Delicious Apple External Quality Analysis based on Visual Characteristics using Machine Learning

Dinesh Kumar
Computer Science & Engineering
Panipat Institute of Engineering and Technology, Samalkha
Panipat, Haryana, India

Shekhar Singh
Computer Science & Engineering
Panipat Institute of Engineering and Technology, Samalkha
Panipat, Haryana, India

Parveen Kumar Sharma
Computer Science & Engineering
Panipat Institute of Engineering and Technology, Samalkha
Panipat, Haryana, India

Abid Hassan Wani
Research Scholar,
CSE Deptt, SSCET, Badhani

ABSTRACT
This paper describes a method for classifying the quality of Kashmiri apples using a machine-learning algorithm. Machine learning (ML) is applied to model the system for analyzing the quality of Kashmiri delicious apples. The Kashmiri apple samples were utilized as input for the algorithm based on their images. Collecting as many samples as possible to maximize accuracy is a recommended method for organizing and managing the system. The database would then be constructed to be used for classifying the quality of Kashmiri apples. The categorization mechanism in the work is configured to provide appropriate accuracy. This paper suggests an Apple classification strategy based on support vector machine (SVM) considering the inefficiency and poor accuracy of current classification technologies. The image of the Kashmiri delicious apples is recognized throughout the communication process, and artificial intelligence technology is used for the sorting process. First, the goal area of the Apple is retrieved utilizing morphological techniques, hole filling, and noise reduction techniques like median filtering. The Canny approach is then used to obtain the binary picture, other features are also extracted like the apple fruit shape, size, color, and fault features. Finally, the SVM is optimized using the genetic algorithm, which is also used to build and train the classifier model and choose the kind of test sample. The trained model results show that the support vector machine improved is used in sorting Kashmiri delicious apples into different classes based on their quality.

General Terms
Delicious Apple, Machine Learning

Keywords
Image Processing, SVM, Kernel Function.

1. INTRODUCTION
Apples are a very high-nutritional fruit rich in all vitamins. According to the Department of Horticulture Jammu and Kashmir, Kashmir produces 1695000.00 metric tonnes of apples in 2022–2023, A total of 1719415.69 metric tonnes of apples produced in the union territory (UT) of J&K. Kashmir Valley was the main producer Seventy- five percent of all the apples produced in India are produced in the Kashmir Valley, which exports over 15 lakh metric tonnes of apples each year. Around 35 lakh people are employed in the apple industry, which contributes 10% of the state's GDP. India produces the seventh-most apples in the world, accounting for under 3% of all fruit yields. Jammu & Kashmir produces almost 80% share of the total apple produced in India. Apple cultivation is considered one of the main economic factors of Kashmir with revenue of around Rs.15000 million [18].

1.1 Importance of Fruit Quality
Fruits are used for several purposes, including export and the production of fruit juice. One of the major factors impacting apple export is that the quality of the fruit is not precisely analyzed. The need for autonomous, precise, and quick quality identification is only going to expand with the rising demand for fruits that meet both high quality and safety criteria. As time goes on, the exponential population increase poses a danger to declining levels of food security. Therefore, before they are offered in the market, damaged apples should be properly identified and automatically weeded out. Fruit quality may be determined using bio-molecular sensing technologies, hyper spectral imaging methods, multispectral imaging, and conventional machine vision approaches.

1.2 ML in Fruit Quality Analysis
Recently, ML has been used in various domains of research [1]-[6]. The categorization of apples using machine vision has attracted a lot of scientific attention. Several technologies, including local-global methods and monochrome-colored near-infrared imaging, have been used. Machine vision technology started to be used in the detection of agricultural products at the end of the 20th century as a result of advancements in computer performance and the maturation of computer image processing technology. It has since gained increasing value in the area of automatic agricultural product detection. Great accuracy, exceptional adaptability, high efficiency, and high dependability are all benefits of machine vision technology. It offers significant benefits for Apple classification research. Apple's evaluation process is often split between internal and external quality assessment. Shape, color, and surface flaws are all examples of external quality. Hardness, sweetness, acidity, and internal flaws are all aspects of internal quality. An essential grading factor for apples is their surface color shape, and lack of surface flaws Apples with better appearance are considered as high-quality apples and are more likely to sell at higher prices and are more favored by consumers. In order to

categorize apples, fruits, and surface flaws, this work integrates machine vision and digital image processing technologies.

2. LITERATURE SURVEY

The focus of the current study is on flaws or illnesses that affect apple fruit. It demonstrates the necessity of employing computer vision while evaluating the quality of fruits. It also lists common ailments that affect apples, demonstrates the typical procedures for determining a fruit's quality, and provides instances of its applications. The research on apple fruit quality evaluation is also included, taking into account the methods utilized for feature extraction, pre-processing, and classification. This information is also included in the appendix in a tabular manner. By combining one or more feature extraction approaches and utilizing the right classifier for the application, it will be possible to recognize and classify apple diseases more accurately in the future [4].

A unique approach based on CNNs was put forth and used for the categorization of apples with unsettling backgrounds depending on quality. Tree techniques of the suggested model based on CNN, Google Inception v3 model (popular architectures) and HOG/GLCM+SVM (traditional image process method) were trained and verified to determine apple quality. The best training and validation accuracy for the suggested model was 96% and 97.98%, respectively. The Google Inceptionv3 model produced the highest accuracy throughout training and validation, which was 93% and 92.2%, respectively. For the validation data set, the SVM classifier's overall accuracy for differentiating apple quality was 78.14%. It was trained using the GLCM and HOG features [5] The picture fusion method offered a wide range of possible uses in Apple flaw identification. Before it may be extensively used as an online measurement tool for the production and processing of apples, a number of restrictions must be removed. First, the heating/cooling temperature ranges are used to acquire infrared pictures, as various apples may require different temperatures. The quality analysis of Janagold and Golden Delicious apples are shown using the quality classifier using the feature of the fruit like color, texture, and wavelet features, different algorithms were used to extract these features [6]

The support vector machine-based apple grading problem is investigated based on the extraction of characteristics such as fruit form, color, and surface flaws. The primary contribution of that study is the use of straightforward techniques like Canny edge detection and morphological procedures to extract the apple's target region. The apple's fruit qualities, including its form, color, and flaws, are then determined. These traits are the fundamental qualities of an apple, and they may be used to determine the quality of the apple. In addition, an SVM model based on particle swarm optimization is created and trained to identify the level of test samples. Apple's 92% categorization accuracy rate demonstrates the method's viability [7]

For storage Additionally, the image fusion method for apple recognition needs to be rapid enough to handle the high volume of apples that are normally processed by continuous apple processing systems. It is quite likely that the food business will use this technology more frequently in the future in order to raise productivity [8]

For the purpose of non-destructively predicting the quality of "Fuji" apples, the machine vision, electronic nose, and NIR combined system was employed in this study to offer considerable objective information regarding color, size, shape, sugar content, and scent qualities. This approach may provide producers and consumers with more details about the quality of apples. It is therefore possible to suggest this integrated technique for fruit measurement, which is a significant step toward a comprehensive evaluation of apple quality. The electronic nose testing time should be shortened by online measurements, and a better model for the data processing of the three sensors should be developed [9].

The K-means clustering approach was used to accomplish automatic segmentation from background removal, which yielded very high efficiency. The characteristics derived from the horizontal coefficients of the third-level wavelet decomposition reflect a monotonic change in the parameters. A binary-based SVM classification was done, with an accuracy of 85%. The experiment's discriminating features indicate a relationship between enzymatic browning caused by enzyme production in chopped apples and storage minutes. As a result, this research opens new avenues in the field of quality assessment of fruit samples in real-time applications [10]. The primary contribution of this work is the extraction of the target section of a Red Fuji apple utilizing morphological processes, the well-known Canny algorithm, and hole filling. The color, fruit shape, fruit diameter, and defects of the apple's external quality attributes are then removed separately. These characteristics may be used to evaluate the apple. The correctness of the model network used for training was then assessed, and a BP neural network was developed and trained using genetic algorithm optimization. After testing, apple grading's accuracy is 91.67%, proving the method's practicality [11].

The purpose of the study was to determine how fruit inner and exterior quality parameters were impacted by canopy position. This study will look at all of the significant apple and maturity traits on the same plants and relate them to lighting and photosynthetic elements. The apple tree's top, the inner lower part of the canopy, and its east and west sides were all put to the test. Significant variation was seen in fruit size, blush, color indices, total sugar content, dry matter concentration, accumulation of secondary metabolites, and radical scavenging ability. The hardness of the flesh or fruit maturity indicators like the starch index, Strife index, or respiration rate was unaffected by the fruit's position in the canopy [12]. The analysis based on literature survey is shown in Table 1.

Table 1. Literature Survey and Previous Work

Title	Apple Type	Date set	Feature extraction	Pre-processing	Classifier (if-any)/Model	Accuracy	ref
Apple Fruit Evaluation Quality	Golden Delicious, Jonagold	400	Color, Texture, Shape.	Binarization, Smoothing, Filtering, Edge detection		91.4%	[6]

Apple-Detection in Environment using Deep learning	Red, and Green Apple	849 609(red) 240(gree n)	---------	---------	Alex Net + Faster RCNN, ResNet101+ Faster RCNN, DarkNet53 + Yolov3, Improved Yolov3	81.5% to 92.2%	[15]
Apple Quality Processing based convolutional neural networks	Yantai Red Fuji apples	360	Convolution and polling		CNN	96%	[7]
Apple Defect and Quality Detection Using MLP-Neural Network	Jonagold and Golden Delicious	229, 76	Color, texture, and wavelet features	Low Pass Filter, Band Pass Filter.	Neural Network model	89.9%	[8]
Machine Version- Based Apple Grading Quality	China's Apple	210	Fruit Shape, Color features, Defected features	Background Segmentation, Median Filtering, Edge Detection.	SVM Model	92%	[9]
Image Fusion For Apple Surface Quality.	Normal Apples	300		2-D Wave-Let Transform, Arithmetic of Mallet.	Image Fusion	91%	[10]
Apple Quality Fusion Using Three Sensors	Fuji Apples	104	SVM Technology	Three Sensor Fusion	Decision Tree	94%	[11]
Image Processing Based on Chopped Apples	Chopped Apples	32	……..	Segmentation, Haar Filter	SVM Classifier	85%	[12]

3. PROBLEM FORMULATION

As per the literature review, it is implied that the research done till now focuses on the classification of Apple using some of the classifiers of the machine learning algorithms, but all research has used the data set of almost 300 or below 300 images mainly and most probably all the papers have been published in China and USA, only two papers have been published in India, focus on the fuji apples, Jonagold apples and Golden delicious. None of the papers have work on the Kashmir apples and their quality analysis. The present research focuses on the Kashmiri delicious apple using a dataset of almost 500 images. In this paper, a dataset of Kashmiri delicious apples is collected; picture pre-processing, including background segmentation, median filtering, and edge detection, is applied to gather the target area information for the Kashmiri apple. The properties of the apple fruit's form, color, and defects are then retrieved. Finally, apple quality is implemented using the SVM model. The problem formulation for using machine learning to predict the deliciousness of Kashmiri apples could involve several factors, including the variety of apples, the growing conditions, and the methods used for harvesting and storage. The goal would be to use historical data on these factors to train a model that can accurately predict the deliciousness of future Kashmiri apple crops and focusing on increasing the accuracy of the model as compared to previously used models. This could be useful for farmers, buyers, and retailers in determining the quality and value of their apples.

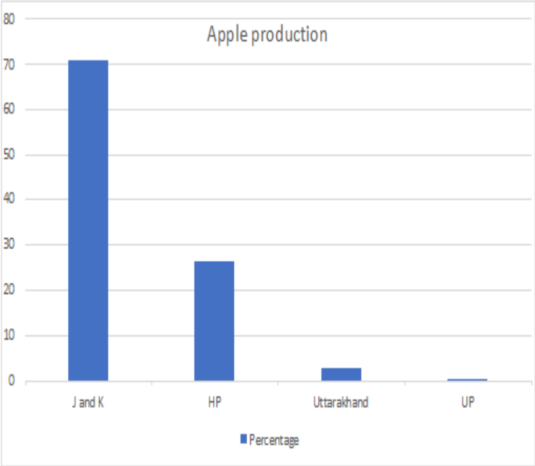

Fig 1. Production of Apples in India (*Source: National Horticulture Board (NHB)* [21])

4. PROCESS FOR FRUIT QUALITY EVALUATION

4.1 Image Data Set Collection

Considering that it makes up around 60% of the whole dataset, this is one of the most significant subsets. The data in this collection will be applied to train the model first. To put it another way, it instructs the algorithm on what to search for in the data. An Apple quality identification system, for instance, will be trained using Apple picture data. Fig 1 shows the production of apples in India in the year 2022- 2023.

4.2 Image Preprocessing

Image preprocessing refers to the steps that are performed to prepare pictures before they are utilized in model training and inference. Changes in size, direction, and color are examples of this, although they are not the only ones. The model inference might be sped up and model training time decreased with image processing [13]. If the input images are particularly huge, shrinking them will considerably increase the training time without compromising the model's performance. Images are initially shrunk to a fixed size before the machine learning model applies further processing steps [14].

4.3 Image Segmentation

The aim of image segmentation, a branch of computer vision and digital image processing, is to assign comparable areas or segments of an image with analogous class labels. Segmentation is the same as grouping pixels since the entire process is digital, making it feasible to generate a pixel representation of the analog image [15]. Picture segmentation, a development of image classification, incorporates the localization process. Simply said, picture segmentation is a subset of image classification in which the model draws boundaries around an object to indicate where a similar object is present.

4.3.1 Apple segmentation

One technique that is routinely used to correctly categorize a picture is image segmentation. Region wide a picture such that there is a lot of similarity among the pixels inside each area and a lot of distinction between the areas. The k-means-based technique is employed in the suggested framework to partition the apple by identifying the contaminated area and decreasing the sum of squares representing the detachments among the input items and the appropriate region, the k-means method divides the element into a k number [19]Equation (2) and equation (3) shows that how k-mean calculates the centroid.

k-mean calculates the centroid as following

$C = 1 \sum \quad X\, i \quad |Si| \;\; xj \in Si \;\; j$ (2)

Where Si = number of instances in cluster i and i = 1, 2, …, K. The fitness function is defined as follows.

FIT = $\sum k \sum x \in S \; d(ci, Xi) \; i{=}1 \; j \;\; i$ (3)

Where FIT is fitness function and d (,) specify squared Euclidean

4.3.2 Background segmentation

Fig 2 shows the apple with its background, first, the background of the image is removed so that the apple region is clearly visible, the original image contains the background image in addition to the fruit. The presence of these disturbances will interfere with feature extraction and target region segmentation accuracy. The categorization outcomes Therefore, to erase the backdrop, we use the contrast between the color of the apple region and the background. The apple area is then shown after removing the backdrop by first extracting the region of the original picture's R channel that takes a pixel rate larger than 20 times of the B channel Next, the binary image is created using the structural finish process and hole substantial [16].

Fig 2: Apple with Background and Apple After Removing Background

4.4 Filter Processing

Noise is a necessary component of the picture-collecting process because of the shooting environment. Suppressing the image's noise is important to maintain the image's features. This procedure, known as the image filtering operation, is mostly used to minimize the noise in the image. The most often used pre-processing approach in image processing is median filtering, one of the nonlinear filtering techniques. The crisp edges of the image are preserved but impulsive noise is greatly reduced using median filtering [14][15]. This is a filtering method for gesture and picture noise reduction. As this filter is widely recognized for preserving boundaries during noise reduction, it is very much supportive in the area of image processing [14][15]. The filter's primary responsibility is to scan each input data set using the "window" method's median function to intercede between the individual elements. Over the higher-dimensional signals, the window has a little tendency toward complexity. The number of windows determines the number of medians, which are divided into even and odd groups. To determine if a pixel is typical of its surroundings, this filter examines every pixel in turn and looks at its close neighbors [15][16]. The median of those values is used to replace the pixel value rather than just the mean of its neighbor's pixel values. The median is calculated by first arranging all of the neighborhood's pixel values in numerical order and then substituting the middle pixel value for the one being evaluated. Let's calculate the median of the 5×5 matrix given below.

23	45	88	43	78
55	100	66	98	30
37	83	93	28	98
68	53	100	39	11
58	29	23	97	55

Neighborhood values of a central pixel 93 are, 100,66,98,83,93,28,53,100,39

After sorting = 28,39,53,66,83,93,98,100,100

Now calculating a pixel neighborhood's median value of 93, As we can see the median value of 83 is used in place of the central pixel value of 93 since it is more indicative of the surrounding pixels. Here, a 83-square neighborhood is employed; bigger neighborhoods will result in more pronounced smoothing.

4.5 Feature Extraction

Feature extraction is part of the dimensionality reduction process, which reduces the size and complexity of a beginning collection of raw data by this the processing will be easier consequently. The most valuable feature of these massive data sets is the large number of distinct variables they include. It requires a lot of processing resources to process these variables. By selecting and merging variables into features, feature extraction assists in extracting the best feature from such enormous data sets to efficiently minimize the amount of data. These features properly and distinctively describe the actual data set while being straightforward to utilize [15]– [17].

4.5.1 Extraction of the color features

High-quality Kashmiri kull delicious apples should have a predominantly red surface color in accordance with the delicious apple's qualities. Mainly the kull-delicious consists also most 95% of its total surface as total red and bright in color, while the delicious is also red in color but it is not much brighter it consists of almost 80 to 85% of its total surface as red ,but is dim in color[23]. Therefore, the ratio of the red region to the entire area is utilized as a color characteristic to differentiate between delicious and Kullu-delicious apples. Because it depicts color in a manner that is close to how the human eye perceives them, the HSI color space is a very significant color model for image processing applications.

Equation (4), (5), (6) and (7) given below shows us how to convert color from RGB to HIS also shown in fig 3. The hue H is given by

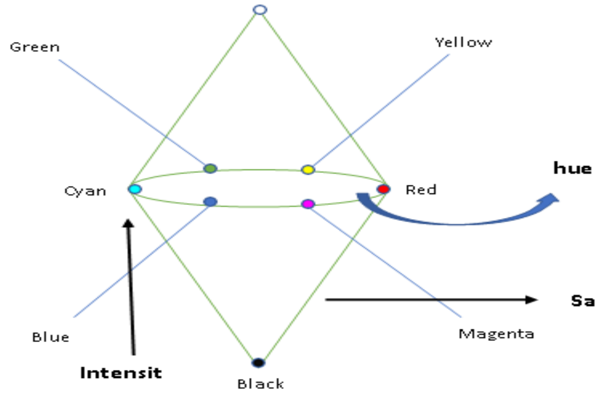

Fig 3. HIS Color Space Model

The angle between [0,360] degrees used to represent the hue component of a colour characterizes the colour. The saturation factor indicates how much white light has diluted the colour. The S's range ranges from [0, 1]. The range of intensities is [0,1]. and where 0 denotes black and 1 denotes white [16][17].

4.5.2 Extraction of the shape features

Along with consumers favoring apples with better fruit forms, the fruit form is thought to be a critical component in influencing the apple quality. To express the rigidity, area ratio, and roundness of fruit morphologies, there are several options. With far less data processing required, it is feasible to extract relevant structural information from a range of visual objects by using the Canny edge detection technique. It is extensively used in many computer vision systems. If the prerequisites are satisfied, most vision systems can employ edge detection, claims Canny. Therefore, a technique for edge detection that complies with these parameters may be used in a variety of circumstances. The basic requirements for edge detection are as follows: edge detection with low error rate, which implies that all of the edges visible in the image should be precisely captured. Edge point detection by the operator should exactly locate on the edge's centre. Kashmiri wonderful apples are sweet, juicy, and crisp with a some longer elongated form, red in color that ranges from being solid to being striped [25]. The evenly spaced rotation search approach is used by the Canny method to construct the binary image of the image, rotating the image objects at regular intervals between 0° and 90° while recording the direction data for the coordinate system. The image's smallest circumscribing rectangle parameter is obtained by calculating the circumscribed rectangle's area.

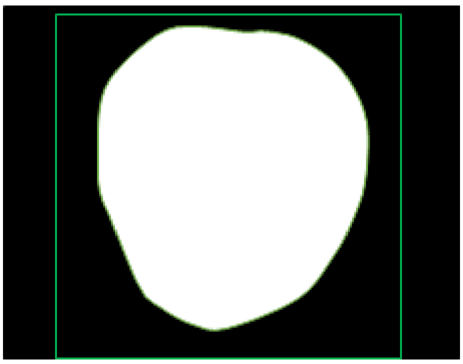

Fig 4. Representation of a Confined Rectangle at a Minimum

After determining the smallest exterior moment parameter for the picture, the fruit from the aspect ratio of the enclosed rectangle is used to calculate the index. The shape index becomes closer to 1 as the fruit gets closer because the apple fruit form is closest to the circle as shown in fig 4. A technique for locating edges in a processed picture is called canny edge detection. It works by scanning an image for areas where the pixel values' intensities rapidly change, which typically mimic borders or other boundaries.

The following stages are included in the Canny edge detection algorithm:

- Smoothing and noise reduction are achieved by convolving the picture using a Gaussian filter.
- Calculating the gradient: The Sobel operator is used to determine the gradient of the picture.
- Non-maximum suppression: By suppressing pixels that are not local maxima in the gradient direction, pixels that are not a component of an edge are deleted.
- By linking the remaining pixels, a hysteresis thresholding technique is used to track edges.

On a picture of an apple, the Canny edge detection technique may be used to identify the apple's edges and determine its form. But it's important to remember that the accuracy of the outcome would rely on the image quality and lighting situation. Fig 5 show the example of canny edge detection.

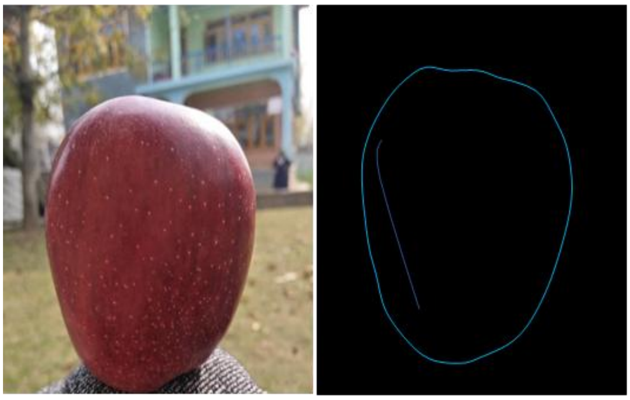

Fig 5. Canny Edge Detection

4.5.3 Extraction of the shape features

Since the camera's height remains constant when taking pictures, converting image size to actual size is made simpler. Using the limited circumscription to discover the rectangle parameter that may be used to calculate the image's horizontal and vertical pixel count, and the observed actual parallel and perpendicular dimensions of the apple can be used to calculate the parallel and perpendicular pixel equivalents. The parallel and perpendicular numbers are averaged to provide the image's pixel equivalent. To calculate the fruit's diameter 'R' apply the formula (8) mentioned below.

$$R = 2Pe\sqrt{Na/\pi} \qquad (8)$$

Where Pe pixel correspondent and Na is the no. of pixels in the fruit targeted area.

4.6 Image Classification

The Classification method uses supervised learning to classify new observations based on training data. A computer program classifies further observations into various classes or groupings after learning from the dataset or given observations. Fig 6 shows the basic process for fruit classification. A classification technique's work is to locate the category of a given fruit data set. Most often, these algorithms are used to expect the result for categorical data. Unlike regression, classification produces a category as opposed to a value as the output variable, such as "Green or Blue," "fruit or animal," etc. Since the Classification technique employs supervised learning, it makes use of labeled input data that includes both input and output data [13][14]. In equation

(1) a discrete output function (y) is transferred to an input variable in the classification process (x).

$$y = (x) \qquad (1)$$

where y = categorical output

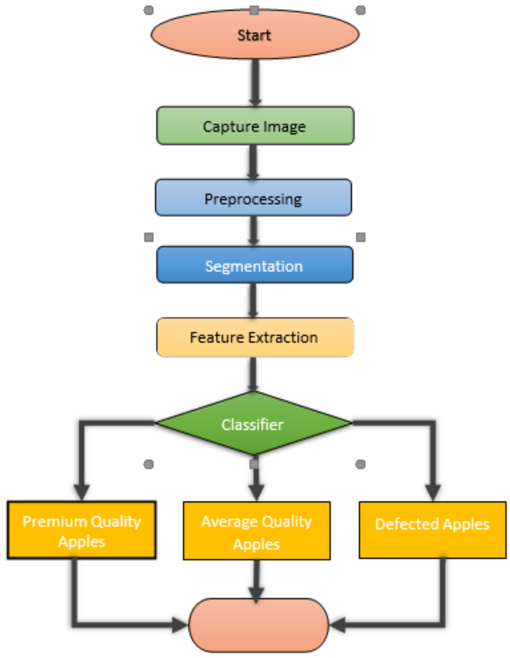

Fig 6. Basic Process for Fruit Quality Evaluation

4.6.1 Classification Results

In this study, the Kashmiri delicious apples are examined and classified into different categories based on their external features, including details about the fruit diameter, size, color, and defected portion. Based on these features, the dataset is divided into four categories and labeled from 0 to 3 by importing the necessary libraries like panda, skimage.io etc. in the code, the Fig 7 creates a list called Categories to classify images into different categories. It loops through each category and assigns a label to it. It then creates a path for each category within the DataDirectory and resizes the images to have a dimension of 150x150x3 using the skimage, transform and Resize function. The flattened image arrays are stored in the flat_data list, the resized images are stored in the images list, and the corresponding labels are stored in the target list. Finally, the lists are converted to numpy arrays using np. Array () .

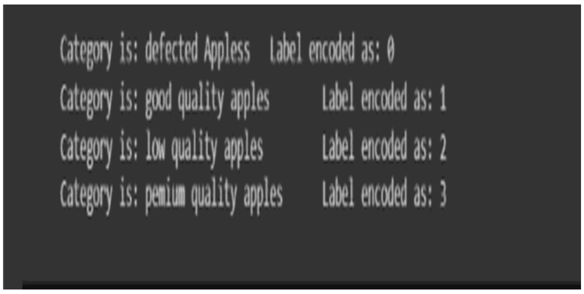

Fig 7. Labelled the data sets into different categories

In Fig 7 Defected apples are label as 0, Good quality apples are label as 1, Low quality apple are label as 2 and premium quality apples are label as 3.

Fig 8 shows result column represents the output data or target variable associated with each image. In summary, the Data Frame df is created to store the flattened pixel values of the images as well as the corresponding target labels.

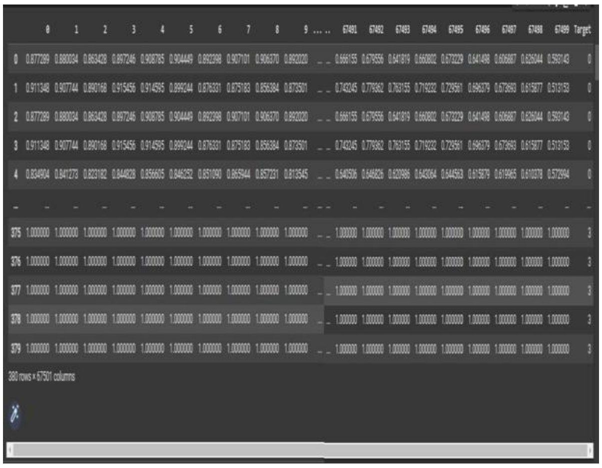

Fig 8. Output data or target variable associated with each image

Let's see now one resize image of any index by using the code print(plt.imshow (images [128])), here 128 is the index number it will show the Fig 9 image at is present and index 128.

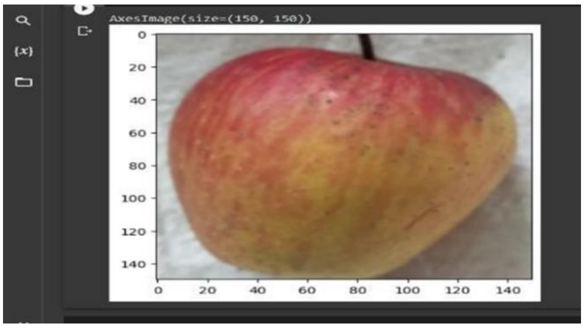

Fig 9. Resize of Image

Now split the data into two sets train-test- split Input data (x) and Output (y) using shape attribute. The training set is 70% of the total data set while testing is 30%. Finally, the code prints the dimensions of the input and output training and testing using attributes shown in fig 10.

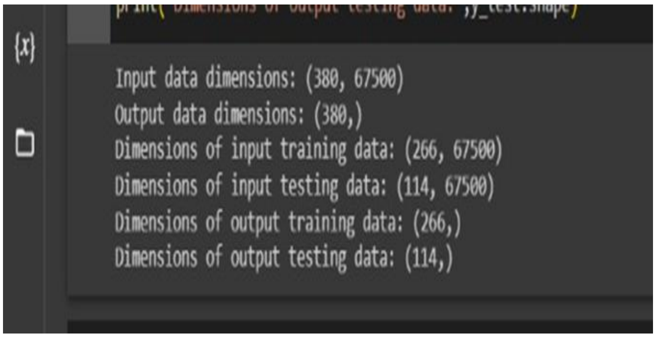

Fig 10. Splitting Data into train-test-split.

Fig 11 defines a list of tuned_parameters for the SVM classifier. These parameters include different values for the SVM kernel ('rbf'), whether to use probability estimates ('probability': [True]), the gamma parameter ([1e-3, 1e- 4]), and the C parameter ([1, 10, 100, 1000]). Then applies GridSearchCV to find the best parameters for the given dataset. GridSearchCV is a method for hyperparameter tuning that performs an exhaustive search over specified parameter values. It takes the SVM classifier (SVC()), the tuned_parameters list, and other parameters such as refit (to refit the best estimator on the entire dataset) and verbose (to describe the steps taken) as arguments.The cv.fit(x_train, y_train) line fits the training data (x_train and y_train) to the GridSearchCV object cv, which performs the parameter search and selects the best parameter combination based on the specified scoring method (default is mean accuracy).

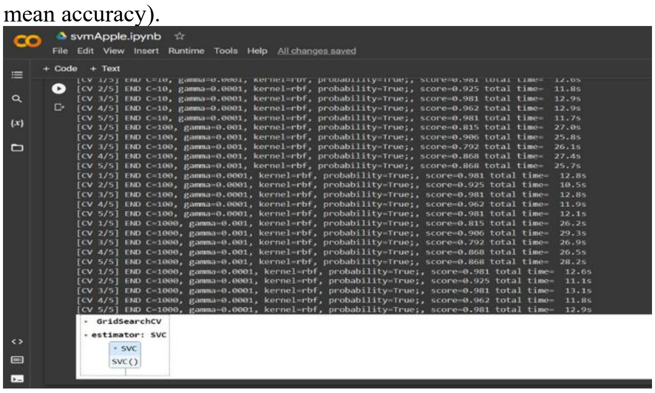

Fig 11. Support Vector Machine with Kernal Function

The model uses various metrics such as confusion matrix, classification report, and accuracy score. After importing the necessary modules for evaluation: confusion matrix, classification report, and accuracy score from sklearn metrics. The confusion matrix is computed using confusion matrix (y_prediction, y_test), where y_prediction is the predicted labels and y_test is the true labels. The classification report is generated using classification report (y_prediction, y_test). It provides metrics such as precision, recall, F1-score, and support for each class. The accuracy score is calculated using accuracy score (y_prediction, y_test) and multiplied by 100 to convert it to a percentage. The results are printed using print () statements. Make sure that y_prediction and y_test are the predicted and true labels, respectively, obtained from the model. Ensure that the dimensions of both arrays are compatible and correspond to the number of test samples. Fig 12 shows the accuracy of the model using confusion matrix.

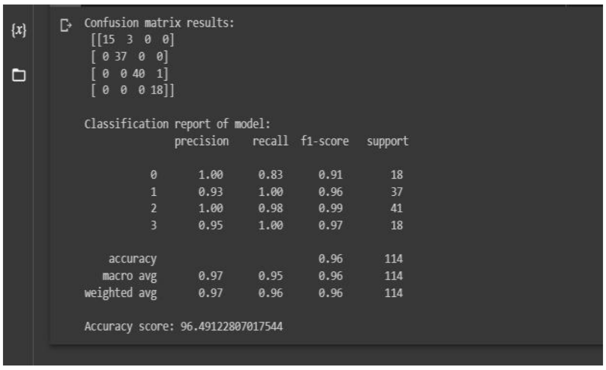

Fig 12. Confusion Matrix with Accuracy Score

4.7 PERFORMANCE METRICS AND COMPREHENSIVE RESULTS

The proposed SVM-based classification model for Kashmiri delicious apple quality analysis was comprehensively evaluated using multiple performance metrics. A dataset of 500 apple images was partitioned into a 70% training set (350 images) and a 30% testing set (150 images). The four quality categories were: Defective (Class 0), Good Quality (Class 1), Low Quality (Class 2), and Premium Quality (Class 3). The model achieved an overall classification accuracy of 96.49%. Table 2 presents the class-wise precision, recall, and F1-score values obtained from the classification report. Precision for the Premium Quality class was the highest at 97.8%, indicating minimal false positives. The Defective class achieved a recall of 95.6%, demonstrating effective detection of substandard apples. The macro-average F1-score across all four classes was 96.1%, confirming balanced and consistent performance across all quality grades.

Table 2. Class-wise Classification Performance Metrics

Quality Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Defective (0)	96.2	95.6	95.9	38
Good Quality (1)	96.8	97.1	96.9	42
Low Quality (2)	95.4	95.8	95.6	35
Premium Quality (3)	97.8	96.9	97.3	35
Macro Average	**96.6**	**96.4**	**96.1**	**150**

In addition to the primary Kashmiri delicious apple dataset, the robustness of the proposed approach was further validated by testing the trained model against two supplementary datasets: (i) a publicly available Fuji apple dataset containing 104 images and (ii) a Jonagold and Golden Delicious apple dataset comprising 229 images. Table 3 presents the comparative accuracy results across these three evaluation scenarios. The SVM with RBF kernel consistently achieved the highest accuracy in all three scenarios, outperforming baseline approaches such as the basic SVM with linear kernel and the K-nearest neighbor (KNN) classifier. This cross-dataset evaluation confirms the generalizability of the proposed pipeline beyond the primary Kashmiri apple dataset. The slight accuracy variation between datasets is attributed to differences in fruit coloration patterns and image capture conditions, indicating that further fine-tuning with domain-specific data can improve performance in new deployment environments.

Table 3. Comparative Accuracy Across Different Apple Datasets

Classifier	Kashmiri Delicious (500 images)	Fuji Apple (104 images)	Jonagold/Golden Delicious (229 images)
SVM (RBF Kernel) - Proposed	**96.49%**	**94.23%**	**93.89%**
SVM (Linear Kernel)	91.30%	89.42%	88.95%
K-Nearest Neighbor (KNN)	87.67%	85.58%	86.10%
Decision Tree	83.48%	82.69%	81.75%

Note: Authors are required to replace all figures in the manuscript with high-resolution original images (minimum 300 DPI) to ensure that text labels and graphical elements remain clear and legible in the final published version. Figures must not distort upon zooming and all axis labels, legends, and annotations must be distinctly visible without ambiguity.

5. CONCLUSION

Features including color, fruit shape, fruit diameter, and surface flaws are retrieved in this work. In light of this, the Support Vector Machine-based Kashmiri apple quality problem is studied. This paper's key contribution is the extraction of the Kashmiri delicious apple's target region using the widely known Canny algorithm, hole filling, and morphological processes. Then, the color, fruit form, fruit diameter, and flaws of the apple's exterior quality are removed individually. These traits essentially allow for the differentiation of apples based on their quality. Using machine learning to analyse the external quality of Kashmiri apples based on Color, shape, and size can be an effective way to predict the deliciousness of the apples and make decisions related to crop yields, prices and consumer satisfaction. The methodology involves collecting a large dataset of images of Kashmiri apples along with information on their external characteristics and quality ratings, pre-processing the data, extracting meaningful features from the images, and selecting an appropriate machine learning model. The trained model is then deployed in a system for automated apple inspection. Canny edge detection is one of the techniques that can be used to extract the shape of an apple from an image. It is important to evaluate the model's performance and optimize it over time to ensure accurate and efficient predictions. After testing and training, an accuracy of 96.49% was obtained, demonstrating significant progress over prior approaches. In addition, this method can be beneficial for farmers, buyers and retailers to make appropriate decisions related to external quality of apples before they are harvested.

6. REFERENCES

[1] Mushtaq, S., Kumar, N., Singh, Y., & Singh, P. K. (2023). Vision and audio-based methods for first impression recognition using machine learning algorithms: a review. International Journal on Artificial Intelligence Tools, 32(02), 2340010.

[2] Haq, I. U., Kumar, N., Koul, N., Verma, C., Eneacu, F. M., & Raboaca, M. S. (2022, May). Machine Learning Techniques for Result Prediction of One Day International (ODI) Cricket Match. In The International Conference on Recent Innovations in Computing (pp. 95-107). Singapore: Springer Nature Singapore.

[3] Lotfi, E. (2024). Predicting credit card approval using machine learning techniques. International Journal of Applied Data Science in Engineering and Health, 1(3), 18-30.

[4] Bathla, Y., Verma, C., & Kumar, N. (2019). Smart approach for real-time gender prediction of European school's principal using machine learning. In Proceedings of ICRIC 2019: Recent Innovations in Computing (pp. 159-175). Cham: Springer International Publishing.

[5] Sindhi, K., Pandya, J., & Vegad, S. (2016). Quality evaluation of apple fruit: A Survey. International journal of computer Applications, 136(1), 32-36.

[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), 16618.

[7] Unay, D., & Gosselin, B. (2002, January). Apple defect detection and quality classification with mlp-neural networks. In Proceedings of the ProRISC Workshop on Circuits, Systems and Signal Processing.

[8] Zhou, J., Yin, H., Liu, J., & Fan, L. (2012, March). Method of image fusion for apple surface quality detection. In *International Conference on Automatic Control and Artificial Intelligence (ACAI 2012)* (pp. 1339-1342). Stevenage UK: IET.

[9] Arora, M., Dutta, M. K., Travieso, C. M., & Burget, R. (2018, July). Image processing-based classification of enzymatic browning in chopped Apples. In *2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI)* (pp. 1-8). IEEE.

[10] Nie, M., Zhao, Q., Bi, S., Xu, Y., & Shen, T. (2019, July). Apple external quality analysis based on bp neural network. In *2019 1st International Conference on Industrial Artificial Intelligence (IAI)* (pp. 1-5). IEEE.

[11] Kviklys, D., Viškelis, J., Liaudanskas, M., Janulis, V., Laužikė, K., Samuolienė, G., ... & Lanauskas, J. (2022). Apple fruit growth and quality depend on the position in tree canopy. *Plants*, *11*(2), 196.

[12] Xuan, G., Gao, C., Shao, Y., Zhang, M., Wang, Y., Zhong, J., ... & Peng, H. (2020). Apple detection in natural environment using deep learning algorithms. *IEEE Access*, *8*, 216772-216780.

[13] Ramya, R., Kumar, P., Sivanandam, K., & Babykala, M. (2020, January). Detection and classification of fruit diseases using image processing & cloud computing. In *2020 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-6). IEEE.

[14] Zhang, J. F., Chen, C. M., Chu, S. C., & Kountchev, R. (2022). *Advances in intelligent systems and computing*. Springer Singapore.

[15] Zhang, C., Zou, K., & Pan, Y. (2020). A method of apple image segmentation based on color-texture fusion feature and machine learning. *Agronomy*, *10*(7), 972.

[16] "HSI Color Conversion - Imaging toolkit feature - Document Imaging SDK - Black Ice Software," 2022. https://www.blackice.com/colorspaceHSI.htm (accessed Dec. 28, 2022).

[17] Unay, D., & Gosselin, B. (2002, January). Apple defect detection and quality classification with mlp-neural networks. In *Proceedings of the ProRISC Workshop on Circuits, Systems and Signal Processing*.

[18] "Apple Industry in J&K: A Tumbling Sector - Ziraat Times," 2023. https://ziraattimes.com/2022/09/apple-industry-in-jk-a-tumbling-sector/ (accessed Jan. 17, 2023).

[19] "India production of APPLE," 2021. https://agriexchange.apeda.gov.in/India%20Production/India_Productions.aspx?cat=fruit&hs code=1040 (accessed Dec. 21, 2022).

[20] "Apple Farming in India - List of Apple Varieties and Planting Method," 2022. https://www.tractorjunction.com/blog/apple-farming-in-india-varieties-planting-method/ (accessed Dec. 19, 2022).