the Gaussian white additive noise in a range images. Next, primitive generation stage is needed for recognition of spherical objects in the range image. Then, parameter and evidence estimation stage is needed to estimate the sphere parameter, included 3-D position, radius and reflectance of fruit, and to estimate the degree of confidence over that estimation.

Lastly, hypothesis generation and verification stage are used to reject the hypotheses that do not have sufficient evidence value. This method can recognize green fruits and Correctness of this method is 80%. On tree fruit recognition using texture properties and color data, a vision based algorithm presents to locate apples in a single image. Texture based edge detection has been combined with redness measures, and area thresholding followed by circle fitting, to determine the location of apples in the image plane. It was shown that redness works for red apples as well as green apples. This increased texture contrast helped to identify apples separately from background. The algorithm worked equally well for close ups as well as distant images of apples. Results show that the accuracy of system is around 90%.

**Chapter 3**

**Implementation**

The data that we will use for this analysis is a combination of a txt file that has been prepared for the purposes of the analysis from kaggle. Txt has list fruit name and it’s subtype along with their mass, width, height and color score. As each fruit has different mass (due to different water and pulp content), width and height, these features are perfect for classifying the fruit robustly into different categories. The different color score helps in determining the edge of the fruits, which makes the model more accurate.

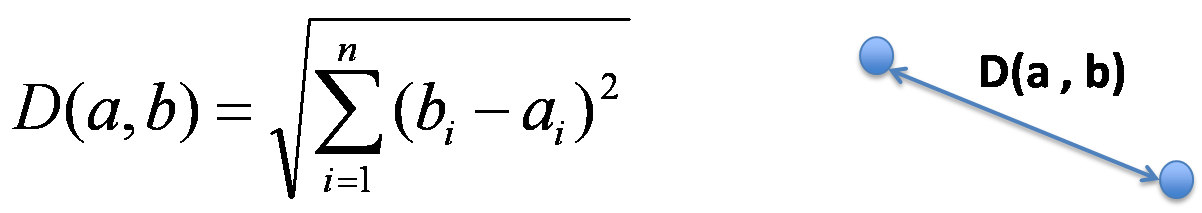


*Fig 1: Classification processes of unknown fruit sample and stored fruit sample*

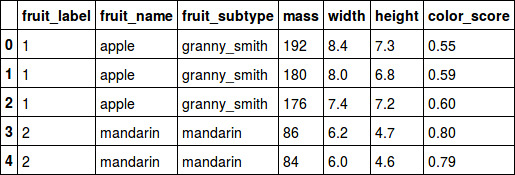
We have done the data cleaning and plotted the by using ColorMap(cm) library of python. CM makes it easy to visualize data that’s been manipulated in Python on an interactive leaflet map. The idea behind choosing a good color map is to find a good representation in 3D color space for your data set.

We have done feature scaling for each fruit in order to convert the fruit name to a specific number assigned to each fruit. In this way we can associate a fruit with a number which helps in classification. The feature value in the feature space are scaled within a range to properly fit the fruit into a category for proper classification.

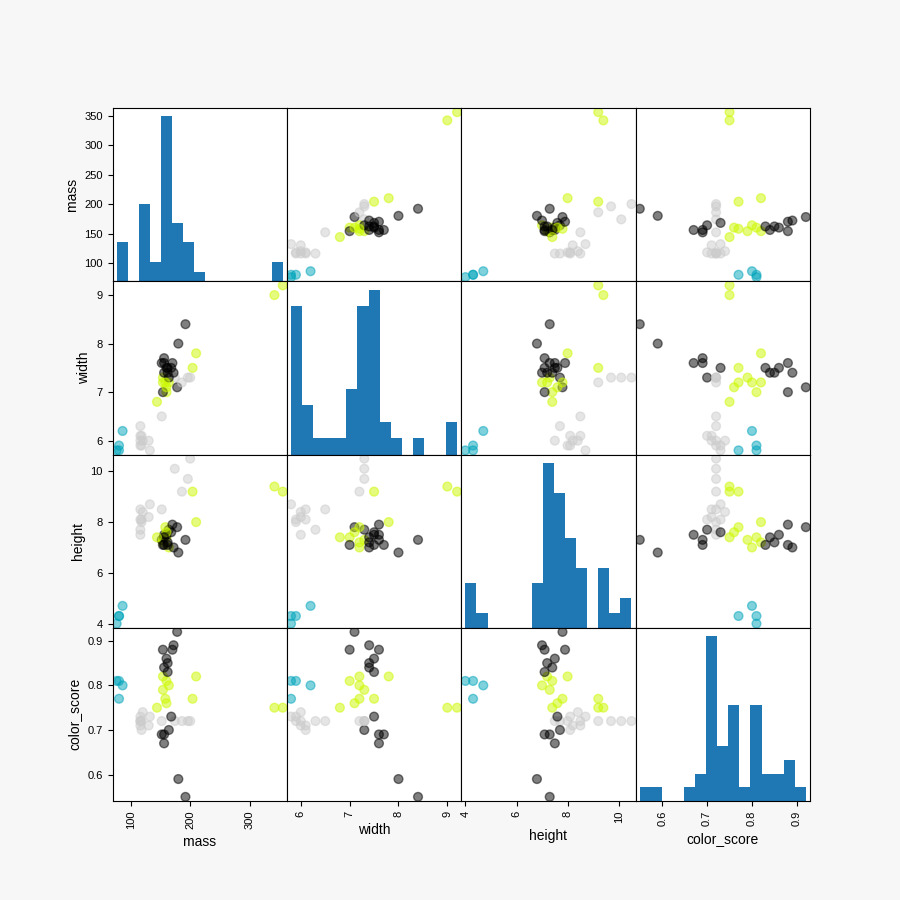
We use the below formula for our project.



Whenever a new point is introduced into the feature space we calculate the Euclidean distance for the point and the point is classified in the space which has the least Euclidian distance with respect to the point .Hence, the nearest neighbor is the classification boundary with the least distance to the point and so, the point falls into that decision boundary.



*Fig 2: Table of feature sets*

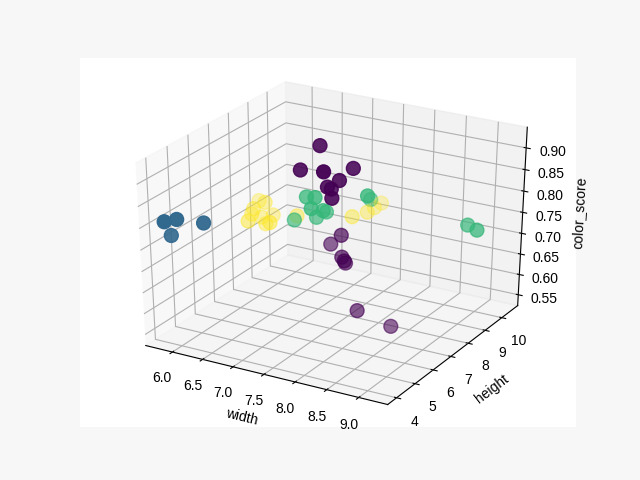
****

*Fig 3: Feature pair plot*

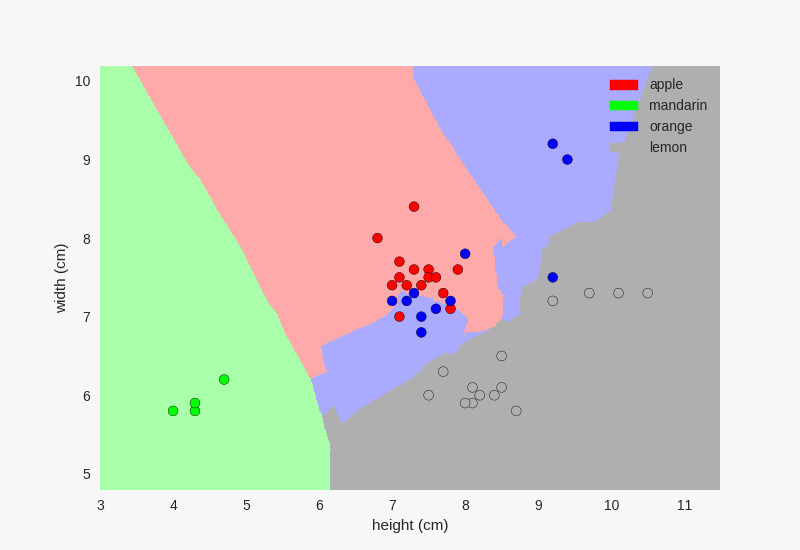
**Chapter 4**

**Result**

The proposed method can classify and identify the fruit images. Which are input and selected to the system based on shape, color, size, texture feature of the fruit. The fruit recognition system has been developed to recognize all the test fruit images. User or tester click the Training Feature Database Creation feature pushbutton, this button is used to extract the feature value of training fruit example after that click the Loading Training Feature Database this button is used to loading the training feature database then click the Load test image pushbutton this button is used to select any test (unknown) fruit image and then Extract Feature of Test Fruit pushbutton this button is used to extract the feature value test fruit image after that click the fruit recognition push button and then recognize the image. The recognition result of the accurate up to 100%.

****

*Fig 4: 3D Feature scatter plot*

****

*Fig 5: Decision boundary of the k-NN classifier*

**Chapter 5**

**Conclusions**

**Finally, we conclude our work and present the results of project work. We were successful in implementing the machine learning algorithm, and classifying the Fruits into different categories using the features like color-score, mass, height width. We also classified the fruits into different categories.**

**References**

[1] N.A. Semary et al., "Fruit-Based Tomato Grading System Using Features Fusion and Support Vector Machine",Advances in Intelligent Systems and Computing 323, DOI: 10.1007/978-3-319-11310-4\_35, Springer International Publishing Switzerland 2015.

[2] Ashwini Awate et al., “Fruit disease detection using color, texture analysis and ANN”, Green Computing and Internet of Things (ICGCIoT), 2015 International Conference on, 10.1109/ICGCIoT.2015.7380603, IEEE, January 2016.

[3] JingLei Tang et al., “Weed identification based on Kmeans feature learning combined with convolutional neural network”, Computers and Electronics in Agriculture 135 (2017) 63–70, January 2017.

[4] M. T. Chowdhury, M. S. Alam, M. A. Hasan and M. I. Khan, Vegetables detection from the glossary shop for the blind, IOSR J. Electr. Electron. Eng. 8, 43–53, 2013.