Freshness Classification of Selected Fruits through Image Processing using K-Nearest Neighbor, Support Vector Machine and Random Forest

A Thesis Proposal by

Karlo Rhommel C. Sta. Ana
Carlos Gabriel E. Del Rosario
Jose David M. De Los Santos

Submitted to the Computer Science and Information Technology Department
Bicol University College of Science

In Partial Fulfillment
of the Requirements for the Degree
Bachelor of Science in Computer Science

December 2021

APPROVAL SHEET

This research proposal entitled **Freshness Classification of Selected Fruits through Image Processing using K-Nearest Neighbor, Support Vector Machine and Random Forest**, prepared and submitted by **Karlo Rhommel C. Sta. Ana, Carlos Gabriel E. Del Rosario, and Jose David M. De Los Santos**, in partial fulfillment of the requirements for the degree **Bachelor of Science in Computer Science** is hereby accepted.

Dr. Benedicto Balilo
Content Adviser

Prof. Arlene Satuito
Programming Adviser

Prof. Ryan A. Rodriguez
Chair, Defense Panel

Dr. Lany L. Maceda
Panel Member

Dr. Jayvee Christopher Vibar
Panel Member

Accepted and approved for the conferral of the degree **Bachelor of Science in Computer Science**.

Ryan A. Rodriguez, MIT, MSCS
Chairperson, CSIT Department

Jocelyn E. Serrano, M.Sc.
Dean, BUCS

BICOL UNIVERSITY
COLLEGE OF SCIENCE
Computer Science and Information Technology Department
Legazpi City

CONTENT ADVISER'S CERTIFICATION

This is to certify that I have checked the manuscript of the Thesis Proposal with the title "Freshness Classification of Selected Fruits Through Image Processing using K-Nearest Neighbor, Support Vector Machine and Random Forest" which is submitted by Karlo Rhommel C. Sta. Ana, Carlos Gabriel E. Del Rosario, and Jose David M. De Los Santos found the following in order as indicated below:

- The TITLE is concise but captures the complete coverage of the study.
- The BACKGROUND OF THE STUDY tells the complete story of the research in an inverted pyramid style covering Global Situations, National Situations, Regional Situations, and Local Situations in relation to the topic at hand.
- The last paragraph/sentence of the BACKGROUND OF THE STUDY defines the GAP of the study as a preparatory concept to the research objectives.
- The OBJECTIVES OF THE STUDY is anchored on the identified GAP.
- The SPECIFIC OBJECTIVES are Specific, Measurable, Attainable, Realistic, and Time-Bound (SMART) and the Verbs used in the statements are based on the learning objectives defined by the Bloom's Taxonomy.
- The paper does not contain personal opinions of the researches and every claims made are properly cited.
- The SIGNIFICANCE OF THE STUDY clearly defines the beneficiaries of the outcome of the research.
- Under the SCOPE AND DELIMITATIONS, the scope indicates the coverage and boundaries of the research, and the delimitations provides the exclusions of certain concepts, features, and modules which are normally and necessarily included but the researchers opted to exclude due to justifiable reasons.
- There are sufficient and up-to-date literatures in the REVIEW OF RELATED LITERATURES that cover all the concepts defined in the Objectives and Gap of the Study and that these literatures are properly cited and anchored on a pre-defined Literature Map which is separately provided.
- The CONCEPTUAL FRAMEWORK OF THE STUDY provides a clear illustration on how the study will be conducted based on the defined objectives.
- The DEFINITION OF TERMS include only terms with operational definitions or with customized definitions which are deemed fit as used in the study.
- The METHOD discusses the detailed activities required to attain the defined Objectives and Concept of the Study. Data Gathering Procedures, Evaluation Method, and Data

Analysis Method are clearly defined and the required forms for their conduct are available and provided.

Issued this 19th day of November, 2021 as part of the requirement for Thesis Proposal Defense.

CERTIFIED TRUE and CORRECT:

Benedicto B. Balilo Jr.
Content Adviser

BICOL UNIVERSITY
COLLEGE OF SCIENCE
Computer Science and Information Technology Department
Legazpi City

PROGRAMMING ADVISER'S CERTIFICATION

This is to certify that I have checked the manuscript of the Thesis Proposal with the title "Freshness Classification of Selected Fruits Through Image Processing using K-Nearest Neighbor, Support Vector Machine and Random Forest" which is submitted by Karlo Rhommel C. Sta. Ana, Carlos Gabriel E. Del Rosario, and Jose David M. De Los Santos found the following in order as indicated below:

- The provided OBJECTIVES OF THE STUDY is sufficient to realize the goal of the study.
- Each member of the group has a well-defined task that will significantly contribute to the success of the project.
- The selected SYSTEM DEVELOPMENT METHOD is appropriate for the selected project proposal.
- The phases of the chosen SYSTEM DEVELOPMENT METHOD iterate the actual activities that need to be conducted in order to realize the objectives.
- The SOFTWARE and HARDWARE requirements provided in the paper, indicates the minimum and ideal specifications to run the target system.
- The attached DIAGRAMS such as UML, Gantt Chart (Appendix A), Flow Chart (Appendix B), Use Case Diagram (Appendix C) and Activity Diagram (Appendix D) are conceptually correct that represents clearly the flow of information of the proposed system. Standard Model is used to present the diagram.
- The attached INITIAL DESIGNS AND FORMS (Figure 3-4 and Appendix E) presents a clear picture of the proposed system.

Issued this 13th day of November, 2021 as part of the requirement for Thesis Proposal Defense.

CERTIFIED TRUE and CORRECT:

Prof. Arlene Satuito
Programming Adviser

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Background of the Study

There are numbers of foods that have gone to waste because of spoilage, the data in the article about wasted food shows that 53% of fruits has the highest percentage of being wasted before being eaten, followed by vegetables that is at 45% and Dairy products at 39% (StudyFinds.org, 2021). Growing this product requires resources, and if a huge amount of it is being wasted, therefore we are not only wasting the product itself but also the resources used in growing it. If you look at land usage, around 1.4 billion hectares of land, which is roughly one-third the world's total agricultural land area, is used to grow food that is wasted (Move For Hunger, 2021). With agriculture accounting for 70 percent of the water used throughout the world, food waste also represents a great waste of freshwater and groundwater resources. It is said that a volume of water roughly three times the volume of Lake Geneva is used just to produce food that is not eaten (Move For Hunger, 2021). Food waste that ends up in landfills produces a large amount of methane – a more powerful greenhouse gas than even CO_2. For the uninitiated, excess amounts of greenhouse gasses such as methane, CO_2 and chlorofluorocarbons absorb infrared radiation and heat up the earth's atmosphere, causing global warming and climate change (Move For Hunger, 2021).

One factor that mostly leads to food wastes are people that are not educated enough to know which fruits are still edible. At most times, people notice their fruits only when they have already rotted which practically would just end in the trash.

According to new research, young adults ranging from 18-24 years old are the greatest offenders, with 92 percent admitting to regularly throwing out fresh food. Overall, 83 percent of people stated they throw away fruit and vegetables because they don't eat it before it spoils.(FreshPlaza, 2021).

Aside from food waste, there is also what we call food loss. Food loss are those foods that are thrown away due to lack of characteristics to pass a specific market standard. Food loss can happen on transportation and distribution, on packing houses, and even on the source of production. On farms, food loss occurs for a variety of reasons. Weather, pests, and illness may prevent food from being harvested. Farmers will sometimes leave their crops unharvested if the market price of produce is lower than the cost of transportation and labor. Cosmetic flaws are another major source of food waste, as consumers are less interested in foods that are misshapen or blemished.(FoodPrint, 2021).

Researchers have been exploring the use of image processing to detect the quality of fruits and vegetables. The majority of their techniques use a hierarchical grading system. Alternatively, they process photos using various filtering algorithms using computer vision tools like MATLAB.(A. M., 2014) Gastélum-Barrios et al. (2011).

Our study aims to help people determine the state of their fruits and create an idea on when it will spoil, to prevent it from being a food waste, improve their quality of consumption and at the same time, even in a small way, help in protecting the environment from the effects of food waste.

Objectives of the Study

To develop an application that will classify whether a fruit is still fresh or not using image processing and machine learning algorithms.

Specific Objectives

1. To gather images of fruits, such as bananas, mangoes, and papayas, and to subject them for image cleaning in order to assess the physical features of each kind of fruit. Images containing skin features will be used for the stale fruit dataset, whereas images without skin features will be used for the fresh fruit dataset.
2. To train the fruit detection and freshness classification models, which are developed using different algorithms such as K-Nearest Neighbor, Support Vector Machine, and Random Forest, and then to implement the most accurate algorithm out of the three.
3. To determine the size, calculated area, nutritional facts/values of bananas, mangoes, and papayas based on the uploaded/captured image.
4. To evaluate the classification model's efficacy in identifying and classifying the fruit using accuracy, precision, recall, and f-1 score evaluation metrics.

Significance of the Study

This study of Freshness Classification of Selected Fruits through Image Processing using K-Nearest Neighbor, Support Vector Machine and Random Forest would benefit the following:

Consumers, to help them choose the best quality of the fruits they buy as well as plan their consumption, so less food would be wasted. Our study can help them acquire an idea on the quality of fruits they eat, in that way, educate them about the responsibilities of food consumption. This will ensure that the fruits they buy are healthy and fresh. By identifying which are fresh fruits, the consumers save money because they can choose which fruit has superior quality .

Local Greengrocers, to help them pick the best fruits to sell; fruits that they locally grow. As well as prove to their buyers its quality and grade.

In The Fruit Production and Marketing Industry, our study can help people in the fruit industry identify which of their products are still fresh and which are not. Quality control will be made easier and faster because there is no need to individually inspect the state of the fruit, a single decaying fruit can contaminate the entire cluster. Our research will employ image processing and machine learning to quickly identify which fruits are fresh and which are rotten. To avoid infection, the rotting fruit can be segregated from the fresh fruits after being identified.

Chefs or Cooks, the people who prepare food must always use the best ingredients for their dishes. This is to help them to always acquire and use the most fresh fruits for their cooking. To help them save time in picking the fresh fruits from their cold storages and to help them maintain the quality of their dishes by keeping the ingredients fresh always.

Fruit Inspectors, they dedicate their life in maintaining and monitoring cold storages of fast food changes, restaurants and other food service industries. Keeping their fruits in top shape gives the food company the ability to produce high quality service. To help inspectors easily determine the freshness of the fruits. To speedup the process of assessment as well as deliver a uniform report for each fruits that is scanned

Nutritionist Dietitian or Nutrition Counselors, their task to create a diet plan for their clients requires a tremendous amount of knowledge about food. This will help them easily check the nutritional facts of each type of fruit. For some Dietitians who provide the meal for their clients, they always need to provide the ones that are fresh specially for fruits and vegetables. In these situations, this is to help them choose the most fresh fruits they can get so everytime they make a meal for their clients, they always give the best possible service.

Fruit Distributors, Not only will the consumer benefit from the research, but also the fruit distributors. They will be able to assure their customers that the fruits they deliver are quality checked. They can attach the result of the detection to the product. This way they can show their buyers about the quality of the fruit they delivered.

Computer Science Students and Future Researchers, our study will benefit future researchers who are taking up the same approach to their research. It will serve as their guide on what methods to undertake in order to complete their work. They can also explore our study and make further improvements on its object detection and freshness classification features.

Scope and Delimitations

Images to use as a dataset must only contain images of papayas, mangoes and bananas. The images were obtained locally by the researchers by means of capturing the images and taken from a popular website named kaggle.com. It's been modified to just show papayas, mangoes, and bananas. The images taken from the website only have white background, while the images captured locally by the researchers have different backgrounds. The physical features of the fruit such as size, calculated area, and skin features are the only things that will be identified in order to give an estimate on the status of the fruit. Upon comparing the K-Nearest Neighbor, Support Vector Machine and Random Forest; the most efficient among the three algorithms will be selected and adapted to the android platform.

Not all fruits will be collected. Not all the variety of fruits will be included in the dataset, only the commonly available fruit in the Philippines, for example Lakatan bananas for bananas, yellow Carabao mango, also known as the Philippine mango or Manila mango for mangoes, and lastly native papayas for papayas. The nutritional facts of the fruit will not be dependent on the size of the fruit. The results of the evaluation metrics that are used for the assessment of this current study does not define the efficiency of the three algorithms generally but for this specific research only.

CHAPTER 2

RELATED LITERATURE AND STUDIES

This chapter contains a compilation of literature and studies collected by the researchers which are conceptually related to this research. A conceptual framework which illustrates the flow of the development is also included in this chapter as well as a list definition of terms that will help the readers determine the subjective meaning of words found in the paper.

Review of Related Literature

Quality of Fruits

Fruits are a staple to every household. Fruits supply essential nutrients for good health and body maintenance. As cited by Vincente et al (2014) in their paper, water, fiber, proteins, lipids, organic acids, and digestible carbohydrates are among the nutritional components derived from fruits and vegetables. In certain areas, starch-based staples provide a key source of energy. Minerals and vitamins are also found in fruits and vegetables. They are the major source of vitamin C in the diet, as well as a good source of provitamin A and vitamin B6. They are high in potassium and low in sodium when compared to other dietary sources which makes them more healthy and life sustaining. With this in mind, according to Musacchi and Serra (2014), Apples are grown in temperate, subtropical, and tropical climates all throughout the world. Commercial production is restricted to latitudes of 25° to 52°.Apples are grown in 63 countries with a wide range of growth conditions and a big number of cultivars. Apple consumption per capita in the world reached 9.2 kg per person in 2013. After all citrus fruits, grapes, and bananas, apples are the world's fourth most important fruit crop**.**

The quality and transportation cost of a fruit determines its price. Fruits purchased in the public market are far less expensive than those purchased at shopping malls. Consumers who have purchased fruits before know what to look for in terms of quality and freshness. The quality of a fruit can simply be assessed by its appearance. According to Lakare & Kapale's (2019) research India's fresh fruit exports are increasing on a daily basis. People are highly health-conscious, and they only eat fresh, high-quality fruit.The three most essential characteristics for determining fruit quality are texture, color, and size. Color recognition is a key part of the ripeness detecting procedure. The identification of ripeness is an external quality element. Texture, on the other hand, is crucial. Fruit with defects can be identified by their texture. The non-uniformity of the fruit's exterior surface is detected via texture analysis. The size is also an essential factor to consider. Moreover According to Linke, Herppich and Geyer(2010), the quality of a fruit like its color, gloss and surface pitting degrades while it is still in the market. Fruit softening , peduncle browning and dehydration was observed on cheries according to the research. Freshness of a particular fruit solely depends on multiple factors that contribute to its quality. One contributing factor is the environment of the fruit and how it is handled upon harvest. Nordey, Davrieux, & Léchaude (2019) stated in their paper that Fruit variability in terms of quality and maturity at harvest is a challenge that needs to be handled across the supply chain in order to prevent postharvest losses and provide consistent quality for customers. Fruits are usually sorted and graded after harvest to generate homogeneous batches, assuming that their post-harvest behavior and quality will be comparable following ripening. In addition as stated by Silapeux, Ponka and Frazzoli (2021), The lack of education and training in post-harvest fresh fruit care has a significant impact on the shelf-life of marketed fruits as well as fruit waste. A large number of fruits are rejected because they do not meet the shape, appearance, color, or size requirements. These quantities of rejected fruits should be seen as nutrient loss; on the other hand, such stocks of nutrients could be beneficial. furthermore According to Kader (2008), the difference between the time of harvesting the tomato to the time it was eaten can give a significant degradation to the tomatoes' smell feature and occurrence of change in its flavor .

Choosing what fruit or vegetable has better quality to take in a market is important to maximize the money you will be spending on these products. The study of Pe´neau, Hoehn, Roth, Escher and Nuessli stated that the result of their study was inline with the conclusion of another researcher they cited, stating that taste, freshness and appearance are the most important attributes for the choice of fruits and vegetables. Only in other instances will you be able to taste a food before buying it. Sometimes it is challenging to determine product quality and freshness only by its appearance, especially if information about that product is not provided. Péneau, Linke, Escher, & Nuessli study states , Consumers' perceptions of freshness are influenced by the amount of information they have about products. As a result, the importance of providing sufficient product information is highlighted. Freshness refers to how close a product is to the original in terms of distance, time, and processing.

Consumers have a common sense of freshness, but they characterize it based on their individual experiences with fruits and vegetables.

Keeping track of a perishable product's freshness is very important in order for it to be delivered, sold and still safe for consumption so that it won't go to waste. Valentino et al. (2021) research states that the fruit also has a freshness period during which it is believed to be fresh. During this time, many fruit supplier firms continue to supply fruit that is unsafe for ingestion due to inaccuracy in the sorting process when the fruit is taken from the plantation and the introduction of other fruit into an incorrect packing. As a result, detecting food rotting from the point of production to the point of consumption is critical. Freshness aside, determining a fruit's maturity is also an important factor to consider for farmers to deliver their product in their best state and their products' marketability. According to Kasim, Abidin, Mangshor, Hamzah, Rahim (2019) A huge export market does not indicate a large growth rate. The maturity of banana bunches is one of the important criteria that determines the quality of fruit during ripening and overall marketability. Farmers will be able to optimize the harvesting phase if they can learn to determine the maturity of fresh bananas, according to Surya & Satheesh. In other words, this skill will help you avoid harvesting bananas that are under-matured or over-matured. Farmers encounter a problem in maintaining fruit quality standards that meet international standards due to a lack of knowledge and skill. While for Pallottino et al (2013) In the field of food science, color is a critical criterion for determining freshness, quality, and consistency. Cool storage of fruits after harvest slows respiration and color ripening. The post-harvest chain, which includes extended transportation transit times, storage, and selection processes, can lead to a loss of fruit quality in terms of appearance, which is a key criterion for consumer preference. Throughout the supply and consumer chain, appearance is used as a selection factor in particular. And in line with it, Narendra & Pinto claims that, to evaluate the overall level of acceptability for clients, consistency in the size, shape, and other quality characteristics of fruit and vegetables is required. Color, size, form, texture, and quantity of faults are the most important characteristics of fresh examination. Various causes, such as rotting, bruising, scab, fungal development, injury, illness, and so on, cause defects or damage in fruits and vegetables. After harvesting fruits and vegetables, proper care should be taken. To avoid cross-contamination and save processing costs, these flaws must be eliminated.

It is important for a perishable product to reach its consumer at its best quality. Environmental factors affect the quality of a perishable product. According to Sanchez et al. (2020) ,one of the main challenges of the food industry is the control of environmental factors. Quality of a perishable product is mainly affected by the temperature. Relative humidity and gas concentration plays an important role in maintaining the postharvest quality of horticultural products. Monitoring environmental factors is important in assuring quality of fruits and vegetable shelf life and evaluating losses. It is almost impossible for a human to manually quality check every fruit in a container in a fast and efficient manner. Having an automated grading system will make these problems way easier. The study of Telang, Shirsath (2017) stated that grading may be done in a variety of ways, including color and size. In the first example, they sort circular-shaped fruits by color, then grade them according to size. The suggested automated categorization and grading system combines three processes: feature extraction, color sorting, and size grading. In this color categorization system and for determining the size of a fruit, software development is critical.

Fruit consumption is essential in one's diet. It provides the body with fresh and healthy nutrients that it requires to function properly. The nutritional value of a fruit is also affected by its state and quality when consumed. Consuming an unripe banana can make a person constipated. On the other hand, consumption of a ripe banana gives plenty of fiber which is vital in good bowel movement. Identifying the quality and state of a fruit is vital to ensure that it is worth buying it. The quality is affected by multiple factors, which are the environment,weather, and temperature. Upon the time of harvest the post life of the fruit is already beginning. Physical inspection of the fruits is crucial since it ensures that they are safe for consumption. The manner of how the fruits are handled upon harvest affect its life and quality. Improper handling of the fruit can damage its skin and the fruit itself.

Image Processing

Computer vision technology has been shown to be a useful tool for fresh fruit and vegetable inspection, and it is now being used in packing houses' automated postharvest fruit and vegetable grading systems. Fruit inspection prior to arrival at the packing lines could provide a number of benefits, including having an accurate fruit assessment to choose between different fruit treatments, as well as cost savings in transportation and marketing non-commercial fruit.

Image processing has greatly contributed in extracting useful information from an image that is used for computer vision, researchers further explained its concept to have a better understanding of how it works and for it to be implemented properly. According to Huang, Schreiber and Tretiak (2002), the area of image processing is concerned with manipulating data that is fundamentally two-dimensional. Image processing techniques are used in a variety of fields, including image augmentation, pictorial pattern detection, and efficient picture coding for transmission or storage. It is also Stated in McAndrew’s Book called Introduction to Digital Image processing with Matlab (2004), Image processing has a wide variety of applications. Techniques may be used in virtually every field of research and technology. Different image processing methods can be divided into broad subclasses for convenience. Different algorithms exist for various tasks and problems, and we frequently would want to identify the nature of the task present.

Implementation of image processing greatly contributed to the classification of products. The study of Dubey & Jalal (2015) stated that In the agricultural sciences, images are a valuable source of data and information. The application of image-processing tools to agricultural operations has far-reaching ramifications. Fruit and vegetable classification is one of the most common applications in supermarkets for automatically detecting the types of fruits and vegetables purchased by customers and determining the appropriate pricing.

Bin, E. (2020) stated that color may be a useful criterion for distinguishing lemons from oranges. Assume that lemons are typically green and oranges are typically orange. If our orange isn't quite ripe and still a little green, we may compare measurements and declare it's an orange if it's larger than a specific measurement; otherwise, it's a lemon. It may appear to be a ridiculous circumstance, yet it is exactly how a Decision Tree algorithm works. It will go into the data and select the most essential attribute that best distinguishes one class from the others, then repeat the process for the other features until it discovers a method to better determine which fruit it is. That is why Decision Trees algorithms may be employed as a tool for feature selection, such as identifying traits that are more important to our identification and removing those that contribute little or nothing to our primary aim.

However, it still poses a challenge regarding how the object in the image will be processed with high accuracy and for the computer vision system to easily recognize it. The study of Ismail and Malik (2021) states that in most vision systems, unwanted high-frequency signals (random noise) damage recorded pictures, causing unpredictable fluctuations in intensity, lighting, and poor contrast. For picture noise reduction, they used a Gaussian filter. A Gaussian filter is a linear smoothing filter with isotropic or circular symmetrical weights derived from a normal distribution in the kernel. The Gaussian filter is a two-dimensional zero-mean discrete function used in image processing that results in uniform smoothing throughout the picture in all directions. Another approach in processing an object in an image is to measure the pixel per metric ration. According to Mustaffa and khairul (2017), the pixel per metric ratio can be used to measure objects in pictures. The ultimate objective is to determine the size of an object in a photograph by measuring it against a reference object. Two mandatory properties should be present in the reference object. To begin, the object must be in a measurable unit (such as centimeter or meter) and have a dimension (such as width or height). The reference item in a photograph is identified by its position (for example, the reference object is always placed in the top left corner of an image) or by its look (for example, by being a distinctive color or form, uncommon, and distinct from all other objects in the picture).In any instance, the reference should be easily recognizable in some way.

Images have been a valuable source of data and information, through Image processing this data and information can be extracted in order to be used in different fields of research and innovation. Images have different illumination and some undesirable high frequency signals, this affects in extracting a particular information from an image resulting in inaccurate outcome. Through the help of these studies, we will be able to apply these techniques and methods to extract information from an image containing a fruit in order to determine its freshness based on its external features with a desirable outcome and accuracy.

Different Approaches on Quality Detection of Fruits

Object detection and Object recognition is the fundamental purpose of image processing. Its ability to understand and interpret objects the same way with us people. In terms of fruit detection, computer vision can determine the state of the fruit and detect various diseases present. According to Li, Feng, Liu and Han (2021), there are different approaches that have already been tested to be effective in detecting the quality of fruits. Those are Bio-molecular sensing technology, hyperspectral imaging techniques, multispectral imaging and traditional machine vision technology. It was also stated that the huge amount of data and uneven brightness is a common issue specially on hyperspectral imaging techniques. On the other hand, traditional machine vision based on color cameras can achieve most quality inspections of fruits. As per Siswantoro, Arwoko, & Widiasri (2019), Fruit recognition using images can be used in a variety of industries, including the food business, food retail, and medicine. To lower the dimension of input features and improve classification performance, a simple feature selection method based on variance was used. A feature was excluded from feature space if its variance was less than a predetermined threshold. To dive even deeper into image processing, Deep Learning Technology gives a big leap to the accuracy of detection. Koirala, Anand, Walsh, Kerry, Wang Zhenglin, McCarthy and Cheryl (2019) stated that advances can come from the agricultural sector, for example, the discipline of near infrared spectroscopy was developed to assess forage and grain quality, but they are more likely to come from better-resourced sectors like medical and security, and diffuse into the agricultural sector. Deep learning (many layer neural networks) in machine vision is an example of the latter, as it represents a step forward beyond algorithms focused on handcrafted features like color, shape, and texture.Deep learning applications in agriculture are still in their infancy, with Kamilaris and Prenafeta-Bold (2018) citing only 40 published papers (using a broad definition of deep learning) and four reports on the issue of in-field fruit counting. One capability of DL is Convolutional

Neural Network. State by Kang, J., & Gwak, J. (2021) The convolutional neural network (CNN) is a type of artificial neural network that filters inputs for usable information using convolutional layers. The output of neurons associated to local regions in the input is computed by CNN's convolutional layers and convolutional filters. It aids in the extraction of temporal and geographical characteristics. But aside from CNN there are a broad range of deep learning algorithms and API's. As stated by Fu, Yuhang (2020) In his project, where he examined various algorithms, YOLO for detecting region of interest in digital images, ResNet, VGG, Google Net, and AlexNet as the base networks for freshness grading feature extraction, and ResNet, VGG, Google Net, and AlexNet as the base networks for freshness grading feature extraction. Fruit decays over time, hence this characteristic is added for freshness grading by evaluating chronologically-related fruit decaying data.

A lot of researchers have come into interest on topics just like ours and many of them share and exhibit different approaches. Many studies about object detection, specifically fruit detection, have been performed by a lot of researchers, and the more of them who gain interest the more its algorithm becomes more accurate.The study of Yogesh , Dubey, Ratan and Rocha (2019) tackles an approach in order to detect the quality of fruit using computer vision that is composed of 7 levels of processing. It starts with Image acquisition then image reprocessing, color processing, segmentation, feature extraction, classification and lastly defect detection. Another approach is introduced In the initial study of Bhargava and Bansal (2018), they used an approach that involves 5 steps of processing an image. It involves image acquisition, pre-processing, image segmentation, feature extraction and classification. Dubey and Jalal (2015) proposed a different approach in finding physical features of a fruit. Precise image segmentation is necessary for the fruit disease classification problem; otherwise, the features of the non-affected region will dominate over the features of the diseased region. K-Means based image segmentation is preferred in this approach to detect the region of interest, which is only the diseased part. Features are taken from the segmented image of the fruit after segmentation. Finally, a Multi-class SVM classifier is used for training and classification. According to Sa, Ge, Dayoub, Upcroft, Perez, and McCool (2016), it is also possible to use Conditional Random Field (CRF) for Fruit detection. The CRF takes into account both color and texture. The color characteristics are created by translating RGB values to the HSV color space directly. The NIR channel is used to retrieve visual texture information. Because NIR images are more reliable than color imagery, they are utilized to calculate texture features. Another study, the study of Nayak, Manjesh and Danusha (2019) stated that their framework mainly consists of two phases because of the different natural conditions of captured images. textual features are extracted from fruit in the first phase and fruit is classified as detected fruit in the second phase. Input to the SVM classifier are the measurements obtained from the study of textural features which is used for training its classifier. Lastly with the most identical research to ours came from Zawbaa, Hossam M.; Hazman, Maryam; Abbass, Mona; Hassanien, Aboul Ella (2014), in which they concluded that The experimental results show that the accuracy of similarities on shape groups archives the lower accuracy among the three groups. For orange and strawberry groups, the height accuracy is achieved (100%) with SIFT feature extraction and KNN classifier. The height accuracy achieved in the second group is for 96.97% apple (with RF classifier), and for 78.89% orange (with SVM classifier). Classify third group achieves 96.97% for apple when using SIFT feature extraction. While strawberry high accuracy is 92.31% when using SIFT as feature extraction and KNN as classifier. In general, RF based algorithms provide better accuracy compared to the other well known machine learning techniques such as K-NN and SVM algorithms.

Other researchers focused on analyzing colors in order to detect a fruit and its characteristics. A method was created by Chandini and Uma (2018) to capture fruit pictures. The work is split into two halves. The first step is to separate the backdrop from the foreground. Only the B component map from the RGB color model is used for background separation, and the Otsu method is applied to the B component. The defective region and the healthy area must be separated from the citrus fruit's surface for the second stage. A GB chromatic map from RGB color space was chosen for distinguishing the healthy and defective areas, and a circularity threshold was used. The settings of the higher threshold, lower threshold, and Gaussian filter were established in the circularity threshold technique to identify the contour of the outer surface of the fruit. During the second stage, the radius of the fruit was also defined. When segmentation was done utilizing the circularity threshold approach, the performance was improved by 15.58%.

The most common processes involved in detection of fruits through image processing are segmentation and classifier. Both of these two are crucial because these are the ones responsible for filtering, masking and classification of the image. According to Ayan, Demirez, Kiziloz, Inci, Isleyen and Erginm (2018), stating about their study, In the marketing of vegetables and fruits, quality is crucial. All of the products coming in from the fields have already been inspected and the rotten ones have been removed. This procedure does not satisfy the highest quality standards, and it wastes time and labor. For detecting spots or rottenness on apples, morphological image processing techniques were utilized, and image processing was done with Matlab software. To recognize the properties of the apples, an image processing technique was built. While ,stated by Choudhary, Khandekar, Borkar and Chotaliya (2017) in their project, to segment photos of fruits taken in natural daylight, edge-based and color-based detection algorithms are commonly utilized. Digitized photographs of mango fruits, as well as their backgrounds, were gathered from the Internet for this project in order to discover a mango in each image and pinpoint its specific location. In terms of masking, Patil, Jadvah, Dalvi and Kulkarni(2020) shared their insight of algorithms in processing and image of a tomato, One of the most essential preprocessing stages is to identify a faulty tomato. The skin defect is computed. For the study, a color picture of a tomato was employed. If the pixel value is less than the threshold value, it is regarded as part of the faulty skin, i.e. low-quality fruit. Any pixel value larger than the threshold value represents pure skin, or high-quality fruit. The image is a mask, and the parts of the image that aren't harmed are shown in black, while the parts that aren't are shown in white. The total number of white pixels is then computed, which equals the total number of pixels that belong to injured skin. While for classifiers as differentiated in the paper of

Almadhor, Rauf, Lali, Damasevicius, Alouffi and Alharabi (2018), Plant diseases can reduce the quality and quantity of agricultural products significantly. Furthermore, postharvest infections might result in significant output losses. Farmers' poor usage of pesticides may result in financial losses due to incorrect diagnosis. To sustain high crop fields, automatic detection of illnesses in plants as they appear on the plants' leaves and fruit is essential. When compared to individual channels, a combination of color and textural features is used to identify and achieve similar results, while disease recognition is accomplished using advanced machine-learning classifiers (Fine KNN, Complex Tree, Boosted Tree, Bagged Tree, Cubic SVM). Horea and Mihai also stated, In order to get an effective classifier, you need a high-quality dataset. As a secondary goal, we've trained a deep neural network that can recognize fruits from photos. This is consistent with the present trend of organizations working in the field of augmented reality. The correct identification of the items is the first stage in developing such an application.

Kamil, I. A., & Fagbadegun, A., (2017) shared in their study that to enhance accuracy and speed of machine learning Convolutional Neural Network is introduced along with image processing. Fruit identification is a critical step in the automated sorting and grading of fruits. Image processing and computer vision techniques are commonly used in automatic fruit recognition. Researchers have lately used a variety of soft computing techniques in the learning module of object recognition algorithms, with Artificial Neural Network (ANN) Backpropagation(BP) algorithms being the most commonly used for fruit recognition. Different ANN Backpropagation algorithms perform differently for different jobs. The goal of this study is to figure out which ANN BP method is best for fruit recognition.The utilization of cutting-edge technologies is critical in the agri-food industry. José Naranjo-Torres et al (2020) stated in their study Artificial intelligence is currently one of the most essential technological tools in modern life. Because of its capacity to create robust representations from images, Deep Learning (DL) has a wide range of applications. The predominant DL architecture for image categorization is Convolutional Neural Networks (CNN). We give a review of the application of CNN applied to diverse autonomous processing tasks of fruit images: classification, quality control, and detection, based on the increased interest in CNNs in recent years.We see that in the recent two years (2019–2020), the adoption of CNN for fruit recognition has skyrocketed, with outstanding results obtained either by employing new models or transfer learning with pre-trained networks. According to Hussain, He and Chen (2018), Because of the similarities between different types of fruits and external environmental changes such as lighting, automatic fruit recognition using machine vision is considered a difficult task. Fruit recognition using a Deep Convolution Neural Network (DCNN) is proposed in this research. Without removing features, photos were used directly as input to DCNN for training and recognition; additionally, DCNN learns optimal features from images through an adaptation process. Example of CNN in image processing is the proposed machine learning approach of Mudaliar and Priyadarshini states that a convolutional neural network (CNN) is a type of artificial neural network that analyzes data using perceptrons, which are a machine learning unit algorithm. Image processing, natural language processing, and other cognitive tasks can all benefit from CNNs. Mudaliar, G., & Priyadarshini, R. B. K. (2021) explains CNN as a multilayer feed-forward neural network (FFNN) that can detect, categorize, and identify any feature in an image fast. It's mostly utilized with visual data, like picture categorization. According to Ananthanarayana, T., Ptucha, R., & Kelly, S. C., (2020) the data used to train the CNN model is quite important. If the training data consists of low-quality photos, the CNN will be unable to accurately identify the item in question and will most likely perform poorly in image classification and object identification tasks. When statistically relevant training pictures are combined with optimized CNNs and inference algorithms on edge platforms, agricultural imaging classification and detection use cases such as freshness of fruits, vegetables, and meat benefit significantly. It is important to pick CNN architectures that are light-weight in order to infer on edge platforms and achieve good performance.While others study and explore algorithms, some researchers study the placement of the fruit and how the data would be taken. Sarkar, Gupta, and Assaad (2019), explained in their paper, Consumers and the food business are becoming increasingly interested in food quality and freshness control. During ripening and storage, for example, the optical absorption and scattering properties of fruits change due to physiological and metabolic changes in the tissues. This work considers the basic spectral feature of fruit reflectance in the visible range for the creation of algorithms to nondestructively monitor the ripening and deterioration of apples using phase information from reflected light. The goal is to create simple, low-cost equipment that could revolutionize food quality monitoring in real time. Arake and Lakshmana (2016), stated based on their experience, The tomato fruit is quite fragile, therefore it must be handled with care while grading. The suggested quality evaluation technique is divided into two phases: hardware development and software development. The technology is designed to take a picture of a tomato and automatically transport it to the proper containers without the need for human interaction. The program analyzes the fruit for flaws and freshness using image processing techniques. Several pictures of the tomato fruit were used in the experiments. The proposed approach was shown to be successful in evaluating the quality of the tomato with 96.47 percent accuracy. Another factor that changes the state of the fruit over time is oxygenation. According to Fu (2020), Fruit freshness grading uses computer vision technology to evaluate the texture, color, and form of the fruit. During the decay process, a fruit undergoes progressive changes, such as the development of black spots owing to oxygenation and shrinking due to the loss of contained water.

Tiago et al. (2019) developed an algorithm for analyzing and processing photographs in order to recognize fruits, especially peaches, and determine their dimensions, such as volume and weight. Image segmentation is used in the proposed approach to extract properties such as color and form. To increase the recognition rate of fruits, these attributes were employed to train a classification algorithm using a Support Vector Machine (SVM). The forecast rate of peaches in orchards (808 trees/ha) had an overall precision of 72 percent. This is the first study to apply these concepts to orchard trees with the goal of predicting production along the fruit development. Orchard trees have other important future applications, not just for production prediction, for this type of algorithm. A similar paper made by Nosseir & Ahmed.(2019) stated a novel method that distinguishes between four types of fruits and distinguishes the deteriorated from the fresh. Fine, Medium, Coarse, Cosine, Cubic, and Weighted K-Nearest Neighbors algorithms are used to separate the different varieties of fruits. 96.3 percent, 93.8 percent, 25%, 83.8 percent, 90 percent, and 95 percent, respectively, are the accuracy percentages for each. The linear and quadratic Support Vector Machine (SVM) algorithms differentiated between decaying and fresh fruits based on the color segmentation and texture feature algorithms values of each fruit image to tell them apart. The linear SVM has a 96 percent accuracy and the quadrat-ic SVM has a 98 percent accuracy. Woo & Seyed(2009) paper proposes a method for measuring the distance between unknown fruit qualities and stored fruit instances using the KNN Euclidean distance metric. The algorithms then locate the closest or most similar samples of the unknown fruit. The store fruit example should be varied in color, shape, and size, as well as captured from multiple angles and positions, to ensure that the system is resilient and capable of recognizing the input fruit image. Ninawe & Pandey(2014) used a k-NN classifier; the suggested method can categorize and identify fruit photos. The recognition results are accurate up to 95%. It can recognize six different types of fruits based on their shape, color, size, and texture. It can classify red apples, green bananas, green guavas, green melon, oranges, and watermelons. Zawbaa et al. (2010) used 178 photos of orange, strawberry, and apple in their study. There are 178 fruit photos in the dataset. Their approach is based on unique shape and color characteristics. The k-NN and SVM classifiers were utilized. They compared the photos of the fruits depending on the feature to evaluate the system. The first group compares the differences in color between the photographs of apples and oranges. They have a comparable shape when taken into account. The second group distinguishes between the form feature and tests it using photos of apples and strawberries that are the same color. The third group examines and contrasts the visuals of oranges and strawberries. They are distinct in terms of color and shape. The accuracy of the third group is the best.

It was stated by Yogitha et al. (2014), the fruit pictures are collected in order to evaluate the surface of the fruit in order to extract color, size, and form characteristics. Prior to reaching the field of vision, the fruit on the conveyor is aligned horizontally to its stem and calyx axes. While moving through the field of view, it continues to spin along its stem and calyx axes. A fruit is caught in many picture frames as it passes through the field of vision. The frame grabber card captures each image when a trigger on one of its digital input lines occurs. Each of the fruit pictures depicts a view of the fruit's surface from one of six rotational positions, known as postures, which reflect rotations in multiples of 60 degrees.This picture collecting technique aids in the analysis of the fruit's whole surface. Rocha et al. (2010) developed a system for classifying fruits and vegetables automatically. They used look, color, texture, and shape criteria to classify the photos of 15 different fruits and vegetables. There is a 3% mistake rate in their system. Gao, Dong, Xiao, Yin and He (2016) introduced a different setup on the placement of the fruit in what they called LED-induced fluorescence spectroscopy technique. Apples were placed on an anodized aluminum plate that a step motor could move. In the measurements, LEDs were used to irradiate the apples, and the fluorescence spectra were acquired automatically at eight different sites. The tiny sensor head, which is a truncated cone-shaped cavity constructed of polished aluminum, was fitted with LEDs at 375 nm and 400 nm with maximum intensities of around 20 mW and bandwidths of about 10–20 nm. According to another group of researchers, Team, D. O. (2018), their automated ripeness detection system consists of a conveyor belt that transports the food in containers to a specific sensor in the automated ripeness detection system. The sensor resembles a standard camera, but it is capable of capturing data that is undetectable to the naked eye. By inputting new product variants on a daily basis, they teach the machine what good and bad products look like. The products are photographed and then made available to the machine as data. As a result, the computer gradually learns the quality criteria. “OK,” “Damaged,” “Badly damaged,” and “Expired” are the four classifications of fresh food. The machine is randomly provided with containers of various ripeness grades. Employees, too, are only informed of which products they must fill the full container with via a screen. This is to avoid having “expired” produce in every third container and “OK” fruit in every fourth container. The team researchers really prohibit the computer from learning a pattern, rather than enabling it to learn a sequence instead of doing thorough product tests.

Researchers had already been studying the capabilities of image processing for years. Many shared their perspective and experience in giving a computer the ability to see and recognize things. Object Detection, specifically fruit detection was not left unraveled. A lot of studies show remarkable accuracy on determining the state of the fruit, its physical characteristics and even the disease the fruit might contain. These studies serve a big headstart to our research. We plan on delivering these algorithms to our handheld devices. In that way more people will have access to a better quality of living.

Conceptual Framework of the Study

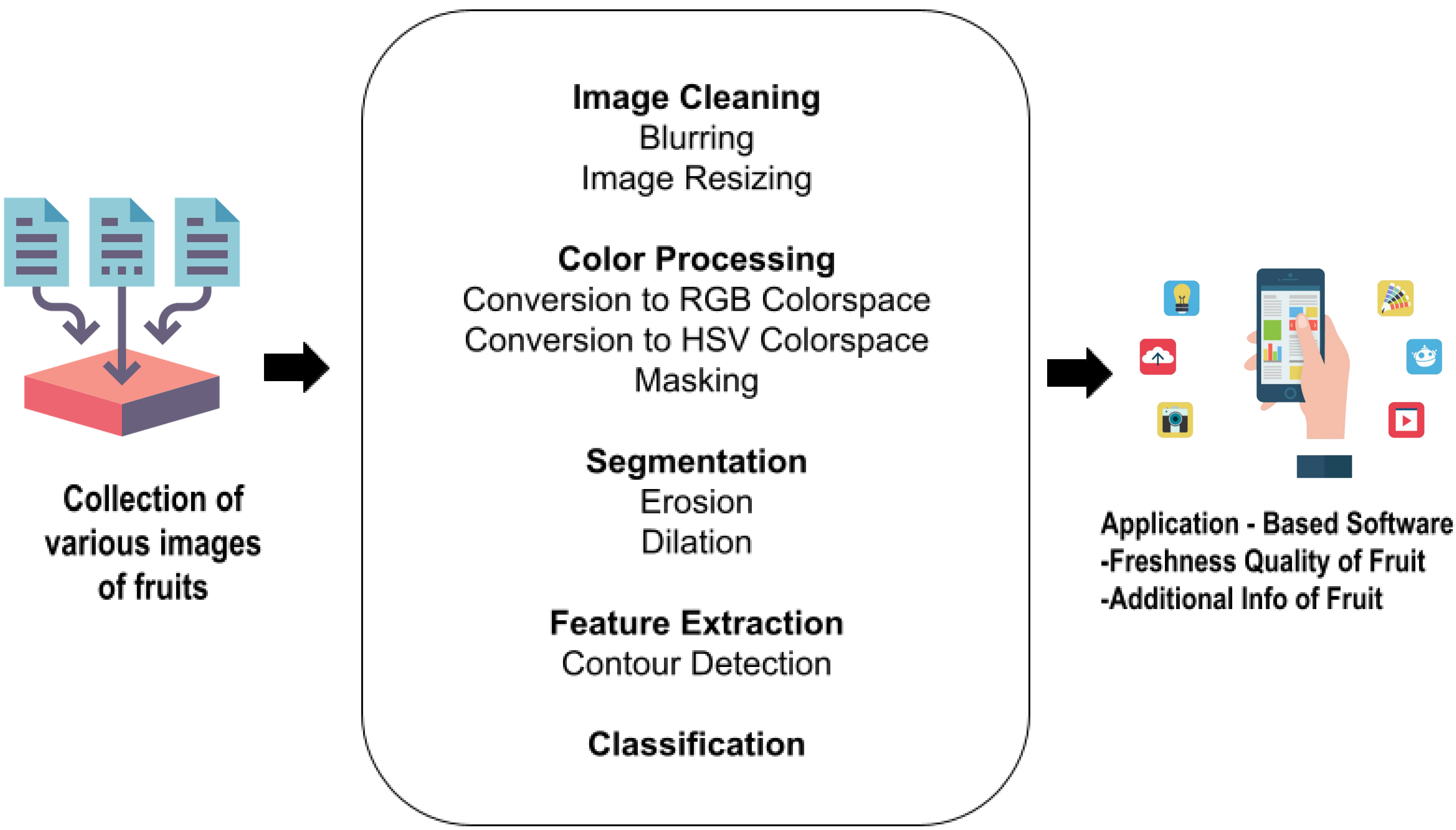


Figure 1

Images of apples, oranges and bananas are collected through any type of digital camera but without using any type of digital filter. The images are then cleaned by removing its background, blurring and resizing in order to keep every data uniform. Next is to convert the images to other color spaces to make the input easier to handle by the algorithms. The inputs are now ready to be processed for segmentation. The image gets stripped off of visible noise. Segmentation isolates the object of interest. Feature extraction process detects all the possible object-like features in the image. Then , the fruit in the image is now classified on what fruit it represents. Lastly, the algorithm will classify the freshness state of the fruit. As output, the system will state the Freshness state of the fruit and additional info about the fruit.

Definition of Terms

Dataset - a collection of fruits, namely papayas, mangoes, and bananas, employed to train the Freshness classification and fruit detection model

Skin Features - external feature of a fruit where it will be used as one of the categories to determine the freshness of a fruit

Filtering Algorithm - image processing algorithms used to clean, convert and transform dataset images into more suitable input.

Machine Learning Algorithms - refers to the K-Nearest Neighbor, Support Vector Machine, and Random Forest algorithms that will be utilized to classify and detect the freshness of the fruits.

Evaluation Metrics - pertains to the algorithms' accuracy, precision, recall, and f-1 score evaluation.

Classification Model - referred to the Freshness Classification and Fruit Detection model.

Freshness Classifier - assesses the freshness of the fruit based on the input.

Fruit Detection - distinguishes the kind of fruit based on the input.

Freshness - Used as the basis to determine the state of a fruit if it is Fresh or Stale.

Nutritional Facts - The nutrition facts label is a label that shows what nutrients and other ingredients are in the food.

CHAPTER 3

METHODOLOGY

Research Methodology

This chapter describes all the methods used by the researchers. The researchers will use an approach that best suits the efficiency of development. The researchers will follow the flow of the created diagrams.

Combined Methodology of Rapid Application Development and Image Processing WorkFlow

The researchers will be using a combined methodology of Rapid Application Development and modified Image Processing Workflow at document level in order to achieve this research's goal of software development and data analysis.

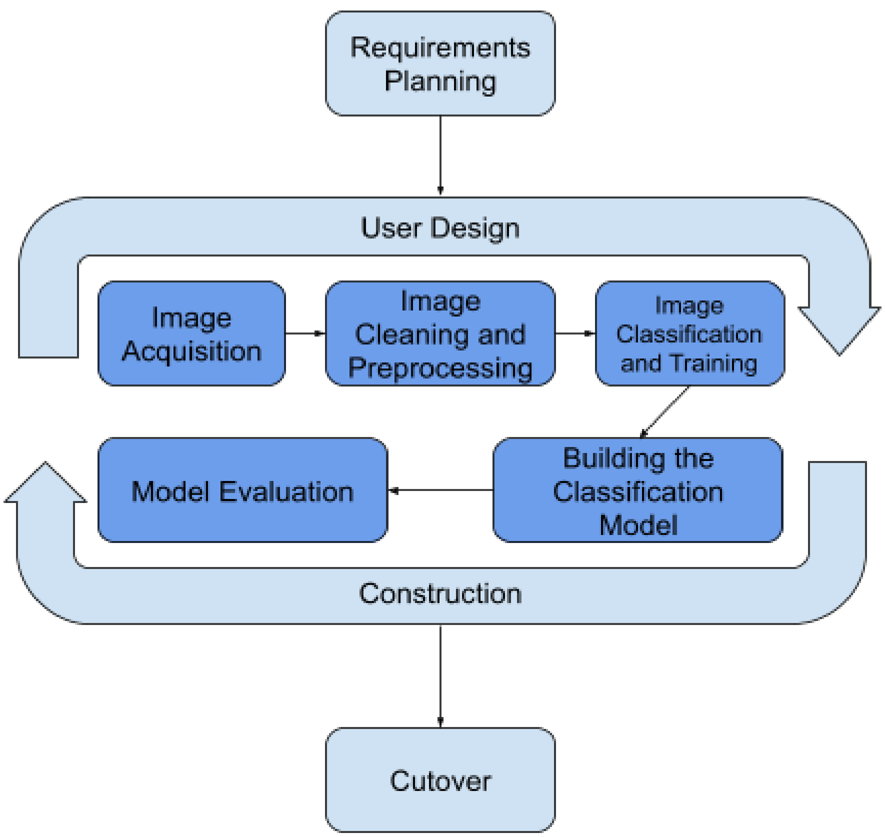


Figure 2

Phase 1: Requirements Planning

The first phase of Rapid Application Development (RAD) methodology contains the tasks of the researchers to imagine and picture the road map of the study.

The researchers set and agreed upon the goal as well as the specific objectives of the research. The researchers discussed what type of fruits will be collected. The researchers also agreed on the content of the image as well as how the fruit is positioned in the image to keep the dataset uniform.

The researchers will be using Rapid Application Development along with Image Processing as methodology for the study. The software that will be developed will contain several modules that will be distributed evenly to all the researchers. The completion of each module will be updated to every other member of the group in order to guarantee the smooth transition of the process in development.

A Gantt chart will be used to monitor the timeline and keep track of each development. The Gantt Chart will provide the timeline of work from documentation to development and until the testing period.(See Appendix A).

An Activity Diagram is created to show the basic flow of data of our proposed classification model. The Diagram shows the main parts of our model which starts from the image selection going to Fruit detection then Freshness Classification and ends in the output by displaying the acquired result. (See Appendix B).

A more detailed diagram is created using a Flow Chart. The Flow Chart also consists of the main modules from the Activity Diagram but shows more info on where the data goes before and after reaching another module. (See Appendix C).

A Use Case Diagram is also created in order to show the connection of the user to the system. Entities connected to the user show the parts that the user can only see when using the tool. While disconnected parts are shown to be a background process or task that is done by the computer.(See Appendix D)

Phase 2: User Design

In this phase, the researchers will begin the development of the prototype that will be used to show the user interface of the application. The researchers will first make an image blueprint of the user interface for the application using Adobe Photoshop Software (See Appendix E). The researchers will then develop a Fruit Detection prototype using opencv-python. The prototype will be able to detect the type of fruit shown in the picture and display its prediction in the console of the pycharm environment. (See Appendix F) One member of the research team will be responsible for making the user interface. The application is composed of two interfaces, the image selection interface and the content interface.

Phase 3: Construction

The construction phase takes the prototype from the design phase and converts them into a working model.The researchers prepare the tool to be used in developing the application as well as the collected data to be manipulated along the process. This phase oversees the entire process, from data collection to coding, and finishes a fully functional Android application.

1. **Image Acquisition**

The data to be collected are images of apples, oranges and bananas. The images will be collected from kaggle.com. The collected images will be categorized into three groups which are Apples, Bananas and Oranges. Each group will then be again grouped into two categories, One for “fresh” and the other for “stale”.

2. **Image Cleaning and Preprocessing**

The collected images should be cleaned and processed in order for the classifier to assess. Image cleaning and processing is important in order to remove the noise and unnecessary artifacts in the image. The researchers will use traditional opencv image processing algorithms to clean the image and isolate the fruit and its physical features.

3. **Image Classification and Training**

In training, the dataset will be divided into two sets, data for training and data for testing with a 70%-30% ratio respectively. The data processed will be used to train the classifiers in recognizing a fruit. Along with it is the classification on the freshness state of the recognized fruit, the classification will be based on its skin features and other factors that are present on the processed data that will be used to determine a fruit's state. Classifiers will be tested and modified along the process of training with the data. The output will be expected to sort the fruit images to their respective classifications. The classifiers are made in adaptation to the K-Nearest Neighbor Algorithm, Support Vector Machine and Random Forest. A more detailed process of the KNN algorithm started started by vectorizing the test image as well as each image in the dataset. The images will now be in the form of a 1D array of pixels. Our next step is to compute for the Euclidean Distance of the test image against each image in the dataset. The Euclidean Distance gathered from each image in the dataset is now subjected for voting following the value of K. The image with the most votes will determine the type of fruit of the one in the test image.(See Appendix G)

3.1 K-Nearest Neighbor

The K-Nearest Neighbor is one of the simplest supervised Machine Learning algorithms mostly used for classification (Figure 3). The KNN algorithm assumes that similar things exist in close proximity. Things at close proximity will be responsible for making a voting process. A K is determined before to estimate the prediction of the algorithm. K is a parameter that refers to the number of nearest neighbors to include in the majority voting process.

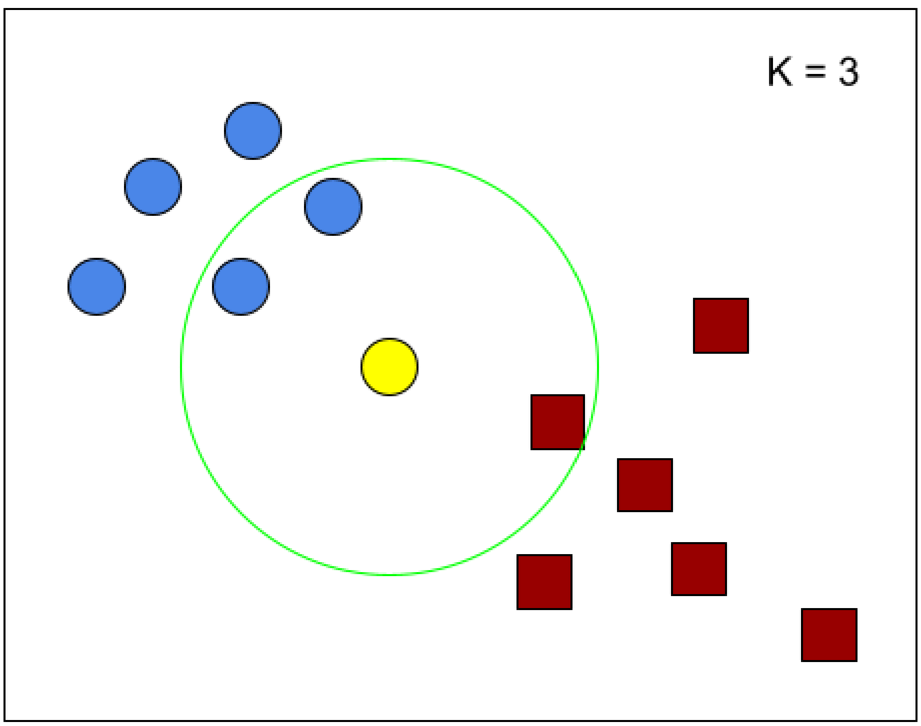


Figure 3

Choosing the value of K

Larger K value takes a lot of computation time and resources due to multiple verification of each node while with smaller K value, the algorithm may capture training data as well as all the noise concluding into a bias result. K is computed by Sqrt(n), where n is the total number of datapoints. After that is choosing an odd value of K which can avoid the confusion between two classes of data.

3.2 Support Vector Machine

The SVM or the Support Vector Machine algorithm is a simple classification algorithm which is observed to be performing best on 2 group classifications.The main advantages of SVM are higher speed and better performance with a limited number of samples. This makes the algorithm very suitable for text classification problems, where it's common to have access to a dataset of at most a couple of thousands of tagged samples.An SVM can be explained by a graphical representation of 2-group classification problem. The left side consists of data points for one of the classes which we can call here as "circles" while on the right side it consists of "squares". The main task of the SVM algorithms is to create the best hyper plane that divides the two classes resulting in the widest margin. The margin on the other hand is determined as the distance between the hyper plan and the nearest data points.(Figure 4)

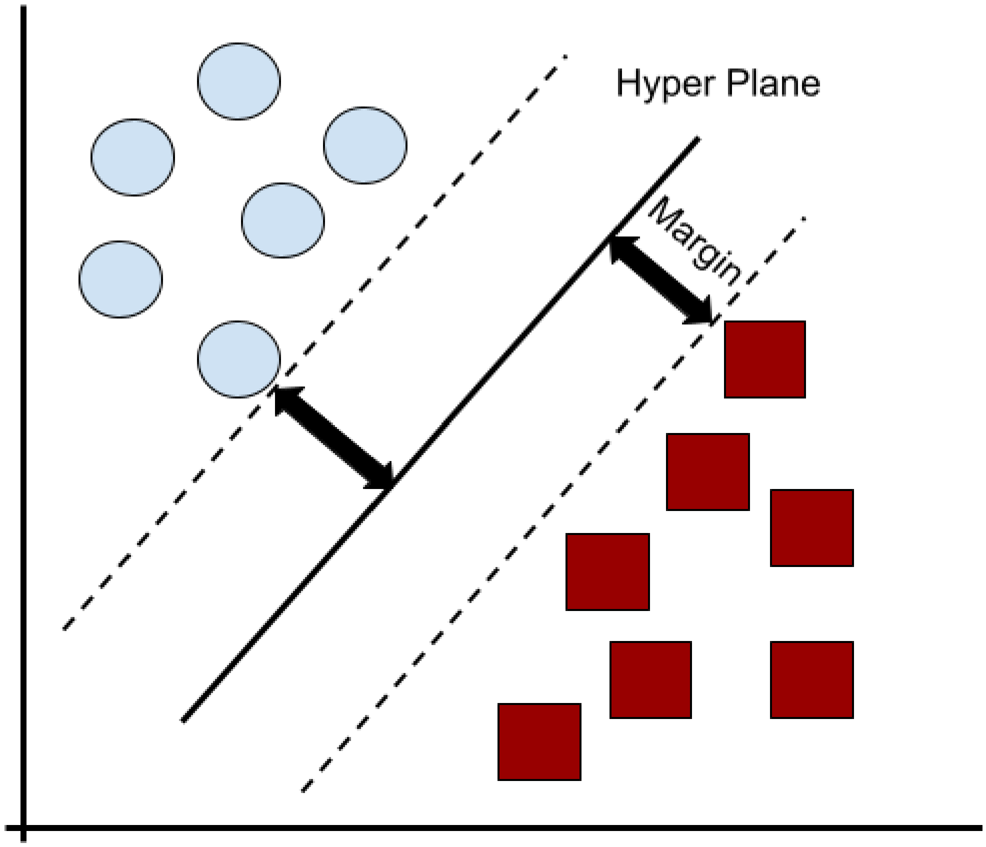


Figure 4

3.3 Random Forest

A Random Forest or Decision Trees is one of most frequently and widely used supervised machine learning algorithms that can perform both regression and classification tasks. The intuition behind the decision tree algorithm is simple, yet also very powerful. For each datapoint in the dataset, the decision tree creates a new node where the most important attribute is placed at the root node. For evaluation we start at the root node and work our way down the tree by following the corresponding node that meets our condition or "decision". This process continues until a leaf node is reached, which contains the prediction or the outcome of the decision tree. (Figure 5)

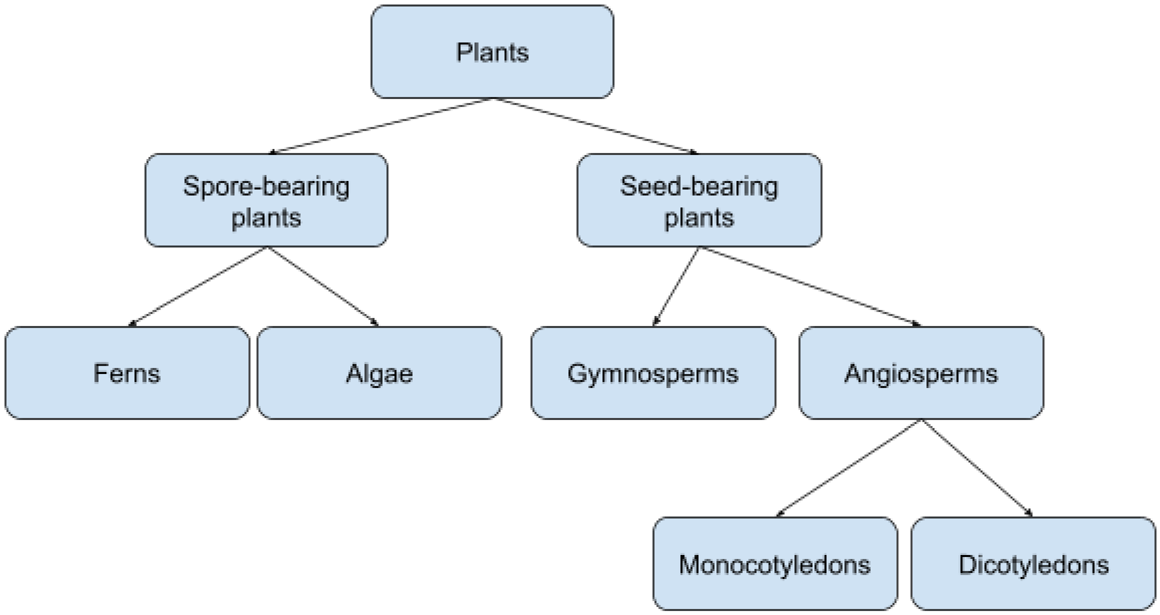


Figure 5

4. **Building the Classification Model**

To build the entire model, two sets of classifiers will be necessary. One for fruit detection and another for freshness state detection.Each classifier will be trained and tested individually. The trained classifiers will then be compiled into a classification model where it will be trained again. All the testing done will be utilizing the confusion matrix.

5. **Model Evaluation**

Model Evaluation is the final step in the image processing workflow. The model will be evaluated for the last time using the evaluation metrics. The results of the three algorithms are compared. The algorithm with the highest test result will be used to be implemented into the android platform. The next step is to be transferred and adopted into an android application where it will be using the android device as a medium.

5.1 Confusion Matrix

A confusion matrix is a table that is often used to describe the performance of a classification model or a classifier on a set of test data in which the true values are known. The confusion matrix itself is relatively simple to understand, but the related terminology can be confusing. The result of a training data rendered on a confusion matrix will result in four classification outputs which are represented by a table. Figure 4 represents an example of a tally for predicting true or false in a value of 10 data points.

N = 10	Predicted : False	Predicted : True
Expected: False	3	1
Expected: True	2	4

Figure 6

true positives (TP): These are cases in which we predicted true and is actually true.

true negatives (TN): These are cases in which we predicted false and is actually false

false positives (FP): These are cases in which we predicted true but should actually be false.

false negatives (FN): These are cases in which we predicted false but should actually be true.

There are various rates that can be computed using a confusion matrix for a binary classifier.

Accuracy - how often the classifier predicts a correct result

$(True\ Positives\ +\ True\ Negatives)\ /\ Total$
$(4\ +\ 3)\ /\ 10\ =\ 0.7$

Miss classification rate - how often the classifier makes a mistake of prediction

$(False\ Positives\ +\ False\ Negative)\ /\ Total$
$(1\ +\ 2)\ /\ 10\ =\ 0.3$

Phase 4: Cutover

After the completion of the construction phase, the developed application will now be tested and used in an actual environment. The researchers will have a full scale testing on the overall accuracy of the fruit detection, on whether it can still detect a fruit that is affected by different environmental factors. This last phase is the end of the development but the developers as well as other users will continue to look for bugs and modifications in order to further improve the accuracy and stability of the application.

Ethical consideration

Ethical considerations are of utmost importance when it comes to research. Data gathering procedures can be affected when ethical considerations are not established for a research. This study will observe confidentiality. The participants who provided the product to gather data on will not be explicitly mentioned in this research. This will ensure that the participant’s integrity and reputation will not be compromised and to respect their privacy. Before beginning the data collection process, the participants' full consent will be acquired to formally start the data gathering. The researchers will not disclose any personal information about the participants to ensure and protect their privacy. The data gathered will only be used for research purposes. Any communication between participants and researchers will be honest and transparent, which will be crucial for objective results and discussions. The researcher will avoid false information and primary data findings. Disclosing the data findings will stay confidential, the researchers will ensure that the data findings are sent to the designated participants.

REFERENCES

A. M. (2014). THE QUALITY IDENTIFICATION OF FRUITS IN IMAGE PROCESSING USING MATLAB. International Journal of Research in Engineering and Technology, 03(22), 92–95. https://doi.org/10.15623/ijret.2014.0322019

Almadhor, Ahmad, Hafiz T. Rauf, Muhammad I.U. Lali, Robertas Damaševičius, Bader Alouffi, and Abdullah Alharbi. (2021). "AI-Driven Framework for Recognition of Guava Plant Diseases through Machine Learning from DSLR Camera Sensor Based High Resolution Imagery" Sensors 21, no. 11: 3830.

Ananthanarayana, T., Ptucha, R., & Kelly, S. C. (2020). Deep Learning based Fruit Freshness Classification and Detection with CMOS Image sensors and Edge processors. Society for Imaging Science and Technology. doi.org/10.2352/ISSN.2470-1173.2020.12.FAIS-172

Arakeri, M. P., & Lakshmana. (2016, January 1). Computer Vision Based Fruit Grading System for Quality Evaluation of Tomato in Agriculture industry. ScienceDirect.426-433.doi.org/10.1016/j.procs.2016.03.055

A. Rocha, A., D.Hauagge, J.Wainer, J., and S. Goldenstein.(2010). Automatic Fruit and Vegetable Classification from Images. Computers and Electronics in Agriculture 70:1. pp.96–104. https://doi.org/10.1016/j.compag.2009.09.002

Ayan, O. , Demirez, D. Z. , Kiziloz, H. K. , Inci, G. , Isleyen, S. & Ergin, S. (2018). The Detection of Spoiled Fruits on a Conveyor Belt Using Image Processing Techniques and OPC Server Software . International Journal of Computational and Experimental Science and Engineering , 4 (1) , 11-15 . DOI: 10.22399/ijcesen.398335

Bhargava, A., & Bansal, A. (2018). Fruits and vegetables quality evaluation using computer vision: A review. Journal of King Saud University - Computer and Information Sciences. doi:10.1016/j.jksuci.2018.06.002

Bin, É. (2020, December 5). Random Forest and Decision Trees - Towards Data Science. Medium. https://towardsdatascience.com/random-forest-and-decision-trees-d8f7cd06803b

Chandini, A. A., & B, U. M. (2018, September 1). Improved Quality Detection Technique for Fruits Using GLCM and MultiClass SVM. IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/8554876

Choudhary, P., Khandekar, R., Borkar, A., & Chotaliya, P. (2017). IMAGE PROCESSING ALGORITHM FOR FRUIT IDENTIFICATION. Volume: 04 Issue: 03

Cubero S., Aleixos N., Francisco A., Torregrosa A., Ortiz C., García-Navarrete O. L., Blasco J. (2014) , Optimised computer vision system for automatic pre-grading of citrus fruit in the field using a mobile platform

Dubey, S. & Jalal, A. (2015). Application of Image Processing in Fruit and Vegetable Analysis: A Review. Journal of Intelligent Systems, 24(4), 405-424.

F. Pallottino, C. Costa, F. Antonucci, P. Menesatti (2013), SWEET CHERRY FRESHNESS EVALUATION THROUGH COLORIMETRIC AND MORPHOMETRIC STEM ANALYSIS: TWO REFRIGERATION SYSTEMS COMPARED, Acta Alimentaria, Vol. 42 (3), pp. 428–436 (2013)

Febrian Valentino et al. (2021) IOP Conf. Ser.: Earth Environ. Sci. 794 012110

FoodPrint. (2021, June 11), Food Waste Is a Massive Problem—Here's Why. https://foodprint.org/issues/the-problem-of-food-waste/

FreshPlaza. Retrieved September 10, 2021, 90% of young people throw out fruit and veg.(n.d.).from https://www.freshplaza.com/article/2183962/90-of-young-people-throw-out-fruit-and-veg

Fu, Y. (2020.). Fruit Freshness Grading Using Deep Learning.

Gao, F., Dong, Y., Xiao, W., Yin, B., Yan, C., & He, S. (2016, September 1). LED-induced fluorescence spectroscopy technique for apple freshness and quality detection. ScienceDirect.https://www.sciencedirect.com/science/article/pii/S0925521416300795

Gastélum-Barrios, A., Bórquez-López, R., Rico-García, E., Toledano-Ayala, M., & Soto-Zarazúa*, G. (2011). Tomato quality evaluation with image processing: A review (14th ed., Vol. 6). Academic Journals. https://doi.org/10.5897/AJAR11.108

H. M. Zawbaa, M.Abbass, M.Hazman, and A. Hassenian.(2014).Automatic Fruit Image Recognition System Based on Shape and Colour. AMLTA Springer International Publish-ing Switzerland. 488, pp. 278–290.

HUANG, T. S., SCHREIBER, W. F., & TRETIAK, O. J. (2002). Image Processing. Series in Machine Perception and Artificial Intelligence, 367–390. doi:10.1142/9789812776952_0015

Ismail, N., & Malik, O. A. (2021, February 11). Real-time visual inspection system for grading fruits using computer vision and deep learning techniques. ScienceDirect. doi.org/10.1016/j.inpa.2021.01.005

Israr Hussain, Qianhua He, Zhuliang Chen (2018) AUTOMATIC FRUIT RECOGNITION BASED ON DCNN FOR COMMERCIAL SOURCE TRACE SYSTEM

José Naranjo-Torres, Marco Mora, Ruber Hernández-García, Ricardo J. Barrientos, Claudio Fredes, Andres Valenzuela(2020), A Review of Convolutional Neural Network Applied to Fruit Image Processing , Vol 10, Iss 3443, p 3443

Kader, A. A. (2008). Flavor quality of fruits and vegetables. Journal of the Science of Food and Agriculture, 88(11), 1863–1868. doi:10.1002/jsfa.3293

Kamil, I. A., & Fagbadegun, A. (2017). Fruit identification employing computer vision based on artificial neural network backpropagation algorithms | Journal of Applied Science, Engineering and Technology. AJOL. https://www.ajol.info/index.php/jaset/article/view/184808

Kang, J., & Gwak, J. (2021, August 24). Ensemble of multi-task deep convolutional. . . Multimedia Tools and Applications. https://link.springer.com/article/10.1007/s11042-021-11282-4?error=cookies_not_supported&code=f52c1bfc-dc2d-4540-a507-823cd70b5550

Koirala, A., Walsh, K. B., Wang, Z., & McCarthy, C. (2019). Deep learning - method overview and review of use for fruit detection and yield estimation - USQ ePrints. Eprints. https://eprints.usq.edu.au/36618/

Lakare S. A., & Kapale N.D.(2019)Automatic Fruit Quality Detection System.IRJET.6(6)

Li, Y., Feng, X., Liu, Y., & Han, X. (2021). Apple quality identification and classification by image processing based on convolutional neural networks. Scientific Reports, 11(1). doi:10.1038/s41598-021-96103-2

Linke, M., Herppich, W. B., & Geyer, M. (2010). Green peduncles may indicate postharvest freshness of sweet cherries. Postharvest Biology and Technology, 58(2), 135–141. doi:10.1016/j.postharvbio.2010.05.014

M. Sarkar, N. Gupta, and M. Assaad, (2019) "Monitoring of fruit freshness using phase information in polarization reflectance spectroscopy," Appl. Opt. 58, 6396-6405

McAndrew, A. (2004). Introduction to Digital Image Processing with MATLAB (1st ed.). Course Technology.

Move For Hunger. Retrieved September15,2021, The Environmental Impact of Food Waste.(n.d.).from https://moveforhunger.org/the-environmental-impact-of-food-waste

Mudaliar, G., & Priyadarshini, R. B. K. (2021, May). A MACHINE LEARNING APPROACH FOR PREDICTING FRUIT FRESHNESS CLASSIFICATION.

Mureşan Horea, Oltean Mihai (2018) in Fruit recognition from images using deep learning Universitatis Sapientiae: Informatica, Vol 10, Iss 1, Pp 26-42 (2018)

Musacchi, S., & Serra, S. (2018, April 14). Apple fruit quality: Overview on pre-harvest factors. ScienceDirect. https://www.sciencedirect.com/science/article/pii/S030442381730780X

Mustaffa, I. B., & Khairul, S. F. B. M. (2017, November 1). Identification of fruit size and maturity through fruit images using OpenCV-Python and Rasberry Pi. IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/8308068

Nadia Kasim, Siti Nazifah Zainol Abidin, Nur Nabilah Abu Mangshor, Hazwa Hanim Mohamed Hamzah, Nurul Zahirah Abd Rahim (2019) STAGE OF MATURITY BANANA FRUIT CLASSIFICATION USING IMAGE PROCESSING Multidisciplinary Informatics Journal Vol. 2, No. 1, Pg. 82 – 90

Narendra V. G., & Pinto A. J.DEFECTS DETECTION IN FRUITS AND VEGETABLES USING IMAGE PROCESSING AND SOFT COMPUTING TECHNIQUES.Proceedings of 6th International Conference on Harmony Search, Soft Computing and Applications (pp.325-337)doi:10.1007/978-981-15-8603-3_29

Nayak A. N.,& Manjesh R., Danusha. (2019). Fruit Recognition using Image Processing.IJERT. (VOLUME 7 - ISSUE 08)

Nordey, T., Davrieux, F., & Léchaudel, M. (2019, July 1). Predictions of fruit shelf life and quality after ripening: Are quality traits measured at harvest reliable indicators? ScienceDirect.https://www.sciencedirect.com/science/article/pii/S0925521418306021

Nosseir, A., & Ahmed, S. E. . A. (2019). Automatic Classification for Fruits' Types and Identification of Rotten Ones Using k-NN and SVM. International Journal of Online and Biomedical Engineering (iJOE), 15(03), pp. 47–61. https://doi.org/10.3991/ijoe.v15i03.9832

Patil, S. V., Jadhav, V. M., Dalvi, K. M., & Kulkarni, B. P. (2020). FRUIT QUALITY DETECTION USING OPENCV/PYTHON. IRJET. (Volume: 07 Issue: 05). https://www.irjet.net/archives/V7/i5/IRJET-V7I51254.pdf

Péneau, S., Hoehn, E., Roth, H.-R., Escher, F., & Nuessli, J. (2006). Importance and consumer perception of freshness of apples. Food Quality and Preference, 17(1-2), 9–19. doi:10.1016/j.foodqual.2005.05.002

Péneau, S., Linke, A., Escher, F., & Nuessli, J. (2009, March 21). Freshness of fruits and vegetables: consumer language and perception - Zurich Open Repository and Archive. Emerald. https://www.zora.uzh.ch/id/eprint/24811/

Sa, I., Ge, Z., & Dayoub, F. (2016). DeepFruits: A Fruit Detection System Using Deep Neural Networks. MDPI. https://www.mdpi.com/1424-8220/16/8/1222

Sanchez, R. T.,Zafra, M. T. M., Castillejo N.,Frutos A. G., & Hernandez F. A. (2020) Real-Time Monitoring System for Shelf Life Estimation of Fruit and Vegetables.Sensors 2020, 20(7).doi:10.3390/s20071860

Silapeux, A. K. G., Ponka, R., Frazzoli, C., & Fokou, E. (2021, January 20). Waste of Fresh Fruits in Yaoundé, Cameroon: Challenges for Retailers and Impacts on Consumer Health. MDPI. https://www.mdpi.com/2077-0472/11/2/89

Siswantoro, Joko and Arwoko, Heru and Widiasri, Monica (2019) Image based Indonesian fruit recognition using MPEG-7 color structure descriptor and k-nearest neighbor. Proceedings Book Enhancing Engineering Innovation Towards A Greener Future: InCITE 2019 International Conference on Informatics, Technology, and Engineering 22-23 August 2019. D84-D90. ISSN 2686-5955

StudyFinds.org. (2021, August 24). Rotten eggs: Americans end up throwing out $1,500 in wasted food every year. Study Finds. https://www.studyfinds.org/americans-food-waste-1500-every-year/

Team, D. O. (2018, May 30). Machine Learning: Using algorithms to sort fruit. EU About Amazon. https://www.aboutamazon.eu/innovation/machine-learning-using-algorithms-to-sort-fruit

Telang S. S., & Shirsath S. M.(2017)Fruit Quality Management using Image Processing.727 –733. CIIIME.5.(6)

Tiago M. Pereira, Pedro Dinis Gaspar, Maria Paula Simões. (2019).Fruit recognition and classification based on SVM method for production prediction of peaches

Valentino, F., Tjeng Wawan Cenggoro, T. W.,Pardamean B. (2020). A Design of Deep Learning Experimentation for Fruit Freshness Detection.IOP Publishing doi:10.1088/1755-1315/794/1/012110

Vincente, A. R., Manganaris, G. A., Ortiz, C. M., Sozzi, G. O., & Crisosto, C. H. (2014). Nutritional Quality of Fruits and Vegetables. Postharvest Handling, 69–122. doi:10.1016/b978-0-12-408137-6.00005-3

Woo Chaw Sang and Seyed Hadi Mirisaee, “A New Method for Fruit Recognition System”, MNCC Transaction on ICT, Vol1- No 1, June 2009.

Yogesh, Dubey, A. K., Ratan, R., & Rocha, A. (2019). Computer vision based analysis and detection of defects in fruits causes due to nutrients deficiency. Cluster Computing. doi:10.1007/s10586-019-03029-6

Yogitha, S., & Sakthivel, P. (2014, March 1). A distributed computer machine vision system for automated inspection and grading of fruits. IEEE Conference Publication| IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/6922281 .

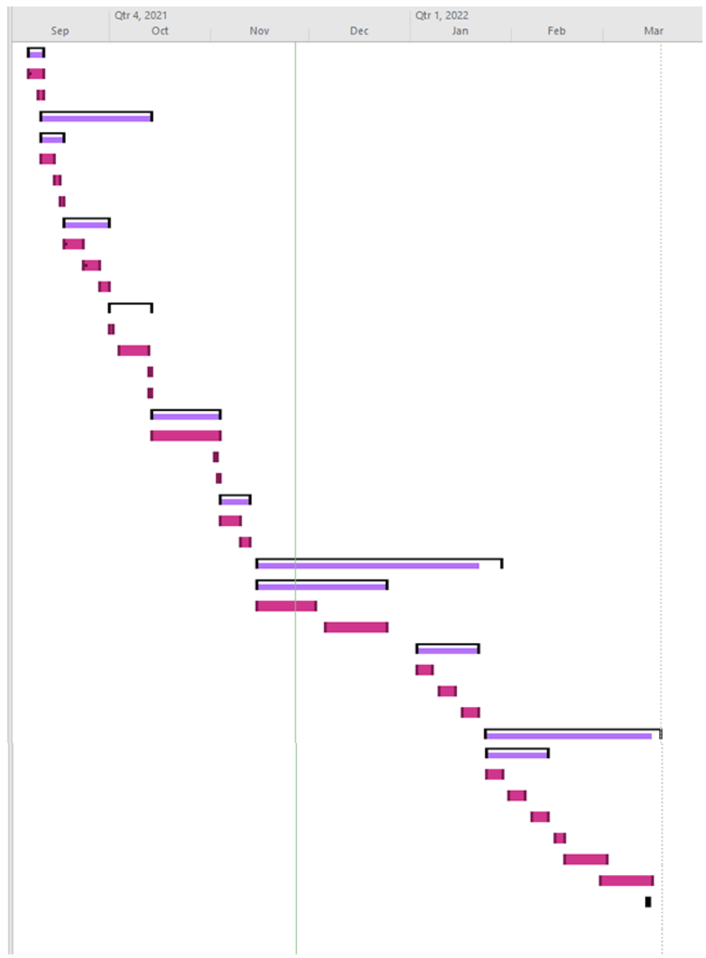
Zawbaa, Hossam M.; Hazman, Maryam; Abbass, Mona; Hassanien, Aboul Ella (2014). [IEEE 2014 14th International Conference on Hybrid Intelligent Systems (HIS) - Kuwait, Kuwait (2014.12.14-2014.12.16)] 2014 14th International Conference on Hybrid Intelligent Systems - Automatic fruit classification using random forest algorithm. , (), 164–168. doi:10.1109/HIS.2014.7086191

APPENDICES

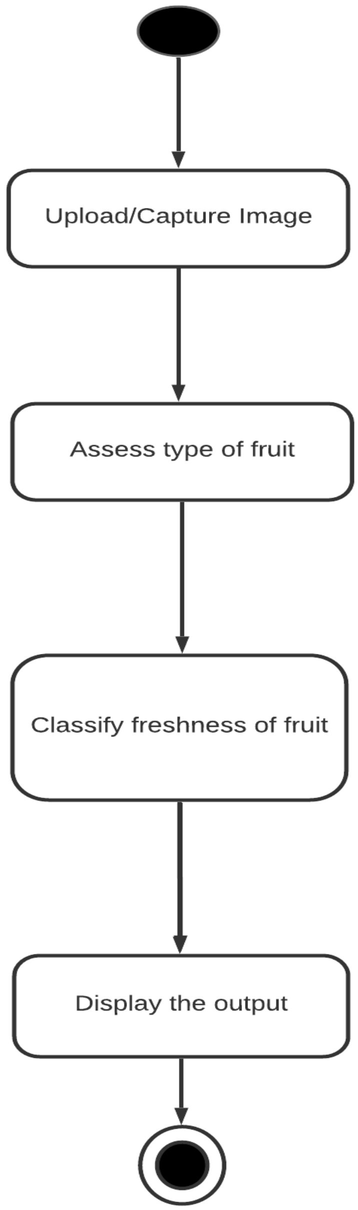
1. Appendix A: Gantt Chart

GANTT CHART

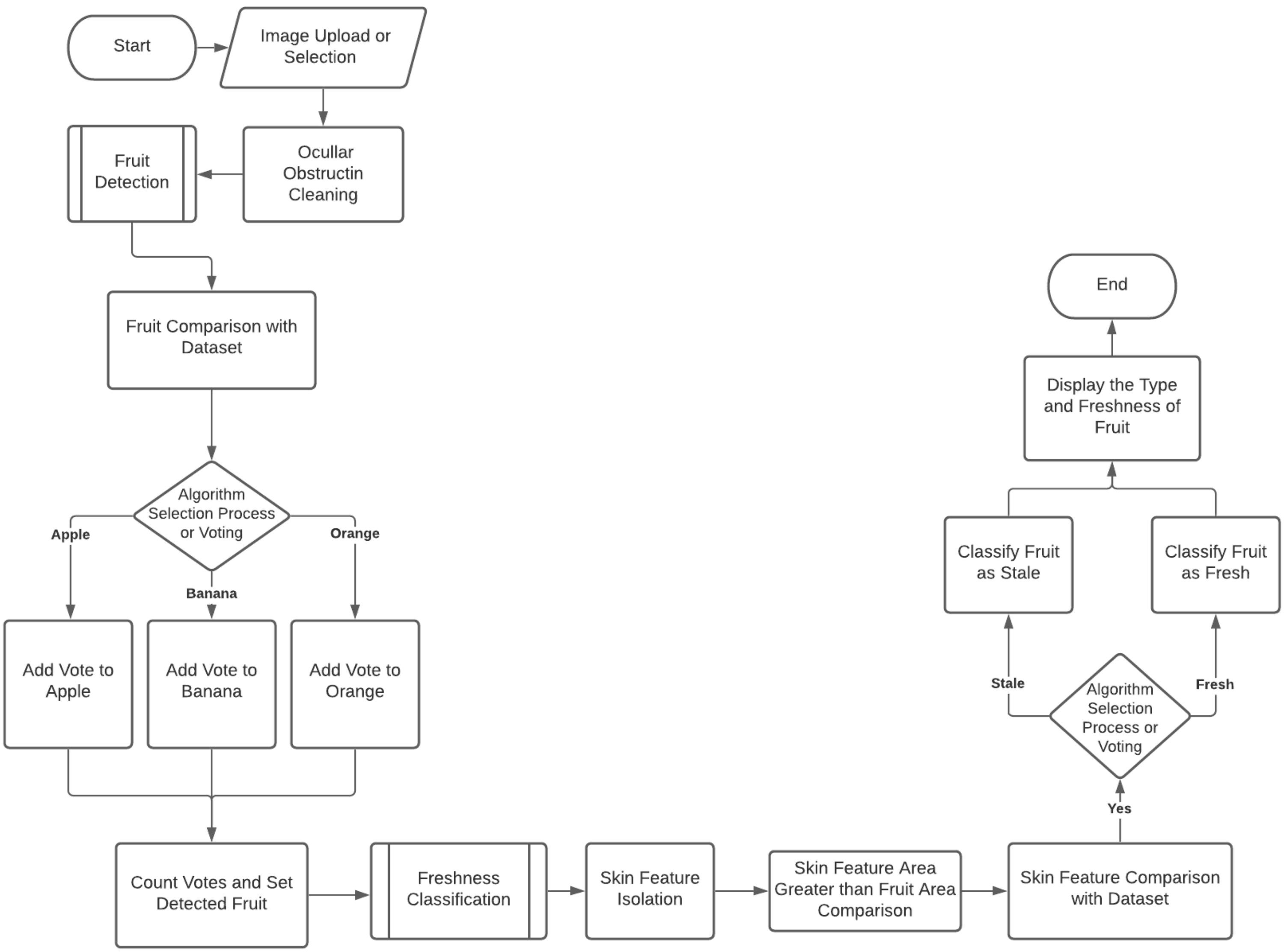
	WBS	Task Name	Duration	Start	Finish
1	1	Planning	5 days	Mon 9/6/21	Fri 9/10/21
2	1.1	Selection of Title	5 days	Mon 9/6/21	Fri 9/10/21
3	1.2	Title Confirmation	2 days	Thu 9/9/21	Fri 9/10/21
4	2	Documentation	24 days	Fri 9/10/21	Wed 10/13/21
5	2.1	Chapter 1	5 days	Fri 9/10/21	Thu 9/16/21
6	2.1.1	Background of the Study	2 days	Fri 9/10/21	Mon 9/13/21
7	2.1.2	Objectives of the Study	2 days	Tue 9/14/21	Wed 9/15/21
8	2.1.3	Significance of the Study	1 day	Thu 9/16/21	Thu 9/16/21
9	2.2	Chapter 2	10 days	Fri 9/17/21	Thu 9/30/21
10	2.2.1	RRL Matrix	4 days	Fri 9/17/21	Wed 9/22/21
11	2.2.2	Review of Related Literature	3 days	Thu 9/23/21	Mon 9/27/21
12	2.2.3	Conceptual Framework	3 days	Tue 9/28/21	Thu 9/30/21
13	2.3	Chapter 3	9 days	Fri 10/1/21	Wed 10/13/21
14	2.3.1	Rapid Application Development	1 day	Fri 10/1/21	Fri 10/1/21
15	2.3.2	Methodology Phases	7 days	Mon 10/4/21	Tue 10/12/21
16	2.3.3	Ethical Consideration	1 day	Wed 10/13/21	Wed 10/13/21
17	2.3.4	References	1 day	Wed 10/13/21	Wed 10/13/21
18	3	Adviser Consultation	9 days	Thu 10/14/21	Wed 11/3/21
19	3.1	Contacting Advisers	9 days	Thu 10/14/21	Wed 11/3/21
20	3.2	Programming Adviser Meeting	1 day	Tue 11/2/21	Tue 11/2/21
21	3.3	Content Adviser Meeting	1 day	Wed 11/3/21	Wed 11/3/21
22	4	Requirements Planning	7 days	Thu 11/4/21	Fri 11/12/21
23	4.1	Creation of Digrams and Timetables	4 days	Thu 11/4/21	Tue 11/9/21
24	4.2	Collection and Sorting of Datasets	3 days	Wed 11/10/21	Fri 11/12/21
25	5	User Design (Iterative Phase)	55 days	Mon 11/15/21	Fri 1/28/22
26	5.1	Construct the Protoype	30 days	Mon 11/15/21	Fri 12/24/21
27	5.1.1	Fruit Detection Prototype	14 days	Mon 11/15/21	Thu 12/2/21
28	5.1.2	Freshness Classification Prototype	15 days	Mon 12/6/21	Fri 12/24/21
29	5.2	Test the Prototype	15 days	Mon 1/3/22	Fri 1/21/22
30	5.2.1	Partial Dataset Testing	5 days	Mon 1/3/22	Fri 1/7/22
31	4.2.2	Pre-Plan Modifications	5 days	Mon 1/10/22	Fri 1/14/22
32	5.2.3	Refine Prototype	5 days	Mon 1/17/22	Fri 1/21/22
33	6	Construction	40 days	Mon 1/24/22	Fri 3/18/22
34	6.1	Finalize Code Stucture	15 days	Mon 1/24/22	Fri 2/11/22
35	6.1.1	Image Preprocessing	5 days	Mon 1/24/22	Fri 1/28/22
36	6.1.2	Image Classification	5 days	Mon 1/31/22	Fri 2/4/22
37	6.1.3	Building the Classification Model	5 days	Mon 2/7/22	Fri 2/11/22
38	6.2	System Testing and Debugging	3 days	Mon 2/14/22	Wed 2/16/22
39	6.3	Online Implementation	9 days	Thu 2/17/22	Tue 3/1/22
40	6.4	Application Development	12 days	Mon 2/28/22	Tue 3/15/22
41	6	Cutover	1 day	Mon 3/14/22	Mon 3/14/22
42	6.1	System Revision			
43	6.2	Monitoring and Support			



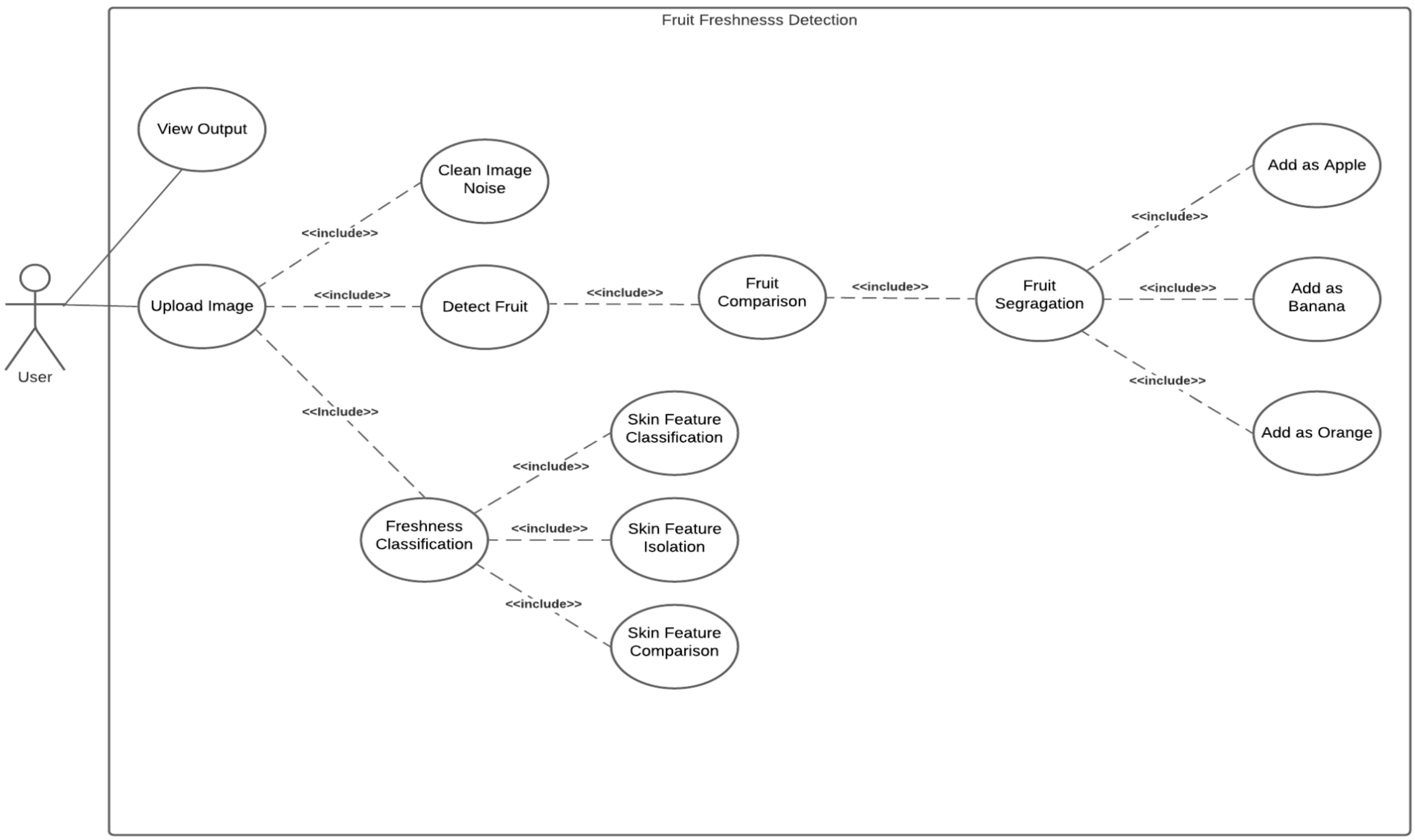
2. Appendix B: Activity Diagram



3. Appendix C: Flow Chart

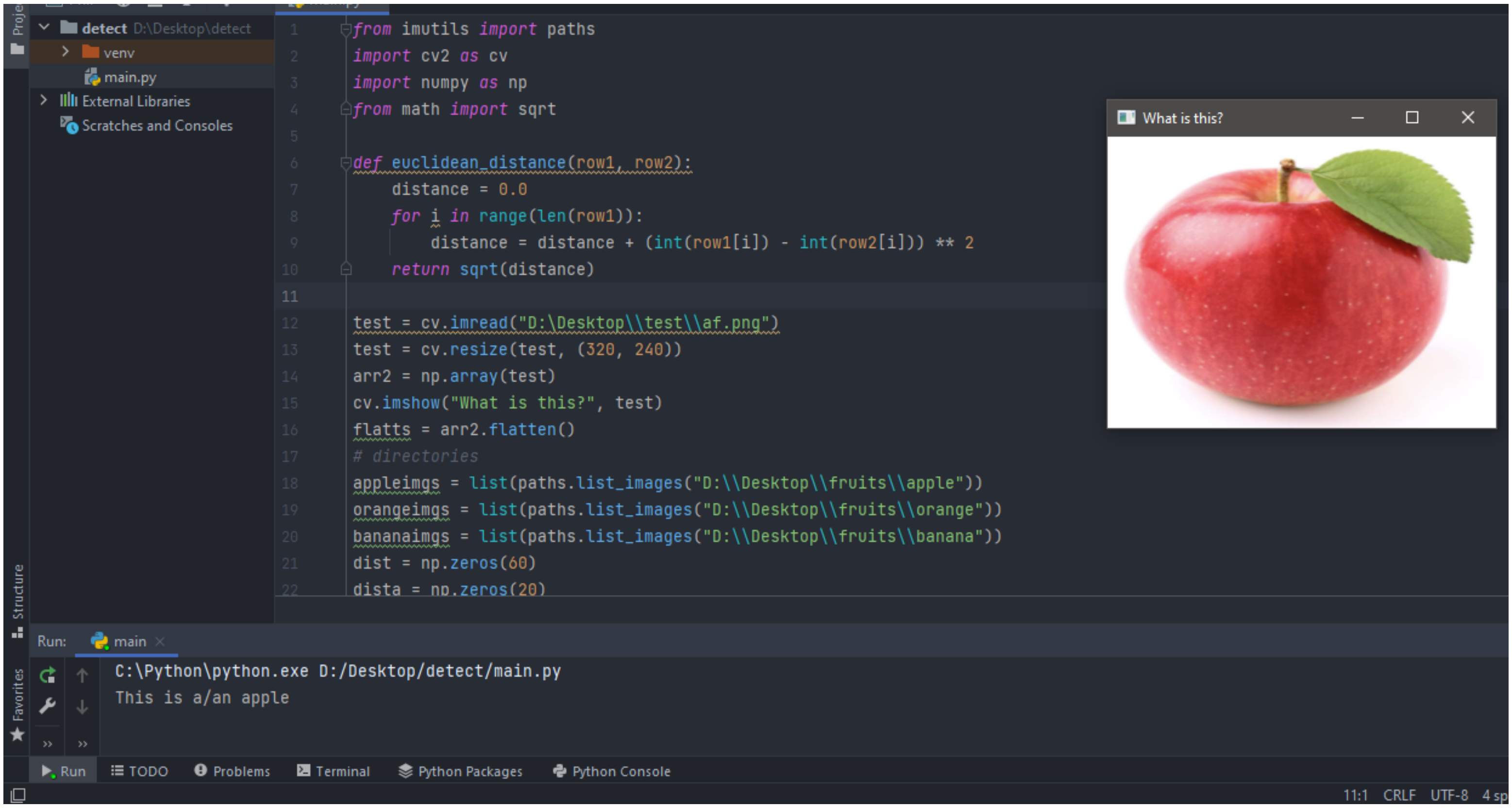


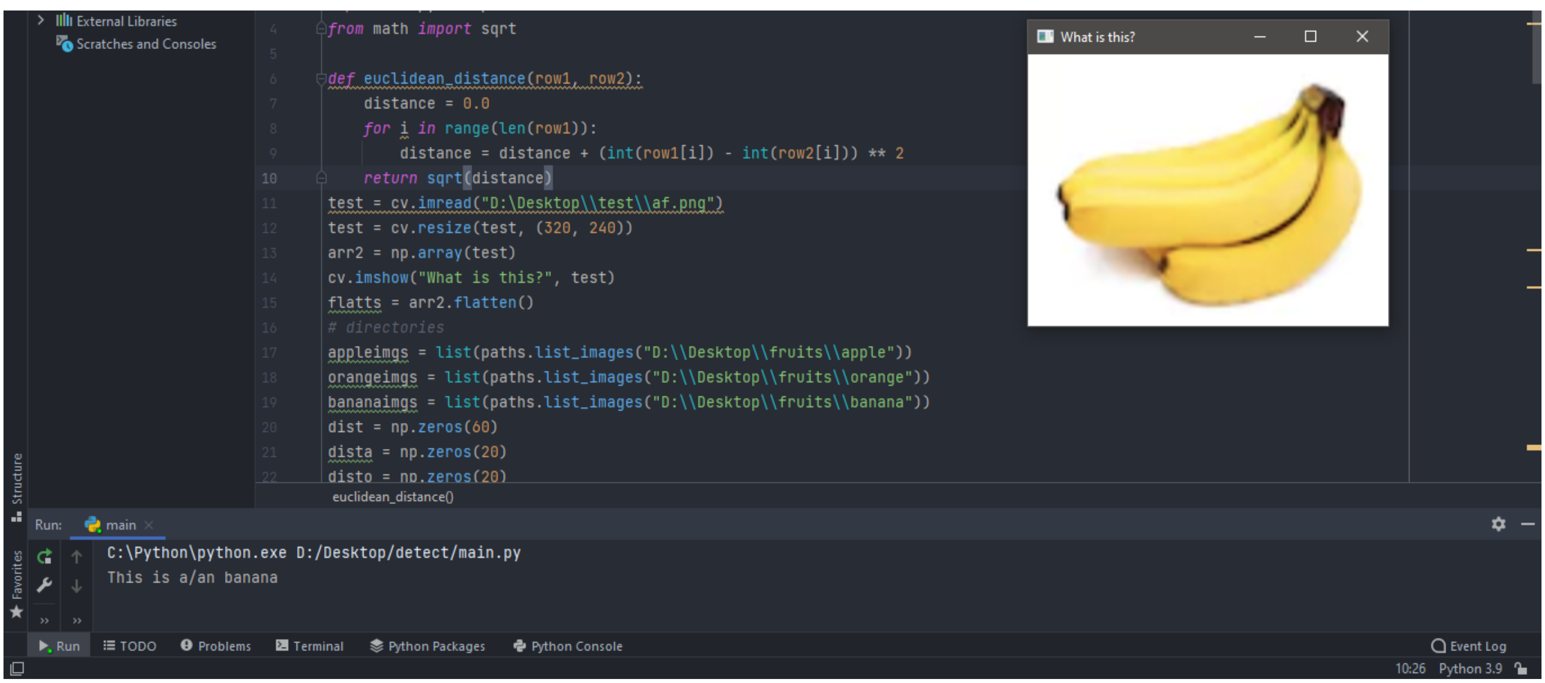
4. Appendix D: Use Case Diagram

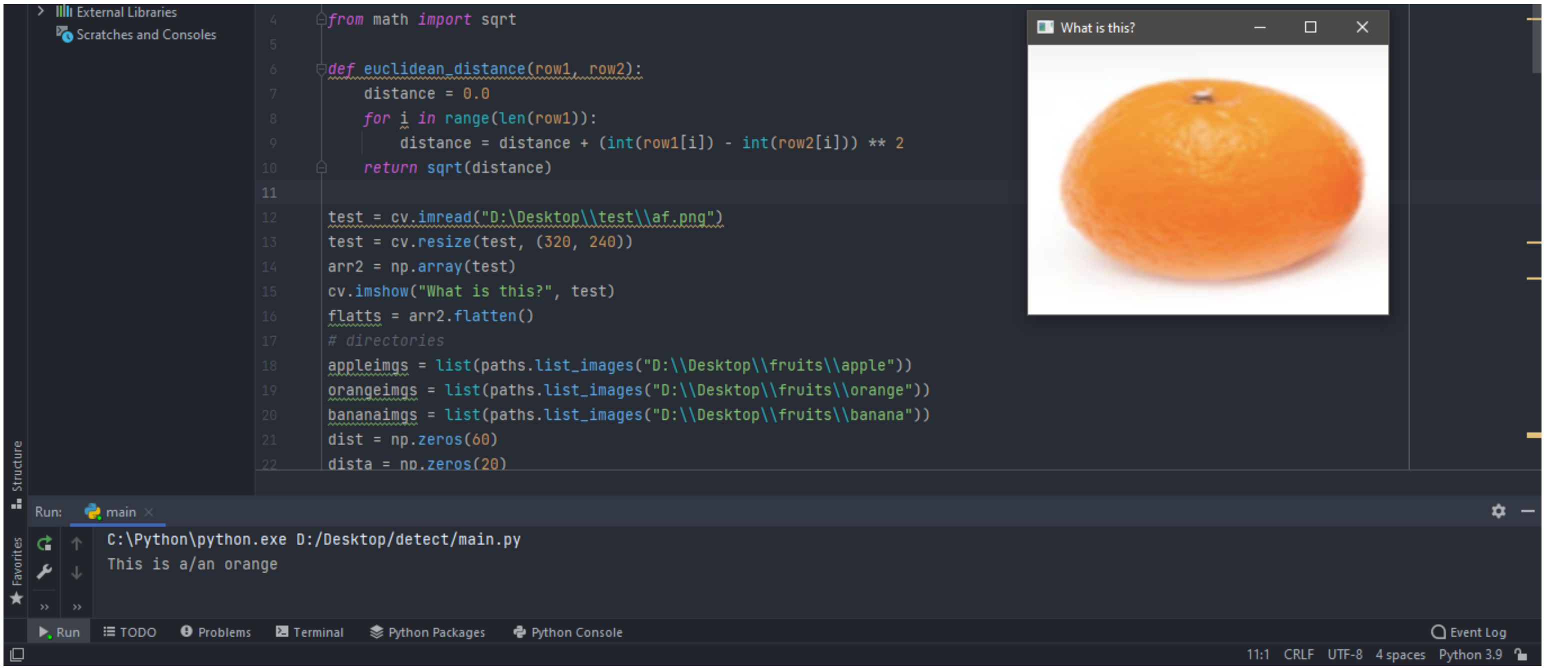


5. Appendix E: User Interface Design

6. Appendix F: Fruit Detection Prototype



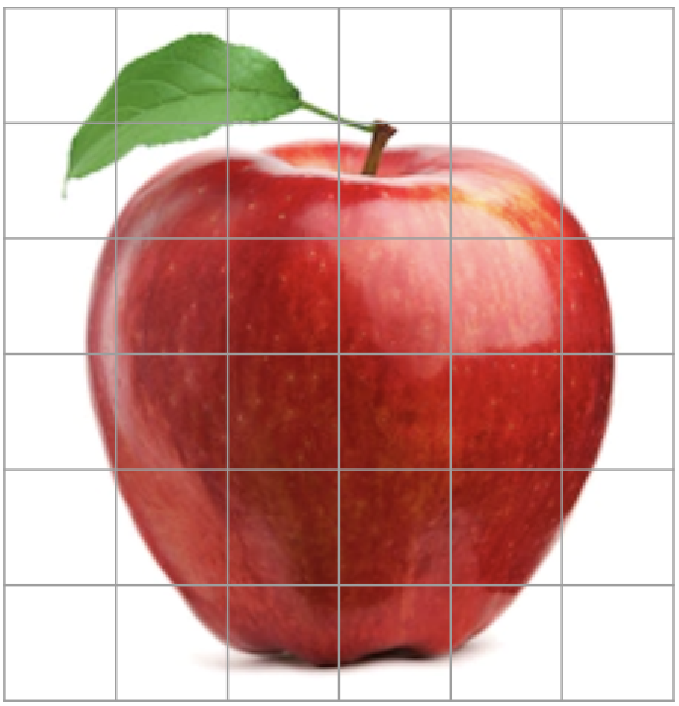



External Libraries
Scratches and Consoles
from math import sqrt
def euclidean_distance(row1, row2):
distance = 0.0
for i in range(len(row1)):
distance = distance + (int(row1[i]) - int(row2[i])) ** 2
return sqrt(distance)
test = cv.imread("D:\Desktop\\test\\af.png")
test = cv.resize(test, (320, 240))
arr2 = np.array(test)
cv.imshow("What is this?", test)
flatts = arr2.flatten()
# directories
appleimgs = list(paths.list_images("D:\\Desktop\\fruits\\apple"))
orangeimgs = list(paths.list_images("D:\\Desktop\\fruits\\orange"))
bananaimgs = list(paths.list_images("D:\\Desktop\\fruits\\banana"))
dist = np.zeros(60)
dista = np.zeros(20)
What is this?
Run: main
C:\Python\python.exe D:/Desktop/detect/main.py
This is a/an orange
Run
TODO
Problems
Terminal
Python Packages
Python Console
Event Log
11:1 CRLF UTF-8 4 spaces Python 3.9

7. Appendix F: KNN Classifier Algorithm

Vectorize the Image

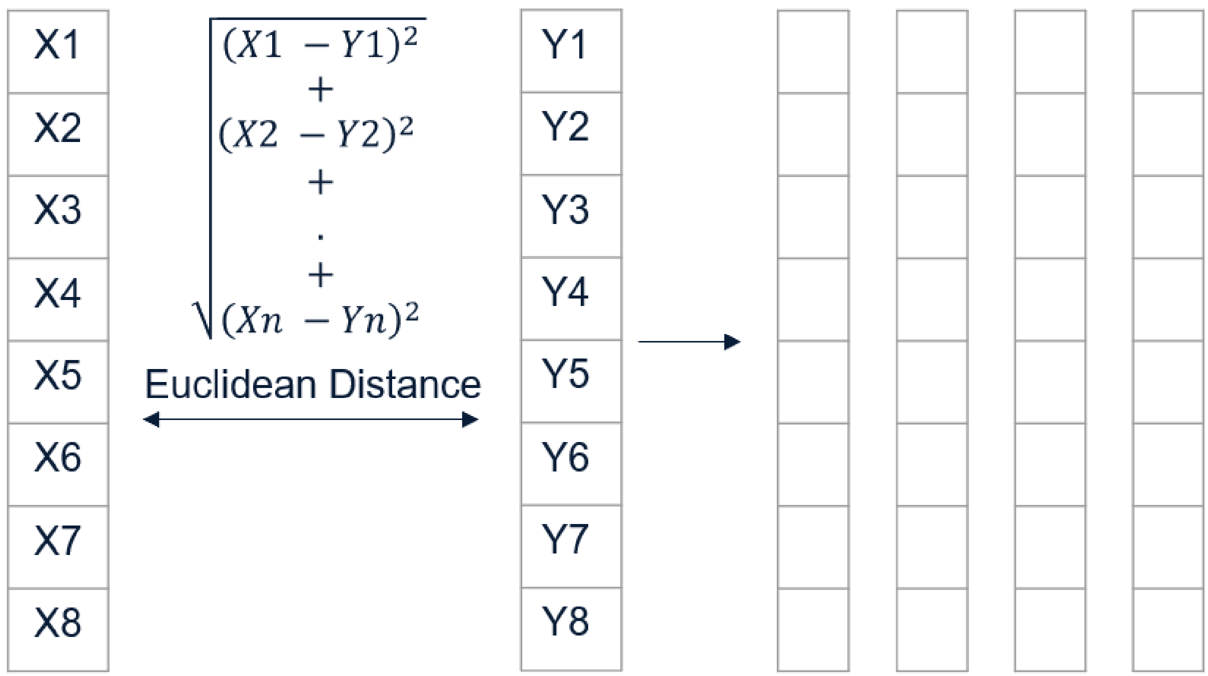
Original Image



Vectorized Image

X1
X2
X3
X4
X5
X6
X7
X8

Compute Euclidean Distance



Test Image

Data Set Images

Subject for Voting

K = 3

Apple	Orange	Apple	Banana	Orange	Banana	Apple	Banana	Orange
2	8	5	5	8	4	1	9	6