FRUIT CLASSIFICATION AND GRADING USING ML

Project Report Submitted

In Partial Fulfillment of the Requirements

For the Degree Of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

Submitted By

SULTAN HUSSAIN (1604-16-733-032)

MOHAMMED MUZAKKIR QADRI (1604-16-733-045)

MOHAMMED FAIZAN MOHIUDDIN (1604-16-733-038)

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

MUFFAKHAM JAH COLLEGE OF ENGINEERING & TECHNOLOGY

(Affiliated to Osmania University)

Mount Pleasant, 8-2-249, Road No. 3, Banjara Hills, Hyderabad-34

2020

Date: 14 / 04 / 2020

CERTIFICATE

This is to certify that the project dissertation titled " *Fruit Classification And Grading Using ML"* being submitted by

1. *Sultan Hussain (1604-16-733-032)*
2. *Mohammed Muzakkir Qadri (1604-16-733-045)*
3. *Mohammed FaizanMohiuddin (1604-16-733-038)*

in Partial Fulfillment of the requirements for the award of the degree Of BACHELOR OF ENGINEERING IN COMPUTER SCIENCE AND ENGINEERING in MUFFAKHAM JAH COLLEGE OF ENGINEERING AND TECHNOLOGY, Hyderabad for the academic year 2019-20 is the bonafide work carried out by them. The results embodied in this report have not been submitted to any other University or Institute for the award of any degree or diploma.

Signatures:

Internal Project Guide   Head CSED

Dr. Uma Dulhare   (Dr. A.A MoizQyser)

External Examiner

DECLARATION

This is to certify that the work reported in the major project entitled " *Fruit Classification And Grading Using ML"* is a record of the bonafidework done by us in the Department of Computer Science and Engineering, Muffakham Jah College of Engineering and Technology, Osmania University. The results embodied in this report are based on the project work done entirely by us and not copied from any other source.

1. *Sultan Hussain (1604-16-733-032)*
2. *Mohammed Muzakkir Qadri (1604-16-733-045)*
3. *Mohammed FaizanMohiuddin (1604-16-733-038)*

ACKNOWLEDGEMENT

We take the opportunity to thank our head of department DR. AHMED ABDUL MOIZ QYSER and instructor DR. UMA DULHARE for the most effective and valuable guidance. They have been the inspiration and were always up front to motivate and encourage us for bringing the project successfully. We would also like to thank our staff members for their valuable support.

We would also like to thank all the people who helped us in this project and whom we might have not mentioned here.

ABSTRACT

The classification and grading of fruits is a burdensome challenge due to the mass production of the fruits. Sorting and grading can be done by human but is inconsistent, time consuming and expensive. The main idea behind the proposed system is to automate the existing system as the automation in agricultural science increases the quality, economic growth and productivity of the country. The algorithm can discriminate three types of fruits (apple, banana, orange) and can also grade them into fresh and rotten categories using machine learning. A classifier is contemplated with three different databases. First database consisted of 4035 apples out of which 1693 were fresh and 2342 were rotten. Bananas database comprised a total of 3805 images from which 1581 were fresh and 2224 were rotten. The last database incorporated 3061 images of oranges out of which 1466 were fresh and 1595 were rotten. SVM (Support Vector Machine) algorithm was used for the classification and grading. Classification using SVM attained an accuracy of 91% whereas the grading of banana, apple and orange acquired 99.1%, 89.7% and 93.8%of accuracies respectively.

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Machine Learning:

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed**.** Machine Learning(ML) can be explained as automating and improving the learning process of computers based on their experiences without being actually programmed i.e. without any human assistance. The process starts with feeding good quality data and then training our machines(computers) by building machine learning models using the data and different algorithms. The choice of algorithms depends on what type of data do we have and what kind of task we are trying to automate.

In machine learning the models are built by training machine with data (both inputs and outputs are given to model) and when the time comes test on data (with input only) and achieves our model scores by comparing its answer with the actual output which has not been fed while training.

1.2 Difference Between Traditional Programming And Machine Learning

In traditional programming, the machine is fed with data (input) and the program (logic), which runs on the machine to get the output whereas in machine learning the machine is fed with the data (input) and the output, which runs on a machine during training to create its own program (logic) which is evaluated during the testing.

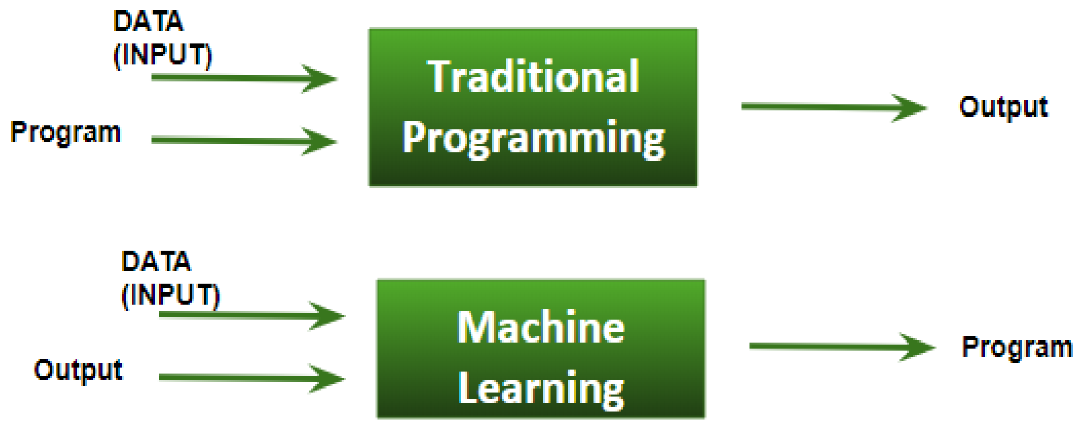


Fig. 1.2 Traditional programming versus Machine Learning

1.3 Types of Machine Learning

1.3.1 Supervised Machine Learning Algorithm

Supervised Learning Algorithms are the ones that involve direct supervision (cue the title) of the operation. In this case, the developer labels sample data corpus and set strict boundaries upon which the algorithm operates.The primary purpose of supervised learning is to scale the scope of data and to make predictions of unavailable, future or unseen data based on 2abellin sample data.The most widely used supervised algorithms are:

- Linear Regression.
- Logistical Regression.
- Random Forest.
- Gradient Boosted Trees.
- Support Vector Machines (SVM).
- Neural Networks.
- Decision Trees.
- Naive Bayes.
- Nearest Neighbour.

1.3.2 Unsupervised Machine Learning Algorithm

Unsupervised Learning is the one that does not involve direct control of the developer. If the main point of supervised machine learning is that you know the results and need to sort out the data, then in case of unsupervised machine learning algorithms the desired results are unknown and yet to be defined.Another big difference between the two is that supervised learning uses 2abellin data exclusively, while unsupervised learning feeds on unlabeled data.

The unsupervised machine learning algorithm is used for exploring the structure of the information, extracting valuable insights, detecting patterns, implementing this into its operation to increase efficiency.

The most widely used algorithms are:

- k-means clustering.
- t-SNE (t-Distributed Stochastic Neighbor Embedding).
- PCA (Principal Component Analysis).
- Association rule.

1.3.3 Semi Supervised Machine Learning Algorithm

Semi-supervised learning algorithms represent a middle ground between supervised and unsupervised algorithms. In essence, the semi-supervised model combines some aspects of both into a thing of its own.Semi-supervised machine learning algorithm uses a limited set of 2abellin sample data to shape the requirements of the operation (i.e., train itself).The limitation results in a partially trained model that later gets the task to label the unlabeled data. Due to the limitations of the sample data set, the results are considered pseudo-labeleddata.Finally, 3abellin and pseudo-labeled data sets are combined, which creates a distinct algorithm that combines descriptive and predictive aspects of supervised and unsupervised learning.

1.3.4 Reinforcement Machine Learning Algorithm

Reinforcement learning represents what is commonly understood as machine learning artificial intelligence. In essence, reinforcement learning is all about developing a self-sustained system that, throughout contiguous sequences of tries and fails, improves itself based on the combination 3abellin data and interactions with the incoming data. Reinforced ML uses the technique called exploration/exploitation. The mechanics are simple – the action takes place, the consequences are observed, and the next action considers the results of the first action. In the centre of reinforcement learning algorithms are reward signals that occur upon performing specific tasks. In a way, reward signals are serving as a navigation tool for the reinforcement algorithms. They give it an understanding of right and wrong course of action.

Most common reinforcement learning algorithms include:

- Q-Learning.
- Temporal Difference (TD).
- Monte-Carlo Tree Search (MCTS).
- Asynchronous Actor-Critic Agents (A3C)..

1.4 Advantages Of Machine Learning:

1. Easily identifies trends and patterns– Machine Learning can review large volumes of data and discover specific trends and patterns that would not be apparent to humans.

2. No human intervention needed (automation) –With ML, you don’t need to babysit your project every step of the way. Since it means giving machines the ability to learn, it lets them make predictions and also improve the algorithms on their own.

3. Continuous Improvement– As ML algorithms gain experience, they keep improving in accuracy and efficiency. This lets them make better decisions.

4. Handling multi-dimensional and multi variety data– Machine Learning algorithms are good at handling data that are multi-dimensional and multi-variety, and they can do this in dynamic or uncertain environments.

5. Wide Applications– You could be an e-tailer or a healthcare provider and make ML work for you. Where it does apply, it holds the capability to help deliver a much more personal experience to customers while also targeting the right customers.

1.3 Computer Vision

Computer vision is the field of computer science that focuses on replicating parts of the complexity of the human vision system and enabling computers to identify and process objects in images and videos in the same way that humans do. Until recently, computer vision only worked in limited capacity.

Thanks to advances in artificial intelligence and innovations in deep learning and neural networks, the field has been able to take great leaps in recent years and has been able to surpass humans in some tasks related to detecting and labeling objects.One of the driving factors behind the growth of computer vision is the amount of data we generate today that is then used to train and make computer vision better.

It is a multidisciplinary field that could broadly be called a subfield of artificial intelligence and machine learning, which may involve the use of specialized methods and make use of general learning algorithms.

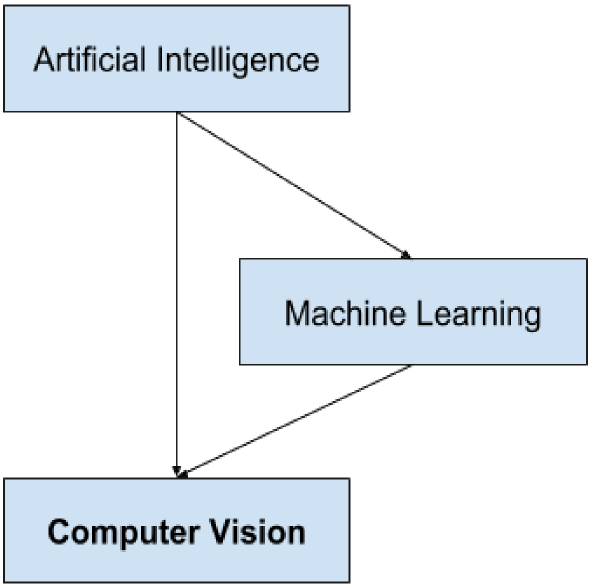


Fig. 1.5 Computer Vision

2. LITERATURE SURVEY

2.1Apples Shape Grading by Fourier Expansion and Genetic Program Algorithm(2008)

DOI: 10.1109/ICNC.2008.703

Authors –Zou Xiaobo, Zhao Jiewen Li Yanxiao, Shi Jiyong, Yin Xiaoping

Apple is a very important fruit in China. The shapeof the apple is important indices for classifications. Animage collecting system was developed and 4 imagesare obtained from each apple in this study. In order to describing the irregular shapes of apples, Fourier expansion was developed to reduce the dimensionality of the edge points. The image of an apple was captured by an image collecting system which consisted CCD colour camera, chamber, colour grabber, cone shape apple roller and a computer. After capturing the image, features were extracted from it by using Fourier Expansion. The features extracted were organised using genetic algorithm. A whole image data collecting system wasdeveloped in this study in order to inspect more fruit surface. This system consists of a cone-shape apple roller device controlled by stepping motor, a chamber, a lighting system and an image grab device. On this system, 4 images of different aspect are obtained from each apple, allowing the inspection of approximately 90% of the fruit surface.The Fourier descriptor, which used boundaryradii and their Fourier transform, was effective indescribing the apple shapes. 33 coefficients of theFourier descriptor were extracted for representing themain shape features of apple, whereas the first twoprinciple components of the 33 coefficients werecalculated. A new method called organization featureparameter based on formula expression tree by usinggenetic program was proposed in this paper. It couldsolve the problem how to getting optimum featureparameters. By applying “step decision tree” descriptorin combination with the new method to identify theshape of apples, the grade judgment ratios for “extra”,“categories II ” and “reject” are high, but the ratio for “category I” is not high.

2.2Fruits and vegetables quality evaluation using computer vision: A review(2018)

https://doi.org/10.1016/j.jksuci.2018.06.002

Authors –AnujaBhargava ,Atul Bansal

This paper highlights the use of image processing and computervision technology in the field of food industry and agriculture. Themost important quality characteristics of agricultural products aresize, color, shape, texture and defect. To replace manual inspection of food, computer vision system is used which provide authentic,equitable and non-destructive rating. The computer vision basedquality inspection comprises of four main steps, namely, acquisition,segmentation, feature extraction and classification. In thispaper, an attempt has been made to explore and compare the variousmethods/algorithms proposed by researchers in each step. Itcan be concluded from the extensive survey carried out in thispaper that although number of researchers have proposed variousmethods for the quality inspection of fruits and vegetables still arobust computer vision based system with improved performanceis required to be built.In the literature the images of fruits and vegetables are capturedmainly from one direction. However, the system performance mayimprove by considering the images of fruits and vegetables capturedfrom different directions. Authors have utilized differentcolor spaces for the color based feature extraction, still one may explore combination/ other color space to improve the performance.It can also be concluded from the work carried out in thispaper that one can include the images from different regions tomake the system regional bias free. In the work reported in literature,fruit and vegetable grading, sorting and disease recognitionare done on single fruit. A generalized system may also be designedto grade or sort and detect the defects of multiple fruits andvegetables.

2.3 Automatic Detection and Grading of Multiple Fruitsby Machine Learning(2019)

https://link.springer.com/article/10.1007/s12161-019-01690-6

Authors –AnujaBhargava ,Atul Bansal

In this work, fully automatic detection and grading of multiplefruits are proposed. The area of fruit is extracted by the splitandmerge algorithm and is segmented by fuzzy segmentation.After segmentation, multiple features (30 features) areextracted which are fed for fruit detection. The fruit detectionresults show that the system classifies four types of fruitsnamely apple, avocado, banana, and orange with a maximumaccuracy of 98.48% (SVM) for k = 10. The system uses geometricalfeatures (12 features) for the detection of fruit type,whereas multiple features (30 features) are used for grading offruit. The system also grades the fruits in Rank 1, Rank 2, anddefected ones with a maximum accuracy of 95.72% (SVM)for k = 10. Results showed that the proposed system with theSVM classifier has better and encouraging performance ascompared with k-NN, SRC, and ANN classifiers.The performance of the systemmay be further improved byconsidering a large number of fruit images, various techniquesfor segmentation, more significant features, and different combinationsof classifier technique andmust be installed in a reallifefruit detection and grading environment.

2.4 Comparative study of transform-based image texture analysis for the evaluation of banana quality using an optical backscattering system(2018)

https://doi.org/10.1016/j.postharvbio.2017.08.021

Authors –NorhashilaHashim, Segun Emmanuel Adebayo, KhalinaAbdan,Marsyita Hanafi

Image processing and pattern recognition has drawn significantattention in recent years as a result of higher demand in robotic andautomation applications, including the agroindustry. Although manystudies have been conducted to explore the feasibility of image processinganalysis, the approaches especially for transform-based imagetexture analysis in the evaluation of the quality of fresh produce is stilllimited. This work investigated the adaptability of LLBI with transformbasedtexture analysis coupled with computational intelligence techniquesi.e. ANN and SVM to evaluate the quality parameters of bananas.Three laser diodes emitting at 532, 660 and 830 nm were employed tocapture images of bananas at six ripening stages. Three transform-basedimage texture analysis methods, namely the Gabor transform, Wavelettransform and Tamura were used to extract features from the capturedbackscattered images of the fruit. The extracted features were used asinput into ANN and SVM prediction models. The results indicated that acombination of all wavelengths with the Wavelet transform approachprovided better evaluation of banana quality as reflected by higher R2.The Tamura transform provided quite promising results that were almostcomparable to the Wavelet transform. Although the Gabor transform recorded the poorest performance, the R2 of the predictionmodel however was at an acceptable value, indicating that the methodcan be adopted in the analysis as well. This signified that LLBI withtransform-based image texture analysis could potentially be used in thebackscattering image processing step as it offers higher accuracy in theprediction of banana quality that could conceivably lead to rapid andnon-destructive evaluation of fruit quality and subsequently possibly beused for online grading of the fruit.

2.5 Predicting the ripening of papaya fruit with digital imaging and randomforests(2018)

https://doi.org/10.1016/j.compag.2017.12.029

Authors –Luiz Fernando Santos Pereiraa, SylvioBarbonJr.b, Nektarios A. Valousb, Douglas FernandesBarbina

Classification performance is a good measure of the reliability of thefeature set. Since samples are classified based on pulp firmness, resultsindicate that this internal property is linked to the color of the fruitpeel. Through color image processing, satisfactory results are reportedfor papaya grading based on pulp firmness. By combining hand-crafted image features with machine learning techniques, it is possible to reachhigher accuracy and improve classification models for the papaya fruit.This particular application of computer vision could enable the developmentof a relatively simple and automated system to sort papayaaccording to different levels of maturity. Digital imaging technologiescontinue to change at a rapid pace and the techniques that have beendeveloped can evolve into industrial applications with the right integrationframework. Further investigation is required for achievingrobust levels of ripening prediction. This can be done through moreelaborate hand-crafted features and fast learning using some extremelearning machine variant, or through the use of deep learning in whichdirectly distilling information from images allows the reduction ofhuman bias.

2.6 Development of an automatic visual grading system for grafting seedlings(2017)

https://doi.org/10.1177/1687814016686265

Authors –Subo Tian, ZifanWang, Jifeng Yang, Zichen Huang, RuiliWang, LipingWangand Jianghui Dong

In this study, the comparative analysis of a vegetablegrafting machine and machine vision technology inagricultural production was conducted. This study canidentify that the blue light illumination can be used tocapture the best images, based on machine vision andimage processing theory. Furthermore, the image processingand corresponding algorithms were designedand developed using vision system software with Open CV image and video libraries on a Visual C++ programmingplatform.For 100 seedlings, an automatic grading test wasconducted based on a machine vision muskmelon scionautomatic seedling grading system. The results showedthat the automatic grading results are basically consistentwith manual measurement results, with a gradingsuccess rate of ;97% and 98% for muskmelon andpumpkin, respectively, a grading error of 22.3% to2.3%, and an average grading error of 20.57%. Thetest results met the requirements of the design system,which laid an important foundation for further developingan automatic vegetable grafting machine

3. SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

1. Sorting and Grading of fruits and vegetables manually.
2. Khojastehnazhand, Omid and Tabatabaeefar (2011) studied on development of a lemon sorting system based on color and size. Basically, two inspection stages of the system can be identified: external fruit inspection and internal fruit inspection. The former task was accomplished through processing of colour images, while internal inspection required special sensors for moisture, sugar and acid contents. In this paper, an efficient algorithm for grading lemon fruits was developed and implemented in visual basic environment. The system consists of two CCD cameras, two capture cards, an appropriate lighting system, a personal computer and other mechanical parts. The algorithm initially extracted the fruit from the background. The samples of different grades of lemon were situated in front of the cameras and were calibrated off-line.
3. Badhe, Singh and Bhatt (2011) studied on development and evaluation of mango grader. A computerized grading machine was developed and evaluated to grade Alphonso mangoes on weight basis in five grades. Logistic software was developed to run the grader. The performance of the grader was evaluated at four speeds (480, 600, 720 and 840 m s), four microprocessor settings (B1, B2, B3 and B4) and their effect was observed on five grades of mango viz., Grade I (326-375 g), Grade II (276-325 g), Grade III (226-275 g), Grade IV (176-225 g) and Grade V (< 175 > 376 g) for single lane. The statistically analyzed data showed maximum capacity of 950 kg .
4. Ukey and Unde (2010) developed a sapota fruit grader. In order to increase the output of fruit grading and save time and labour, a sapota fruit grader based on divergent roller type principle was designed and developed. The best combination of roller speed, its inclination and roller gap was found to be 223 r min, 4.5º and 38 to 64 mm, respectively for highest efficiency of 89.5%. The capacity of machine was 1,440 kg h and costed Rs.11, 450/- (without electric motor).

3.1.1 Disadvantages of existing system

1. Sorting and Grading of fruits and vegetables manually is expensive and time consuming. Human operations may be inconsistent, inefficient and time consuming. Grading operations were effected due to shortage of labour in peak seasons.
2. The setup of electronic image grading is expensive. As it includes internal and external examination, it is time inefficient.
3. The weight based grading required to pass the products on a conveyor belt at multiple times to achieve better performance, which makes it consume ample amount of time.

3.2 PROPOSED SYSTEM

The proposed system covers the software side of the computer vision system. It deals with the external features of the fruits (apple, banana and orange). Support Vector Machine (SVM) , the most commonly used supervised machine learning algorithm is used to train this model because of its promising performance. Other supervised machine learning algorithms include KNN, ANN etc. But, SVM has better accuracy when compared to other algorithms. After the model has been trained, captured images of the fruits that are meant for classifying and grading are used for testing the model. It consists of 5 stages for classification and 5 stages for grading. All the stages are the same for both except the end result. Classification will be classified into apple, banana or orange while grading categorizes into fresh or rotten.

3.3 SYSTEM REQUIREMENTS

3.3.1 HARDWARE REQUIREMENTS

- Processor : Intel(R) Celeron(R) 2957U @1.40GHz
- Installed memory (RAM) : 4 GB
- System type :64 bit Operating System

3.3.2 SOFTWARE REQUIREMENTS

- Operating System : Windows 7
- Coding Language : Python

  Libraries :
  - Numpy
  - Pillow
  - OpenCV2
  - SkLearn
  - Tkinter
- Coding Platform : Spyder(Anaconda)

4 SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE

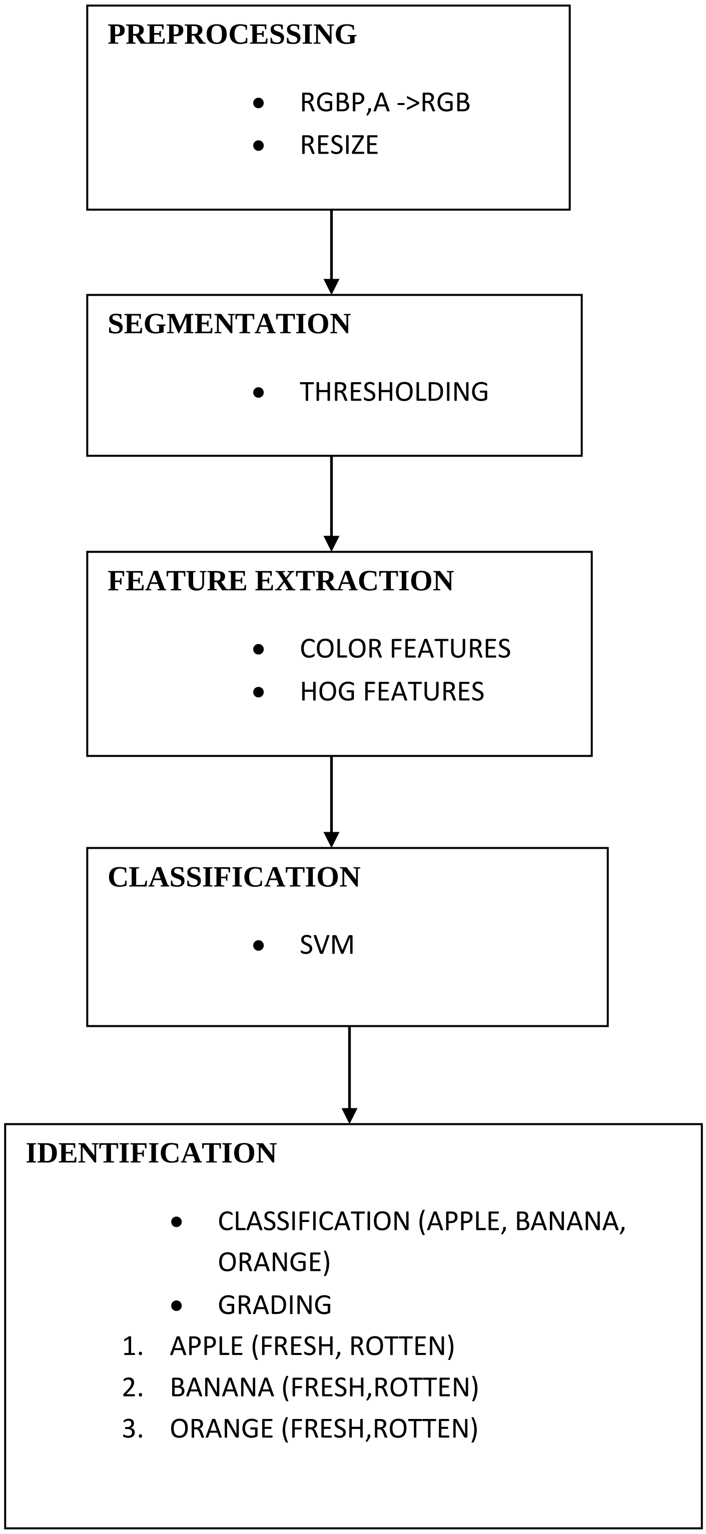


Fig. 4.1 System Architecture

4.2DATA FLOW DIAGRAM

Also known as DFD, Data flow diagrams are used to graphically represent the flow of data in a software model. DFD describes the processes that are involved in a system to transfer data from the input to the file storage and reports generation.

Data flow diagrams can be divided into logical and physical. The logical data flow diagram describes flow of data through a system to perform certain functionality of a business. The physical data flow diagram describes the implementation of the logical data flow.

DFD graphically representing the functions, or processes, which capture, manipulate, store, and distribute data between a system and its environment and between components of a system. The visual representation makes it a good communication tool between User and System designer. Structure of DFD allows starting from a broad overview and expand it to a hierarchy of detailed diagrams. DFD has often been used due to the following reasons:

- Logical information flow of the system.
- Determination of physical system construction requirements.
- Simplicity of notation.
- Establishment of manual and automated systems requirements.

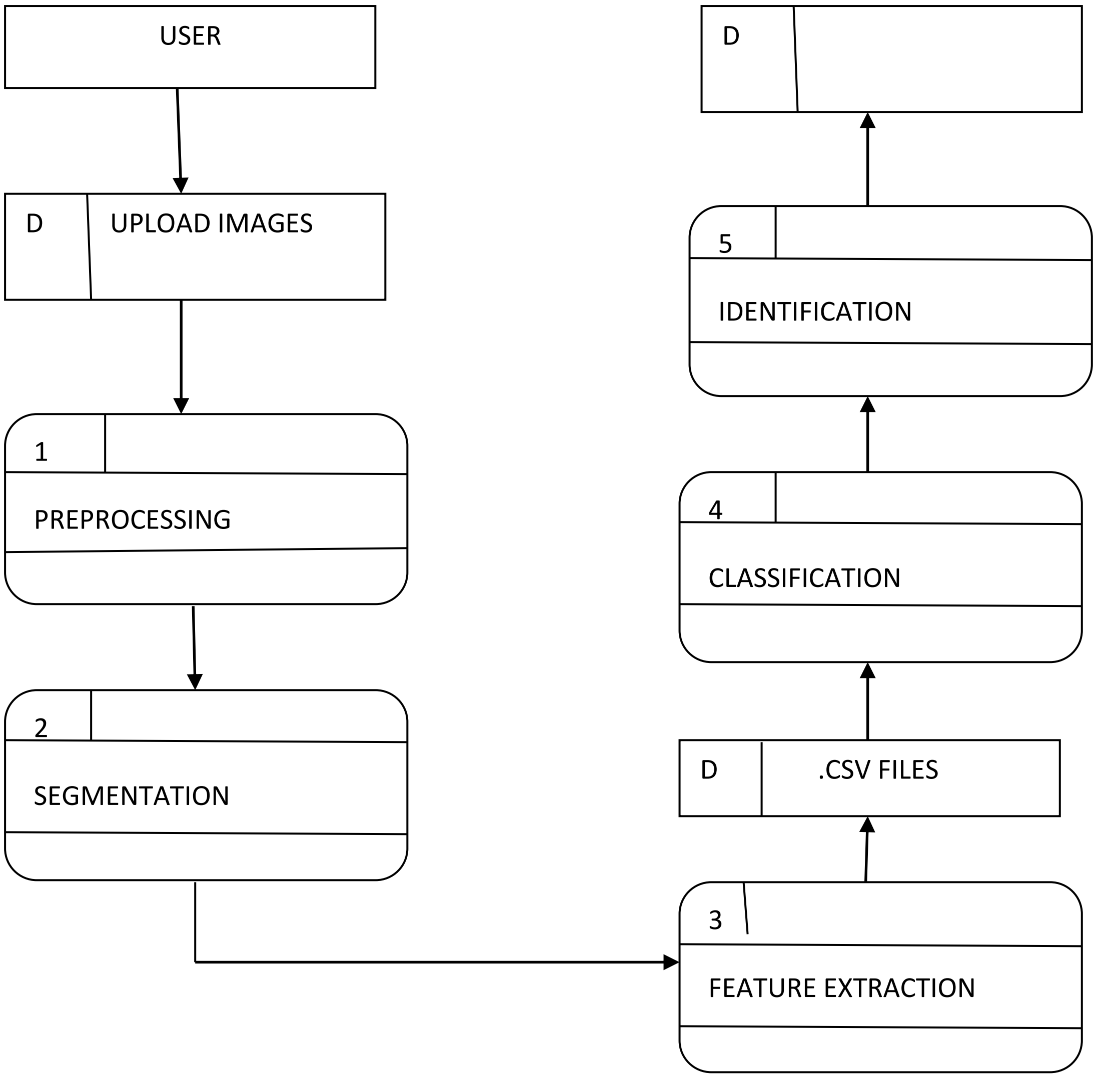


Fig. 4.2 Data Flow Diagram

4.3 UML DIAGRAMS

A UML diagram is a diagram based on the UML (Unified Modeling Language) with the purpose of **visually representing a system** along with its main actors, roles, actions, artifacts or classes, in order to better understand, alter, maintain, or document information about the system.

UML is an acronym that stands for **Unified Modeling Language**. Simply put, UML is a modern approach to modeling and documenting software. In fact, it’s one of the most popular business process modeling techniques.

It is based on **diagrammatic representations** of software components. As the old proverb says “a picture is worth a thousand words”. By using visual representations, we are able to better understand possible flaws or errors in software or business processes.

Goals of UML diagrams:

- Provide users with a ready-to-use, expressive visual modeling language so they can develop and exchange meaningful models.
- Provide extensibility and specialization mechanisms to extend the core concepts.
- Be independent of particular programming languages and development processes.
- Provide a formal basis for understanding the modeling language.
- Encourage the growth of the OO tools market.
- Support higher-level development concepts such as collaborations, frameworks, patterns and components.
- Integrate best practices.

4.3.1 USE CASE DIAGRAM

A use case diagram is used to represent the dynamic behaviour of a system. It encapsulates the system's functionality by incorporating use cases, actors, and their relationships. It models the tasks, services, and functions required by a system/subsystem of an application. It depicts the high-level functionality of a system and also tells how the user handles a system.

The main purpose of a use case diagram is to portray the dynamic aspect of a system. It accumulates the system's requirement, which includes both internal as well as external influences. It invokes persons, use cases, and several things that invoke the actors and elements accountable for the implementation of use case diagrams. It represents how an entity from the external environment can interact with a part of the system.

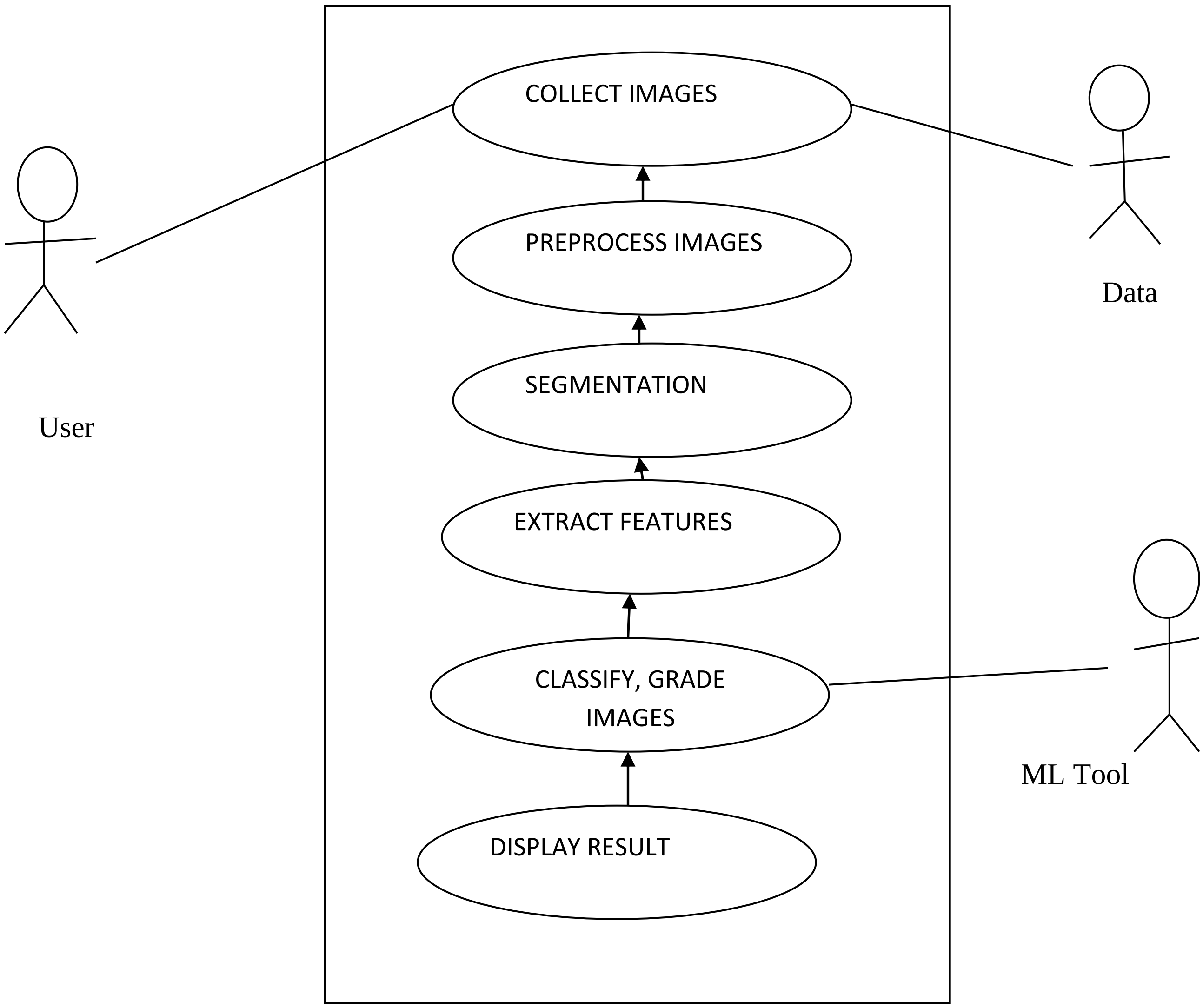


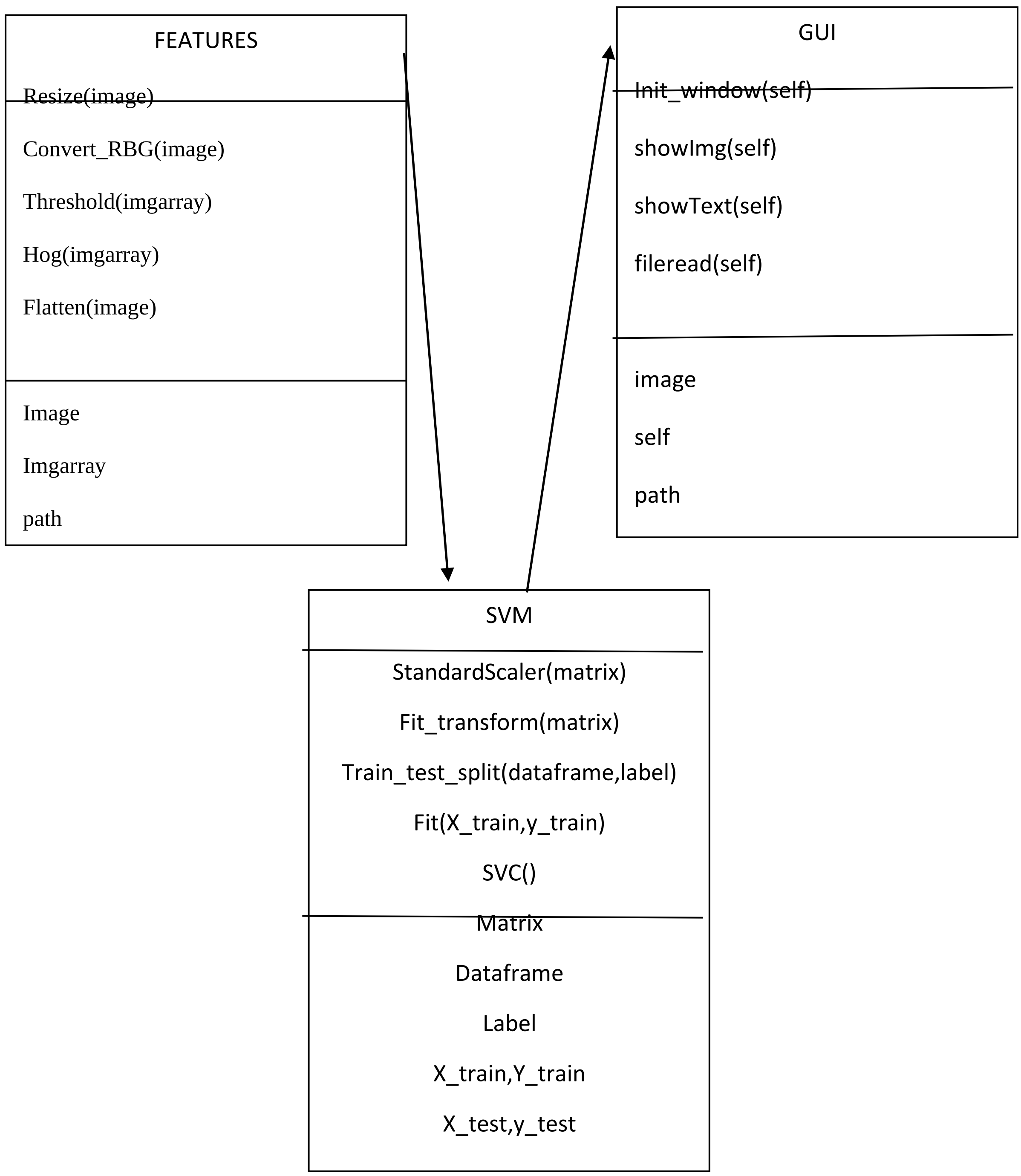
Fig. 4.3.1 Use Case Diagram

4.3.2 CLASS DIAGRAM

The class diagram depicts a static view of an application. It represents the types of objects residing in the system and the relationships between them. A class consists of its objects, and also it may inherit from other classes. A class diagram is used to visualize, describe, document various different aspects of the system, and also construct executable software code.

It shows the attributes, classes, functions, and relationships to give an overview of the software system. It constitutes class names, attributes, and functions in a separate compartment that helps in software development. Since it is a collection of classes, interfaces, associations, collaborations, and constraints, it is termed as a structural diagram.

The main purpose of class diagrams is to build a static view of an application. It is the only diagram that is widely used for construction, and it can be mapped with object-oriented languages. It is one of the most popular UML diagrams.



4.3.3 ACTIVITY DIAGRAM

In UML, the activity diagram is used to demonstrate the flow of control within the system rather than the implementation. It models the concurrent and sequential activities.

The activity diagram helps in envisioning the workflow from one activity to another. It put emphasis on the condition of flow and the order in which it occurs. The flow can be sequential, branched, or concurrent, and to deal with such kinds of flows, the activity diagram has come up with a fork, join, etc.

It is also termed as an object-oriented flowchart. It encompasses activities composed of a set of actions or operations that are applied to model the behavioral diagram.

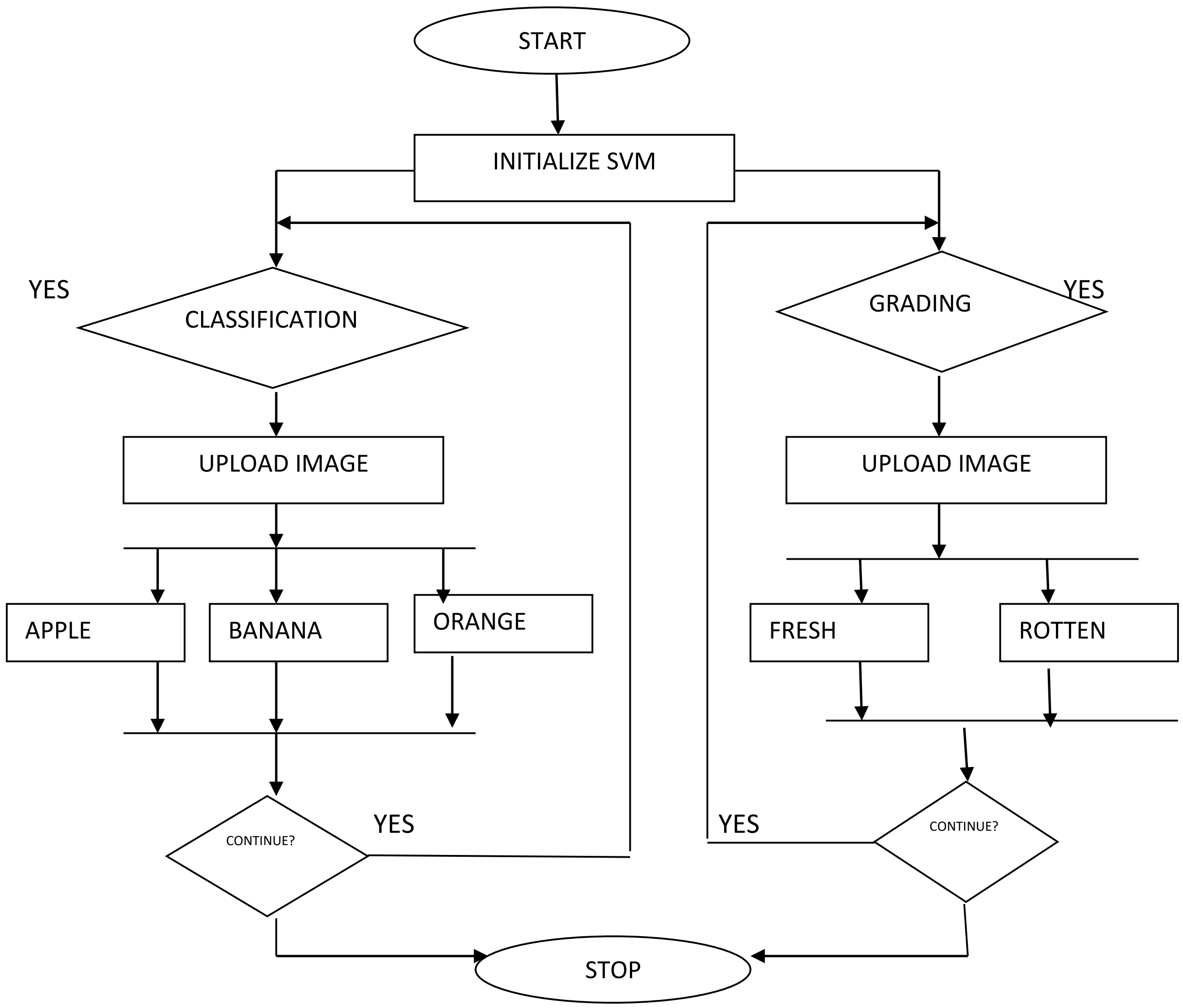


Fig. 4.3.3 Activity Diagram

5 MODULE IMPLEMENTATION

5.1 Module Description

5.1.1 System Construction Module

In first module, we developed the proposed system with required entities for the evaluation of the proposed model. In preprocessing, the initial image is processed for retrieving the analytical data of the image which is transformed into relevant data in the form of a matrix for training the Support Vector Machine model.

Preprocessingmdule alongwith constructed csv (comma seperated) file which contains the classifiers labels and creates an SVM model for fruit classification and grading.

5.1.2 Preprocessing Module

5.1.2.1 Image Formatting -

```
defconvert_Image(path):
    fromos import listdir
    from PIL import Image
    for filename in listdir(path):
    image = Image.open(path+'/'+filename)
    ifimage.mode=='RGBA':
    background = Image.new("RGB", image.size, (255, 255, 255))
    background.paste(image, mask = image.split()[3])
    background.save(path+'/'+filename, "JPEG", quality=100)
print(filename+' done RGBA->RGB')
```

In this snippet, the various colorchannels(P has one channel, RGBA has four channels) of the image are being converted into three channelled RGB images. The original image is being pasted on a three channelled white background to create the new RGB image.

5.1.2.2 Image Resize

```
resizedImage = image.resize((100,100))
resized_images.append(resizedImage)
reimagearr = asarray(resizedImage)
```

The image is resized to a smaller size and converted to an array, to reduce the data elements thereby decreasing the time required for the training of the SVM model. The time consumed by the SVM model is directly proportional to the number of data elements of the image.

5.1.3 Segmentation Module

```
gs_image =resizedImage.convert(mode='L')

dataarr = asarray(gs_image)

RET2,output2=cv2.threshold(dataarr,
th,max_val,cv2.THRESH_TRUNC+cv2.THRESH_OTSU)
thresholded1=output2.flatten()

RET3,output3=cv2.threshold(dataarr, th, max_val,
cv2.THRESH_TOZERO_INV+cv2.THRESH_OTSU)
```

Segmentation is the retrieval of a relevant parts of the image. This can be obtained using various methods one being Thresholding. Thresholding uses a parameter ‘th’ which is used to define the boundary for segmentation. In OTSU’s Thresholding, the th is kept zero. OTSU’s Algorithm determines the boundary itself. Various types of thresholding are TO_ZERO, INV_ZERO, BIN, INV_BIN, TRUNC. As seen in the snippet above TO_ZERO and TRUNC have been combined with OTSU’s Algorithm to achieve segmentation.

5.1.4 Feature Extraction Module

```
color_features = reimagearr.flatten()

color_features1 = np.float64(color_features)

hog_features2, hog_image2 = hog(reimagearr,visualize=True,block_norm='L2-Hys',pixels_per_cell=(12, 12))

hog_features1, hog_image1 = hog(gs_image,visualize=True,block_norm='L2-Hys',pixels_per_cell=(12, 12))

 flat_features1=np.concatenate((color_features1,hog_features2,thresholded1),axis=None)

 flat_features2=flat_features1.reshape(-1, 1)

flat_features3=flat_features2.transpose()

features_list.append(flat_features3)
```

In feature extraction, color features are obtained by joining the three channels of RGB into a single channel. HOG Features (Histogram Of Oriented Gradients) focuses on the high intensity regions as output, hence providing the shape of the relevant parts of the image. The color features, HOG features and the thresholded features are combined into a single array called flat features. The flat features of images in the dataset are combined to get a single features list or matrix for multiple images.

5.1.5 Create SVM Module

```
defCreate_Model(feature_matrix1,labels):

    ss = StandardScaler()

fruits_stand= ss.fit_transform(feature_matrix1)

fruits_pca = ss.fit_transform(fruits_stand)

  X = pd.DataFrame(fruits_pca)

y = pd.Series(labels.genus.values)
```

```
y = labels.iloc[:,0].values
X_train, X_test, y_train, y_test    =
train_test_split(X,y,test_size=.25122124,random_state=1234123)
svm = SVC(kernel='linear', probability=True, random_state=42)
svm.fit(X_train, y_train)
```

The features matrix is standardised to (-1,1) using the standard scaler. The standardised feature list and the labels are split into testing and training sets i.e., X_train, X_test, y_train and y_test using the train_test_split(). The model is trained for X_train and y_train values using SVM algorithm.

5.1.6 Classifier And Grader Module

5.1.6.1 Testing the Model

```
y_pred = svm.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
print('Model accuracy is: ', accuracy)
print("Precision:",precision_score(y_test,y_pred,pos_label='positive',average='weighted'))
print("Recall:",recall_score(y_test, y_pred, average='macro'))
probabilities = svm.predict_proba(X_test)
```

The accuracy, precision and recall of the trained model is calculated.

5.1.6.2 Prediction Module

```
deftestSVMImageClass(svm,image):
    feature_matrix2=preProcessFeaturesImageClass(image)
    X_TEST1 = pd.DataFrame(feature_matrix2)
    y_pred1 = svm.predict(X_TEST1)
    return y_pred1
```

In this module, the image of the fruit for classification and grading is provided to the SVM model after proper extraction of features, to predict the class and the grade of the fruit.

6 SYSTEM TESTING

Software testing is a process, to evaluate the functionality of a software application with an intent to find whether the developed software met the specified requirements or not and to identify the defects to ensure that the product is defect-free in order to produce the quality product. A process of analyzing a software item to detect the differences between existing and required conditions (i.e., defects) and to evaluate the features of the software item.

6.1 Software Testing Types:

6.1.1 Manual Testing:

Manual testing is the process of testing software by hand to learn more about it, to find what is and isn't working. This usually includes verifying all the features specified in requirements documents, but often also includes the testers trying the software with the perspective of their end user's in mind. Manual test plans vary from fully scripted test cases, giving testers detailed steps and expected results, through to high-level guides that steer exploratory testing sessions.

6.1.2 Automation Testing:

Automation testing is the process of testing the software using an automation tool to find the defects. In this process, testers execute the test scripts and generate the test results automatically by using automation tools. Some of the famous automation testing tools for functional testing are QTP/UFT.

6.2 Testing Approaches:

There are three types of software testing approaches.

6.2.1 White Box Testing:

It is also called as Glass Box, Clear Box, Structural Testing. White Box Testing is based on application's internal code structure. In white-box testing, an internal perspective of the system, as well as programming skills, are used to design test cases. This testing is usually done at the unit level.

6.2.2 Black Box Testing:

It is also called as Behavioral/Specification-Based/Input-Output Testing. Black Box Testing is a software testing method in which testers evaluate the functionality of the software under test without looking at the internal code structure.

6.2.3 Grey Box Testing:

Grey box is the combination of both White Box and Black Box Testing. The tester who works on this type of testing needs to have access to design documents. This helps to create better test cases in this process.

6.3 Testing Levels:

6.3.1Unit Testing:

Unit Testing is done to check whether the individual modules of the source code are working properly. i.e. testing each and every unit of the application separately by the developer in the developer's environment. It is AKA Module Testing or Component Testing.

6.3.2 Integration Testing:

Integration Testing is the process of testing the connectivity or data transfer between a couple of unit tested modules. It is AKA I&T Testing or String Testing. It is subdivided into the Top-Down Approach, Bottom-Up Approach, and Sandwich Approach (Combination of Top-Down and Bottom-Up).

6.3.3 System Testing (End to End Testing):

It's a black box testing. Testing the fully integrated application this is also called as an end to end scenario testing. To ensure that the software works in all intended target systems. Verify thorough testing of every input in the application to check for desired outputs. Testing of the user's experiences with the application.

6.3.4 Acceptance Testing:

To obtain customer sign-off so that software can be delivered and payments received. Types of Acceptance Testing are Alpha, Beta & Gamma Testing.

6.4 Test Strategy And Approach:

Field testing will be performed manually and functional tests will be written in detail.

6.4.1 Test Objectives:

- The entry screens, messages and responses must not be delayed.
- All field entries must work properly.

6.4.2 Features to be tested :

- Verify that preprocessing is done on the entire dataset.
- Verify that thresholding achieves proper segmentation of the fruit from the background.
- Every element of the user interface must have proper functionality.

Test Cases : All the test cases mentioned above passed successfully. No defects were encountered.

7 RESULTS

The following accuracies of the classification were obtained by using different algorithms as mentioned in the table below. As per the table, the maximum accuracy was obtained by using SVM algorithm. The following dataset was used for the classification-

https://www.kaggle.com/sriramr/fruits-fresh-and-rotten-for-classification

S.NO	ALGORITHM	ACCURACY
1	KNN(K=5)	71.929%
	KNN(K=10)	75.43%
	KNN(K=15)	78.94%
	KNN(K=20)	87.71%
2	SRC	82.75%
3	SVM	89.23%
4	ANN	88.17%

Table 7.1 Performance of classification by various algorithms

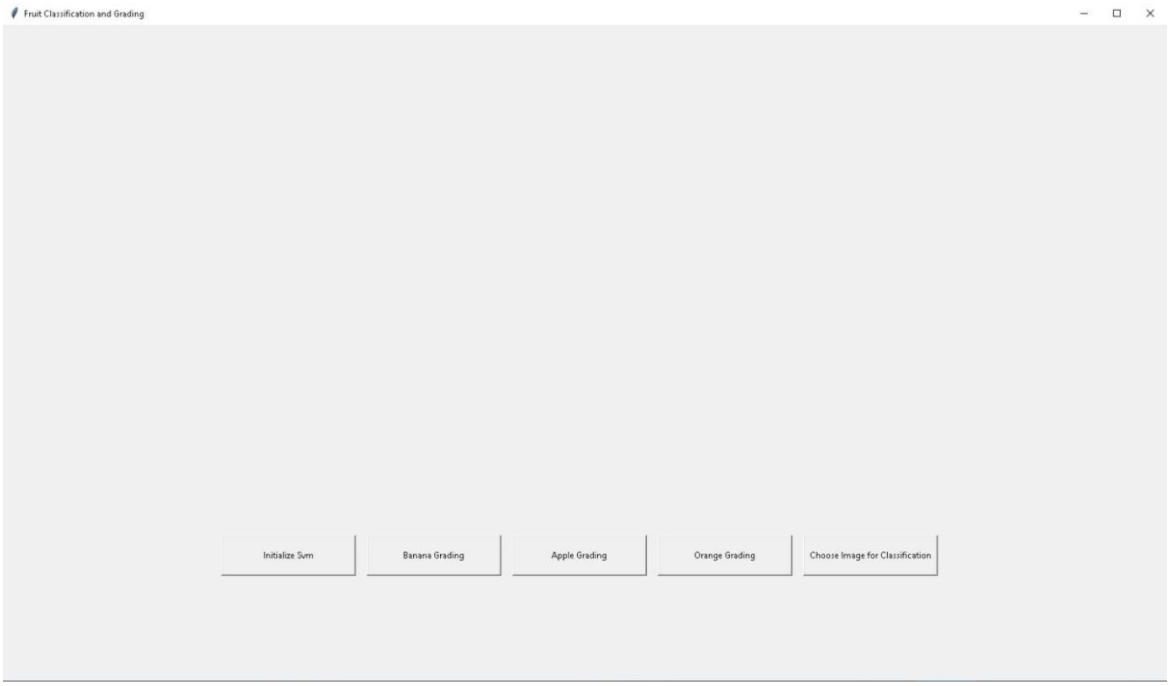
The following table consists of the performance metrics of the grading of apple, banana and orange. It uses the same database mentioned above.

FRUIT	**ACCURACY** = $\frac{TP+TN}{TP+TN+FP+FN}$	**RECALL** = $\frac{TP}{TP+FN}$	**SENSITIVITY** = $\frac{TP}{TP+FN}$	**PRECISION** = $\frac{TP}{TP+FP}$	**SPECIFICITY** = $\frac{TN}{TN+FP}$	**CONFUSION MATRIX** **[TP FP FN TN]**
APPLE	89.349	87.5	87.5	87.089	90.67	[371 55 53 535]
BANANA	99.058	99.224	99.224	98.461	98.945	[384 6 3 563]
ORANGE	93.888	93.010	93.010	94.2777	94.710	[346 21 26 376]

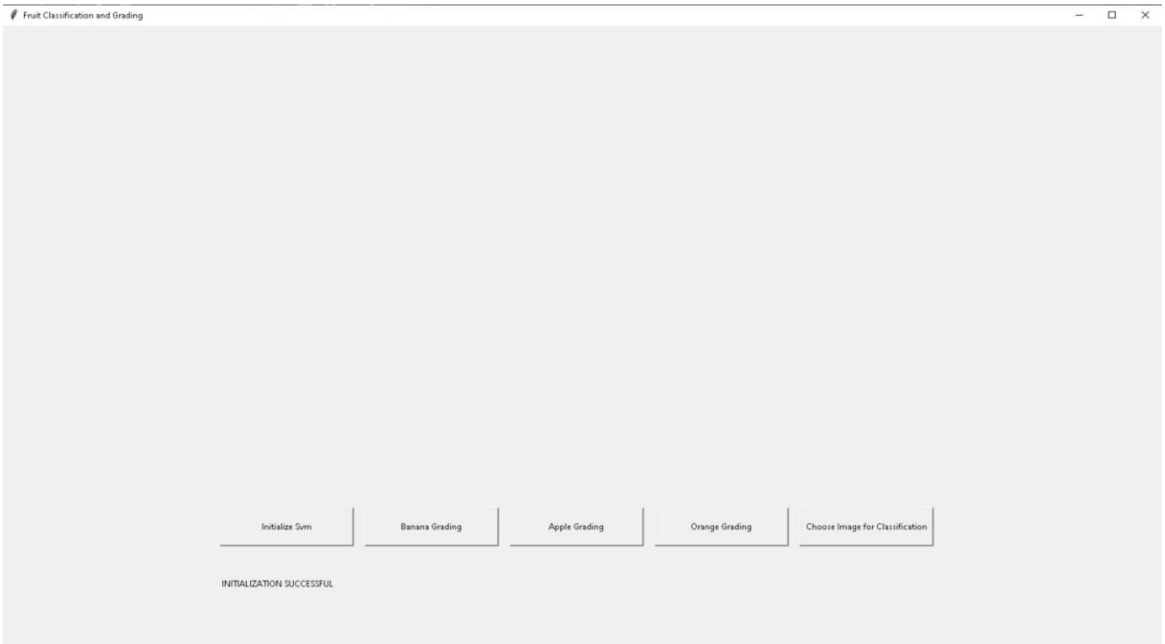
Table 7.2 Performance of grading

8 SCREENSHOTS

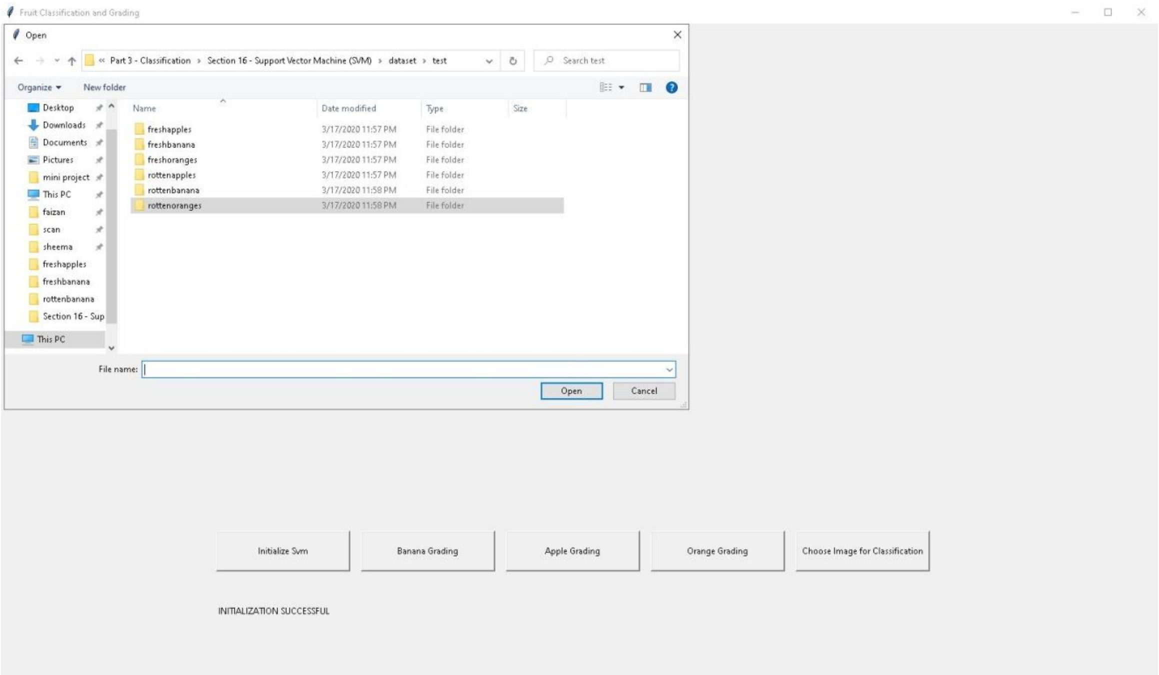
1. The GUI is present below.



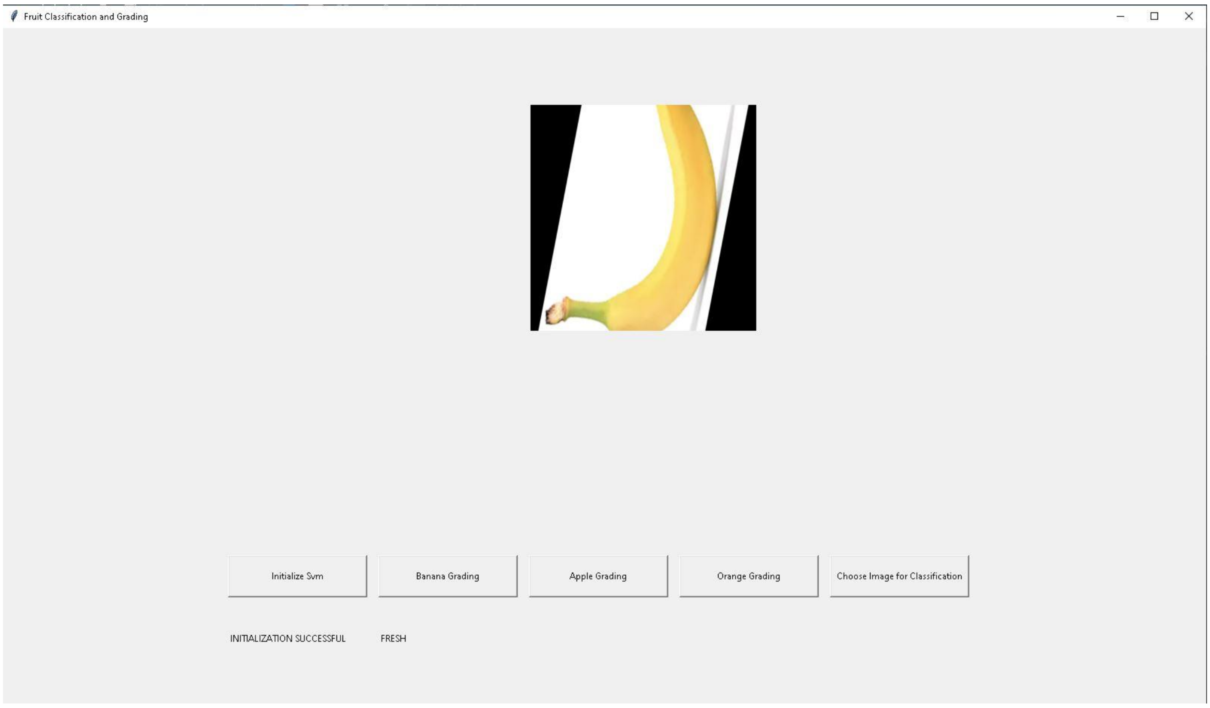
2. Initialization of the SVM is successful as shown below.



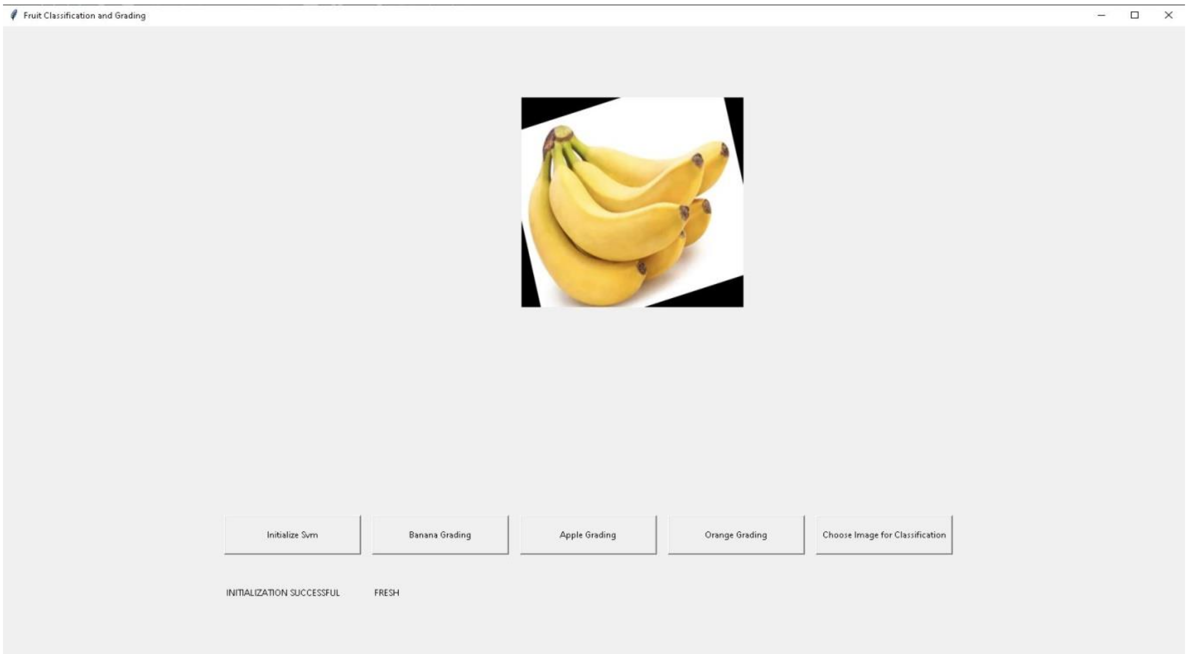
3. An image being uploaded after the initialization of SVM.



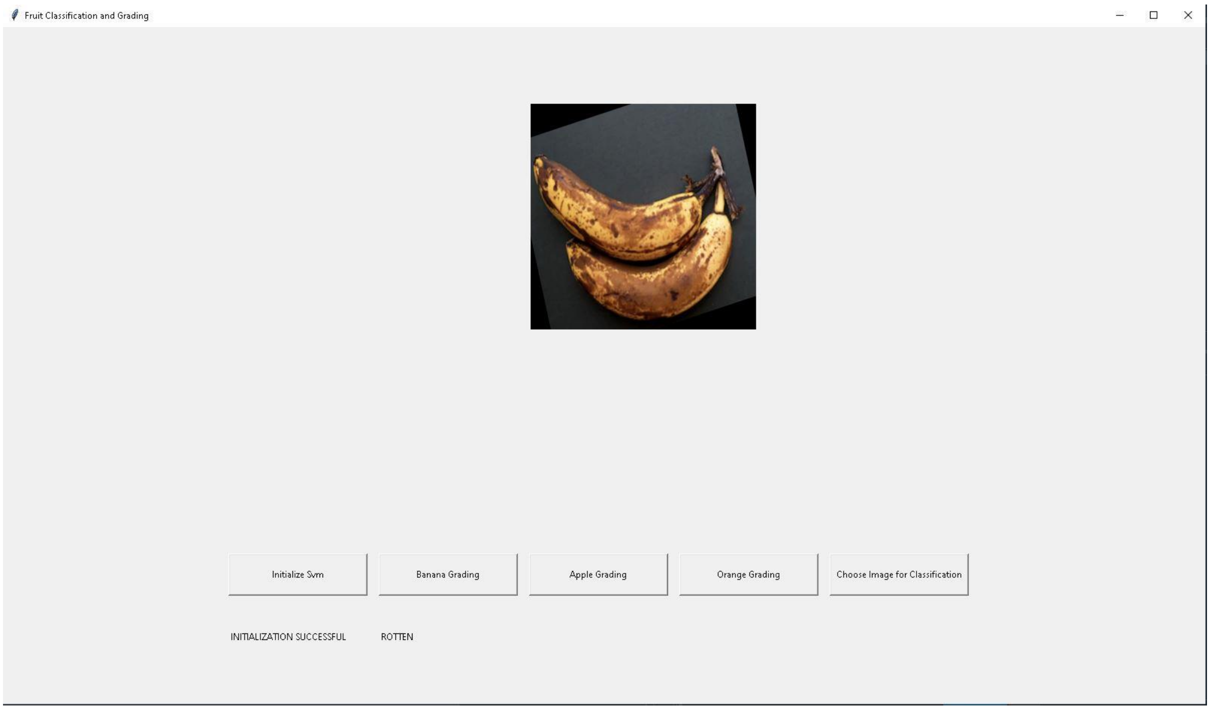
4. SVM is successfully able to identify a fresh banana.



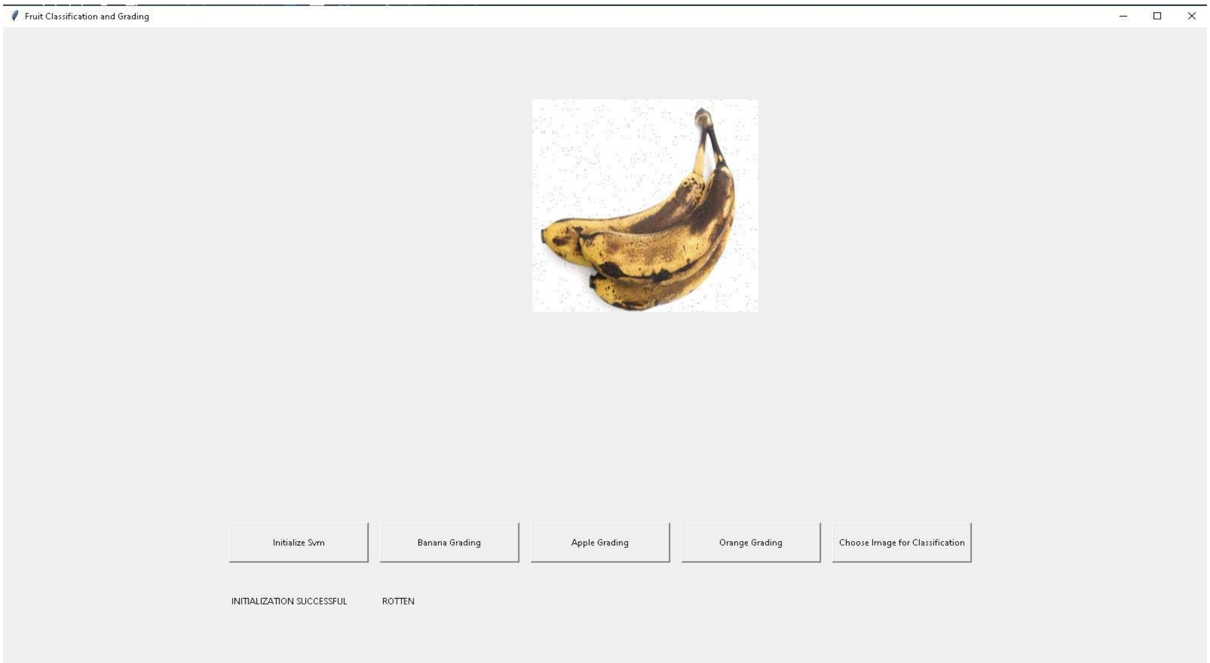
5. If multiple bananas are given, it still identifies them as banana and grades them as fresh.



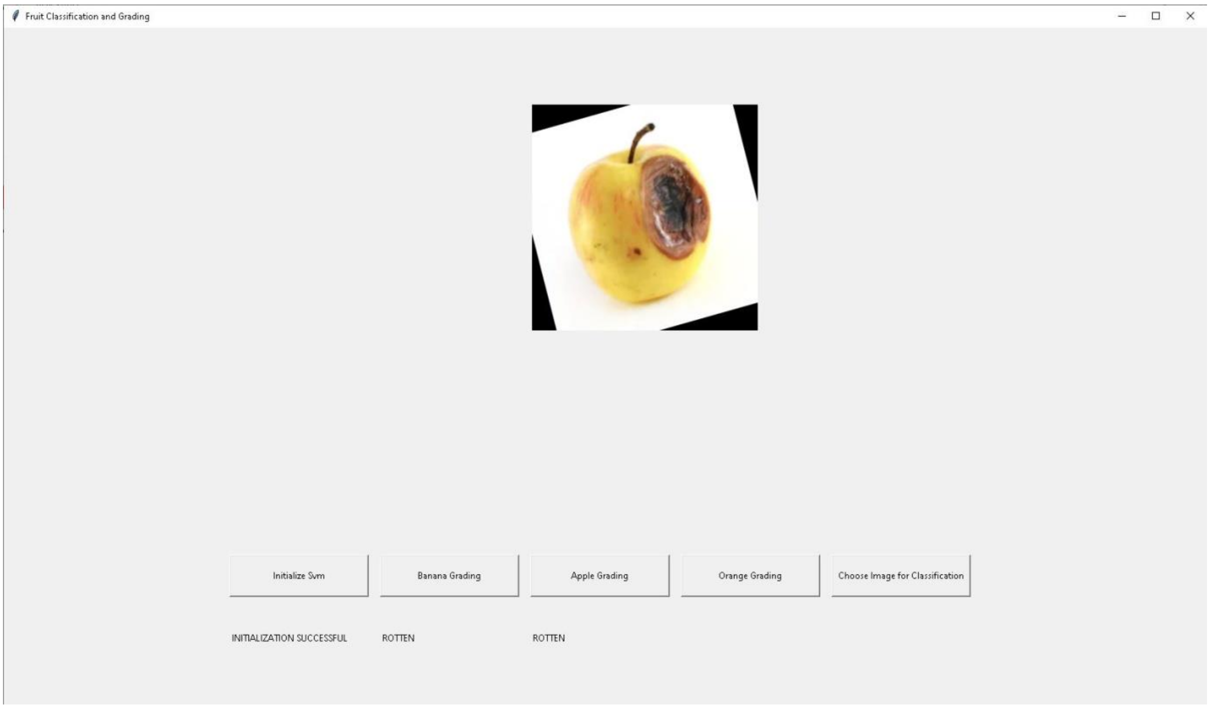
6. It identifies as banana and grades them as rotten.



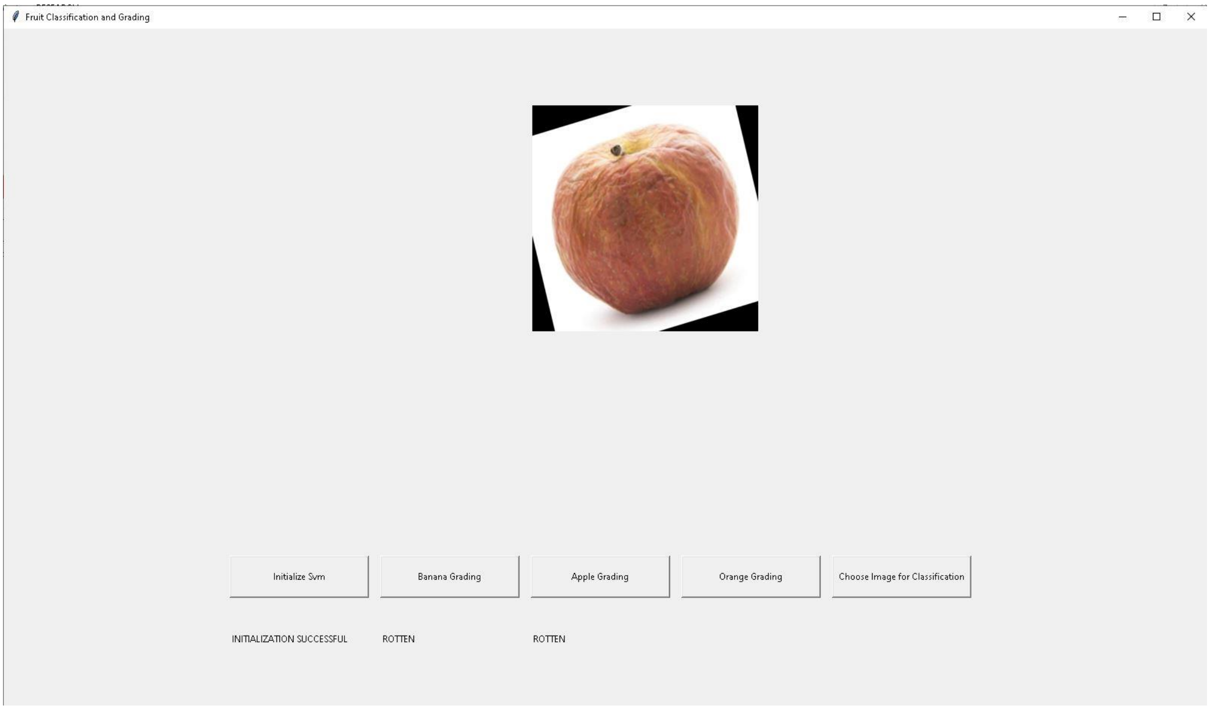
7. Successfully grades banana as rotten.



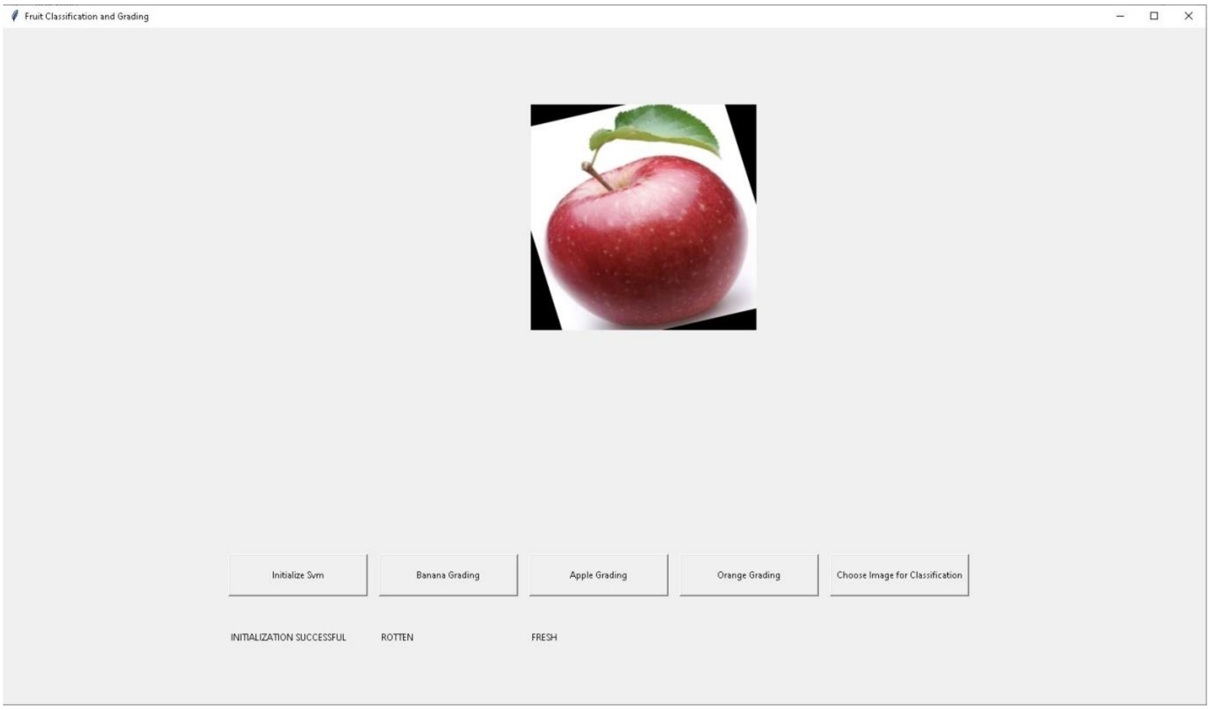
8. SVM identifies it as apple and grades it as rotten.



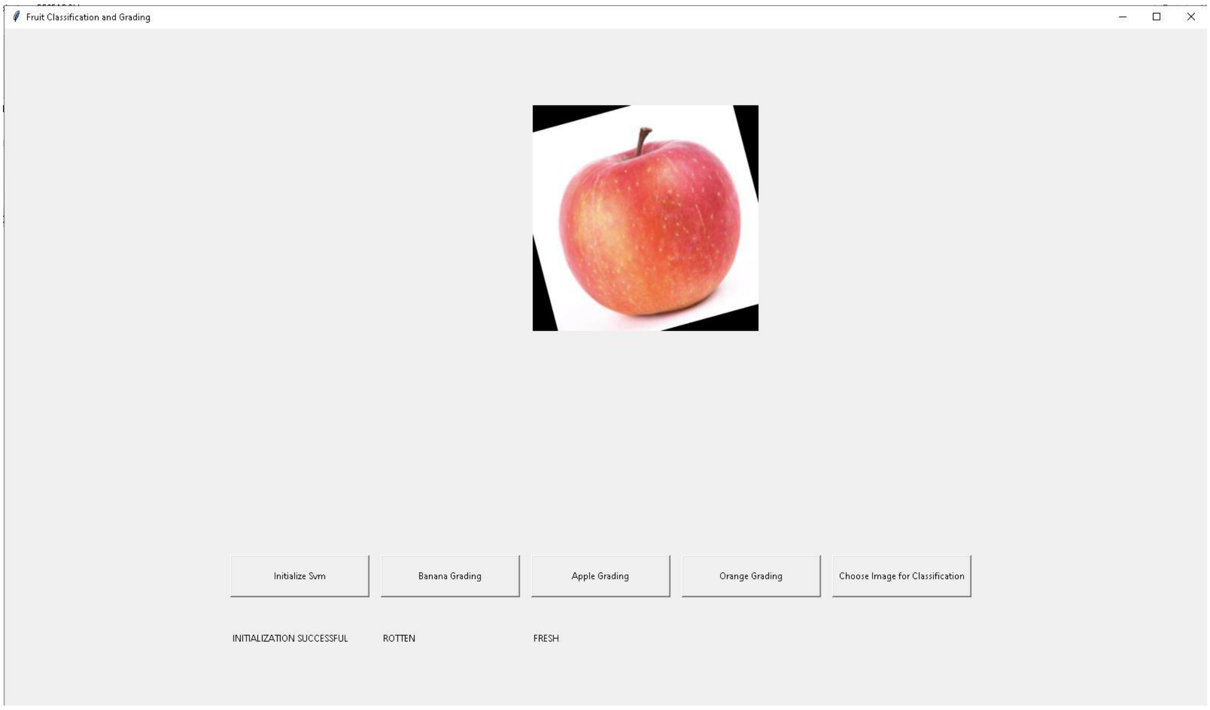
9.SVM grades apple as rotten as shown below.



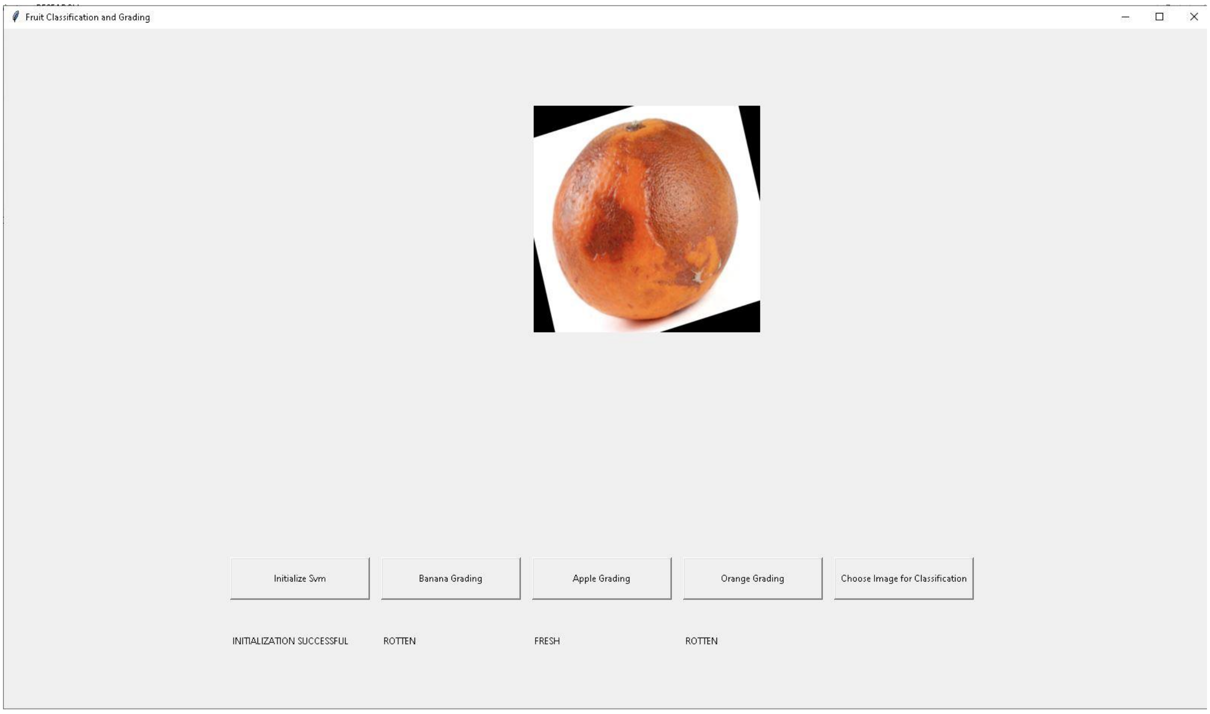
10.SVM grades apple as fresh.



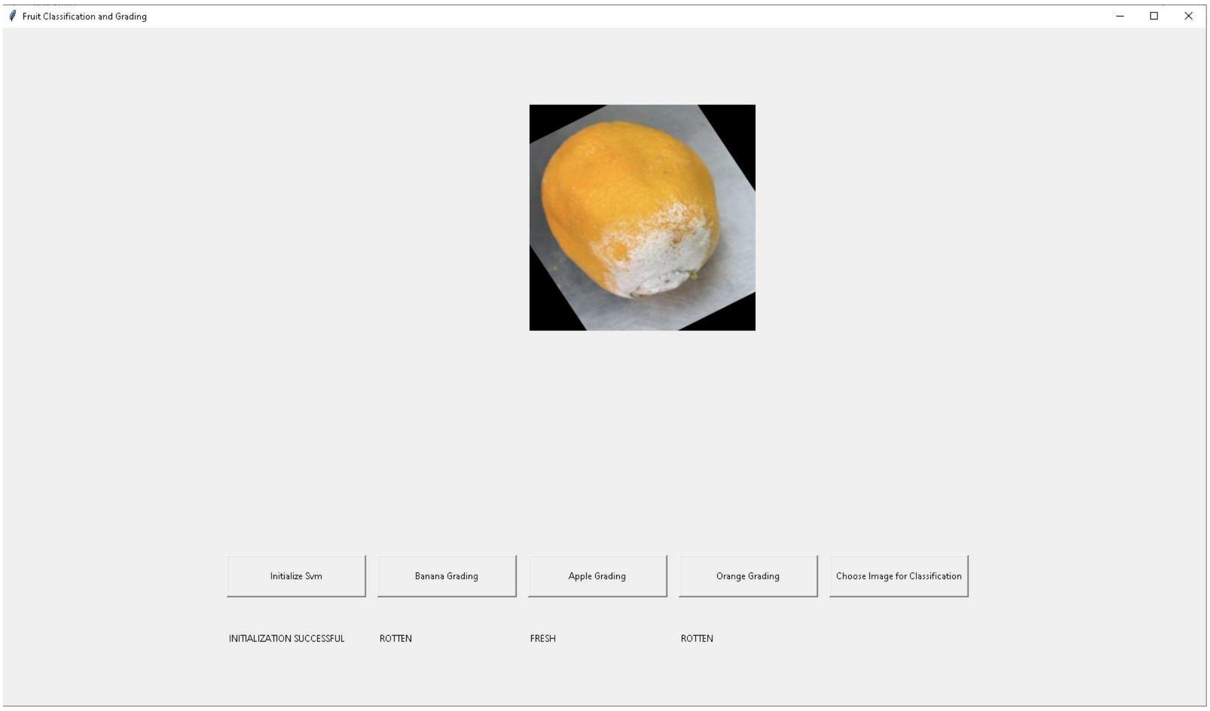
11. SVM grades apple as fresh as shown below.



12. SVM grades orange as rotten.

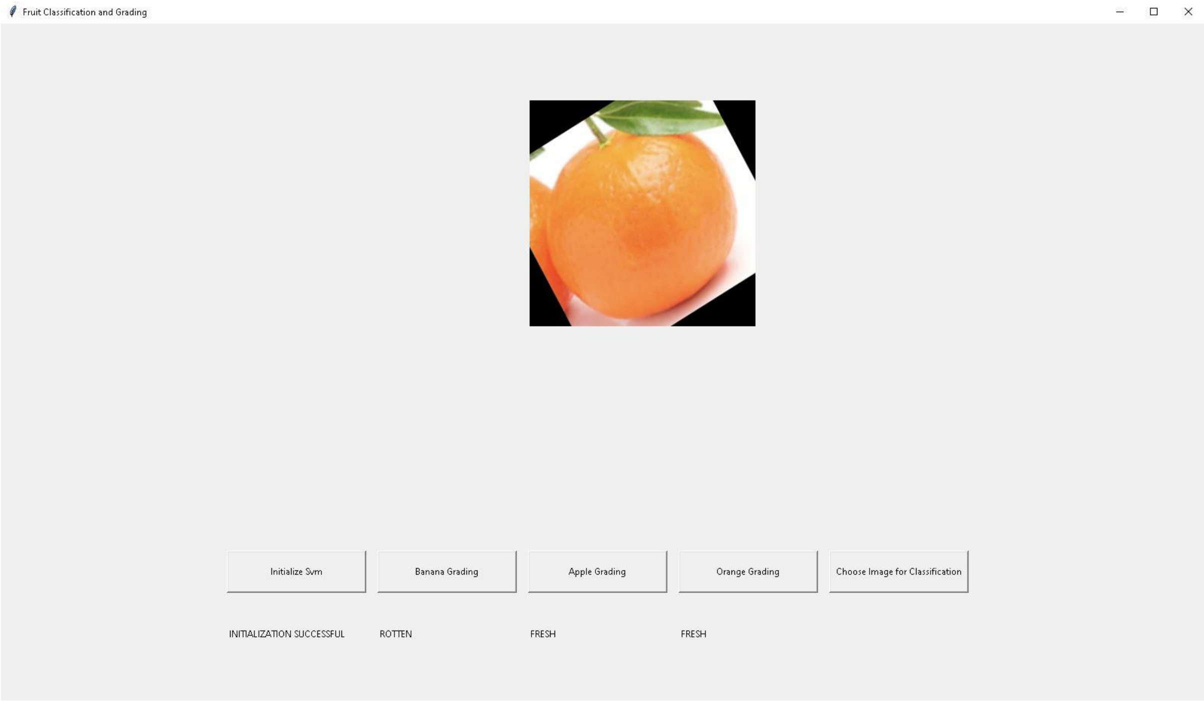


13. SVM grades orange as rotten.

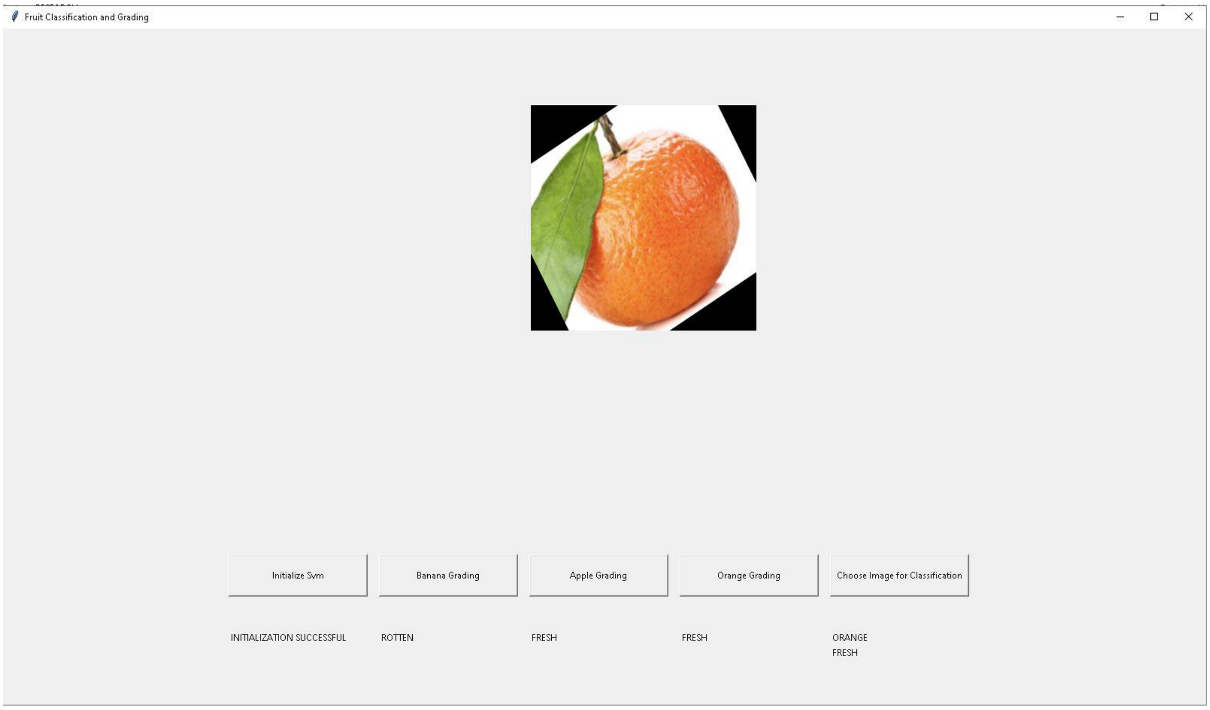


14. If multiple oranges are given, it successfully identifies them as oranges and grades them as fresh as the picture depicts.

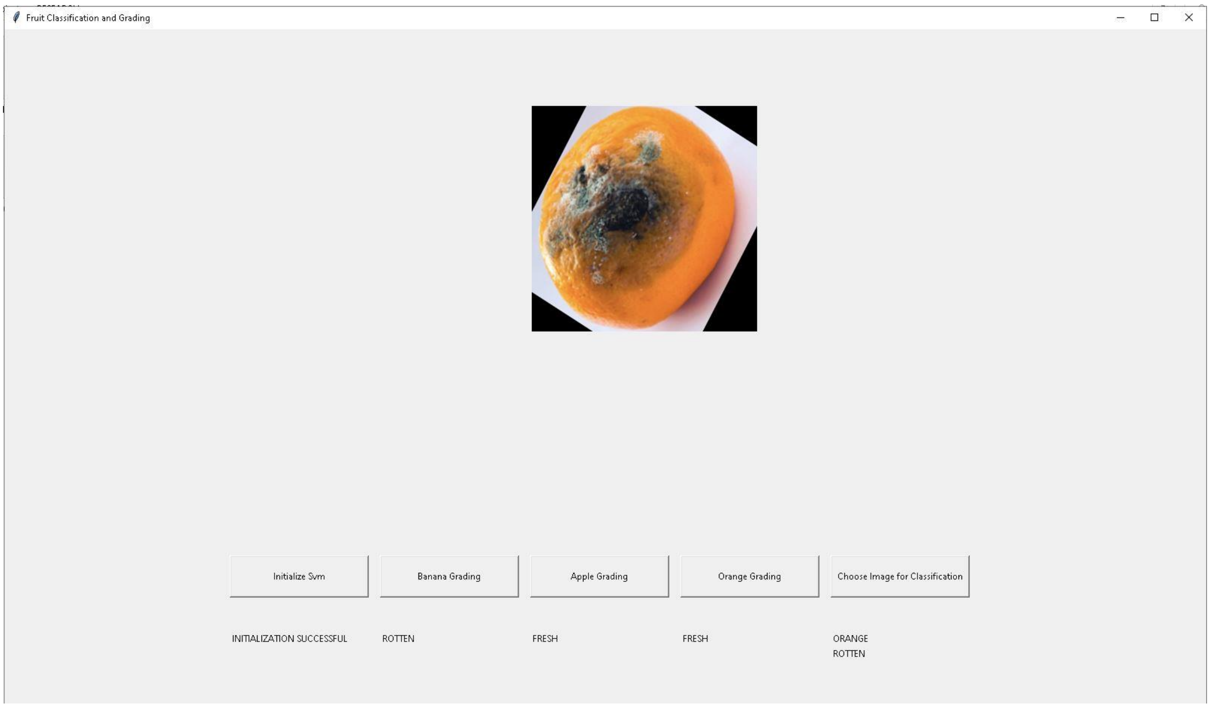
15. SVM successfully grades orange as fresh.



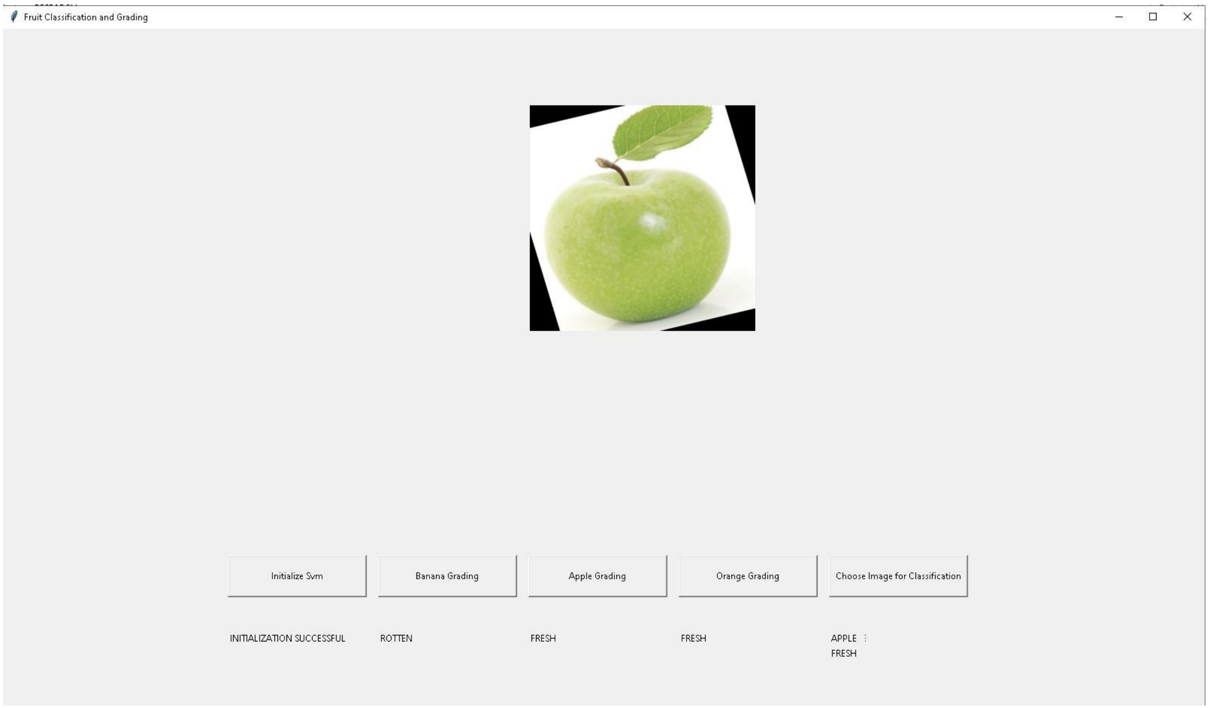
16. SVM successfully grades orange as fresh.



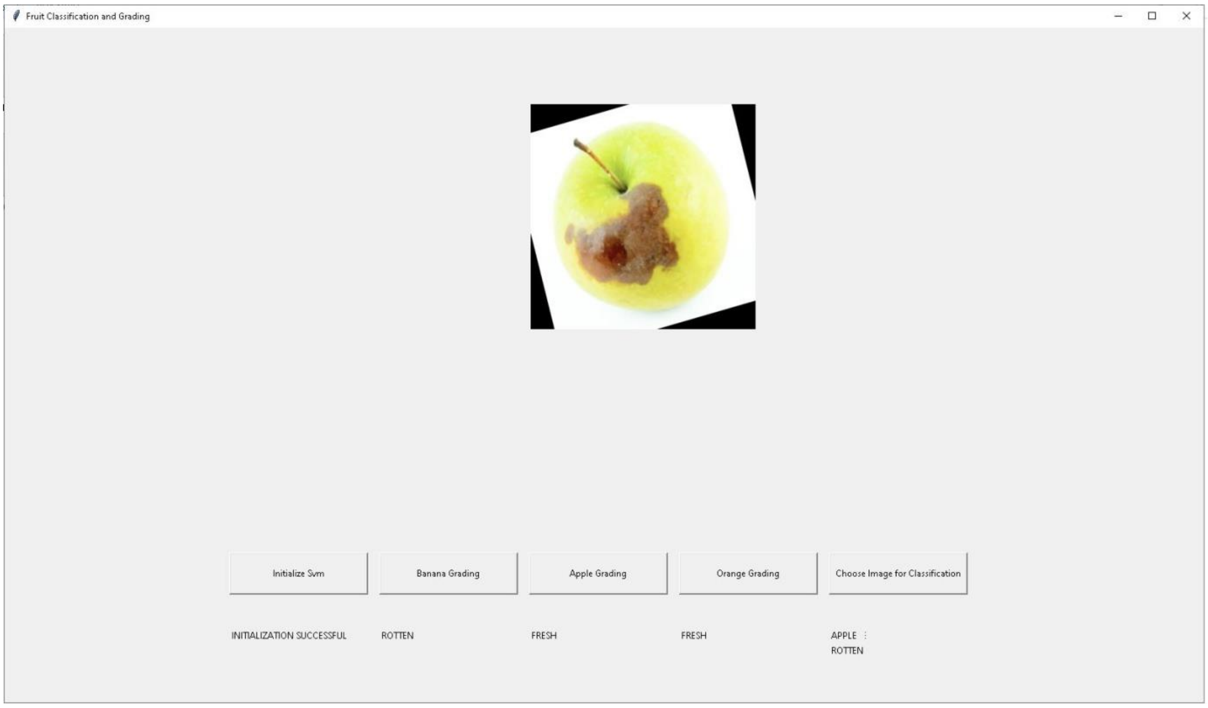
17. SVM successfully grades orange as rotten.



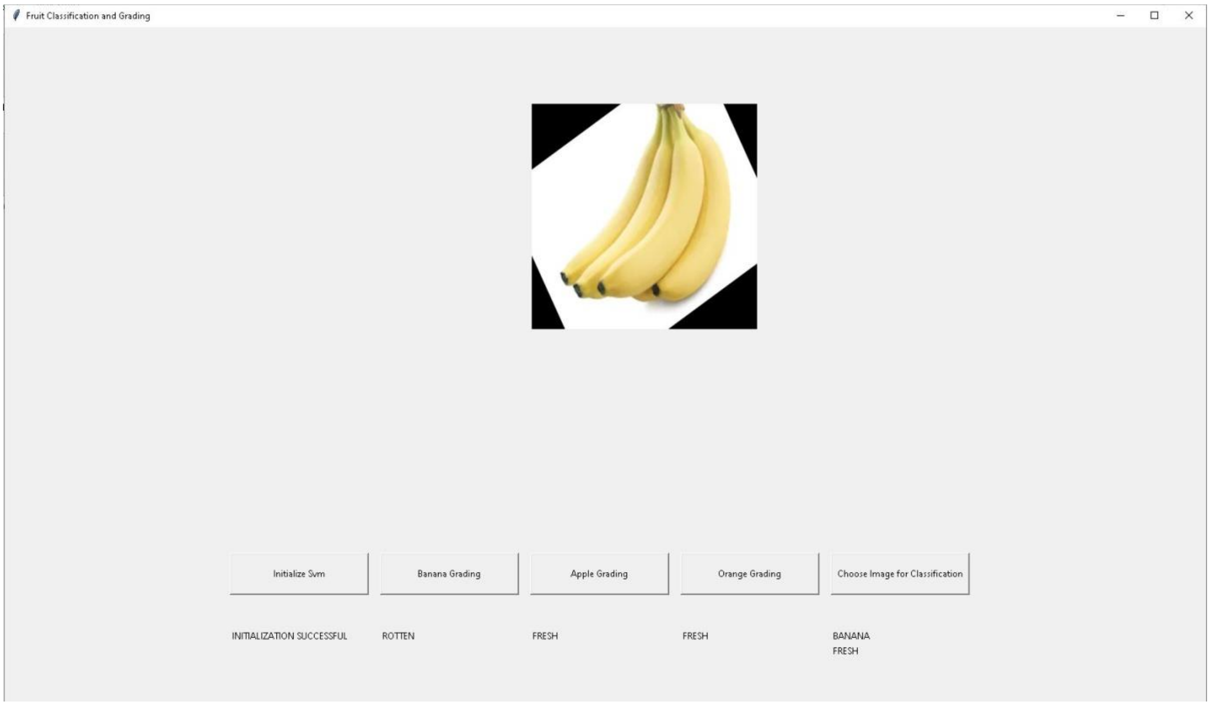
18. SVM successfully identifies green apple and grades it as fresh.



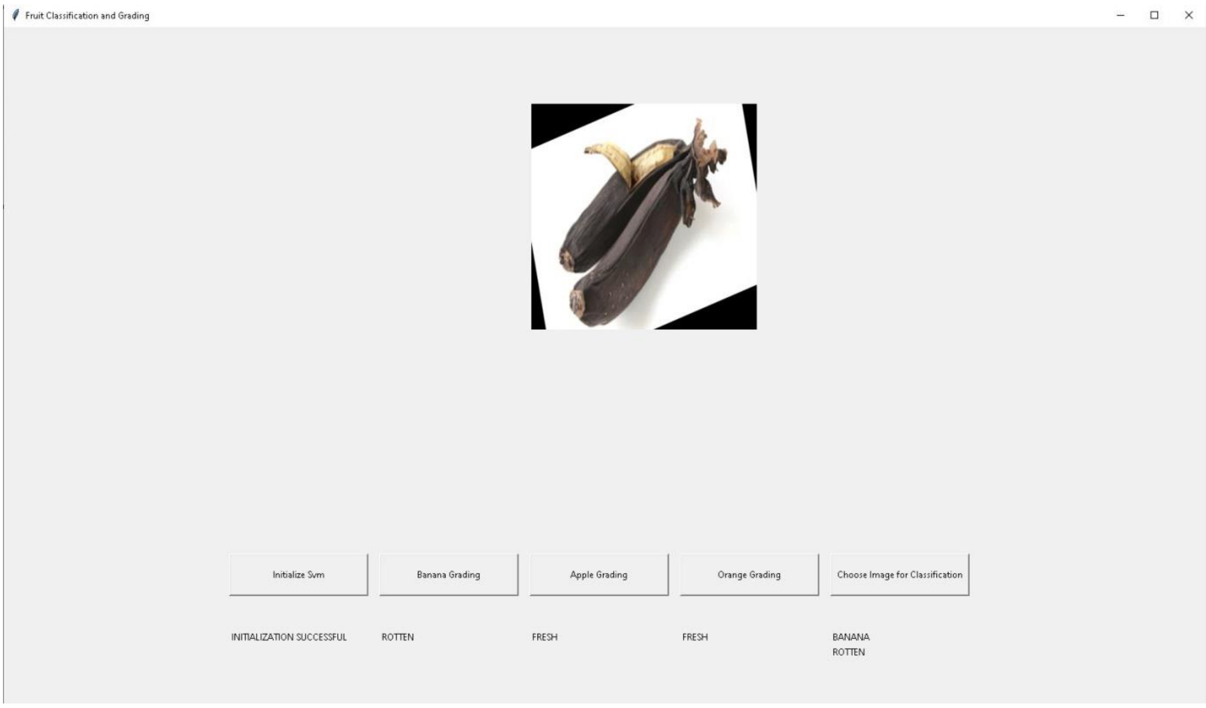
19. SVM grades apple as rotten.



20. SVM grades bananas as fresh.



21. SVM grades bananas as rotten.



9 CONCLUSION

The project was successfully able to distinguish among apples, bananas and oranges and is also able to grade each fruit into fresh and rotten category as desired. Classification using SVM attained an accuracy of 90% whereas the grading of banana, apple and orange acquired 99.1%, 89.7% and 93.8% of accuracies respectively. Various algorithms were used to train the model, but SVM attained the maximum accuracy of around 91 percent in classification.A classifier is contemplated with three different databases. First database consisted of 4035 apples out of which 1693 were fresh and 2342 were rotten. Bananas database comprised a total of 3805 images from which 1581 were fresh and 2224 were rotten. The last database incorporated 3061 images of oranges out of which 1466 were fresh and 1595 were rotten.

10 REFERENCES

1. AnujaBhargava ,Atul Bansal : Fruits and vegetables quality evaluation using computer vision: A review.

2. Anuja Bhargava &AtulBansal;Received: 23 August 2019 /Accepted: 9 December 2019 : Automatic Detection and Grading of Multiple Fruits by Machine Learning.

3. José Antonio Alvarez-Bermejo, Cynthia Giagnocavo, Li Ming,Encarnación Castillo Morales, Diego P. Morales Santos, Yang Xinting :Image processing methods to evaluate tomato andzucchini damage in post-harvest stages.

4. Luiz Fernando Santos Pereira, SylvioBarbon Jr., Nektarios A. Valous,Douglas FernandesBarbin : Predicting the ripening of papaya fruit with digital imaging and randomforests.

5. NorhashilaHashim, Segun Emmanuel Adebayo, KhalinaAbdan, MarsyitaHanafi : Comparative study of transform-based image texture analysis for theevaluation of banana quality using an optical backscattering system.

6. Sergio Cubero&NuriaAleixos& Enrique Moltó& Juan Gómez-Sanchis& Jose Blasco : Advances in Machine Vision Applications for AutomaticInspection and Quality Evaluation of Fruits and Vegetables.

7. Subo Tian, ZifanWang, Jifeng Yang, Zichen Huang, RuiliWang,LipingWang and JianghuiDong : Development of an automatic visualgrading system for grafting seedlings.

8. Zou Xiaobo* Zhao Jiewen Li Yanxiao, Shi Jiyong, Yin Xiaoping*Agricultural Product Processing and Storage Lab, Jiangsu University, 212013, China*<u>*zou_xiaobo@ujs.edu.cn*</u> *:* Apples Shape Grading by Fourier Expansion and Genetic Program Algorithm.