Machine Learning-Based Classification of Fruit Quality: Fresh and Rotten Fruit

Abstract

Fruits play a crucial role in many industries and are good sources of vitamins and minerals, identification of fresh and rotten fruit in the early stage is important to prevent contamination of other fruits and maintain economic stability during export. Industries depending on manual detection of fruit quality require an advanced classification model to reduce human effort and production time. The recent advancement in the field of deep learning has gained popularity for image-processing applications within the realm of computer vision, the classification of fresh and rotten fruits using deep learning is crucial for commercial and agricultural uses within computer vision. However, there are some challenges such as limitations of the dataset, the diversity of fruit types, and environmental factors. This study addresses challenges faced by earlier researchers by developing a methodology for Fresh and rotten fruit classification. In this research, a total of 18,492 fruit images are used, comprising three types apples, bananas, and oranges which are based on six classes. The machine learning models used for the project are Support Vector Machine (SVM), VGG16 Transfer learning, VGG16 with Edge Detection, and Multi-class classification using VGG16 with Edge Detection. VGG16 with Edge Detection performed better with a high accuracy of 97.96% compared to other models.

1 Introduction

In the field of data science, through the use of deep learning algorithms, computers can now identify and classify the objects seen in images and movies. In the 1950s, the idea of computer vision was introduced first when the development of a neural network in its initial stage showcased that it could classify objects based on the detection of edges. [1].The rise of the internet resulted in huge data characterized by the early phase of neural network development. During this era, neural networks showed their ability to object classification by detecting edges, which accelerated rapidly advancing this field. In a decade, these systems have advanced significant improvements due to the large amount of data that is generated each day. The growing demand for goods across all industries has led to an increase in automation, which in turn has accelerated the use of computer vision and its various applications. The effect of this technology is evident across various field that relies on machines for the analysis of images, Videos, and other forms of data. the technological effect is evident in numerous fields that depend on machines for the analysis of images, videos, and different forms of data. computer vision aims to handle complexities and achieve greater efficiency to overcome the limitations of traditional systems. Although there are many applications for computer vision, this study focuses on

[1] https://techsee.me/blog/computer-vision-applications

Machine learning and various deep-learning methods for classifying fresh and rotten fruit. .

The goal of everyone is to maintain safety and a healthy lifestyle. Hence It is very crucial to include fruits in the diet we take, since fruits are rich in nutrients like potassium, vitamin C, folic acid, and dietary fibers, which are vital for maintaining good health. Intake of fruits daily basis reduces the risk of diseases like high blood pressure, diabetes, and different chronic disorders. [2]. It is recommended to include at least one-fourth of the plate with fruits for a nutritious meal. [3]. the application of artificial intelligence and deep learning for fresh and rotten fruit classification introduces many possibilities in different domains. One such practical application in agricultural quality control systems is by using deep learning and computer vision, Farmers and producers can automate the harvested fruits sorting process by ensuring only fresh and good quality fruits are delivered to the marketBargoti and Underwood (2017). Another application is, the classification of fruits in automating inventory management in stores plays a vital role. by using computer vision, retailers can automatically classify and monitor the fruit freshness on shelves. this aids in waste reduction optimizes stock level, and improves overall efficiency.

Rotten fruits often show noticeable changes on their surface like changes in texture, shape, and color. Bad smell from fruits is a sign of rotting that commonly occurs in storage due to different factors such as temperature, moisture, air, light, and the presence of microorganismsAn et al. (2020) Lu Fae (2019). fruits also decay during the transportationSingh and Sharma (2018) Cao et al. (2019). Single rotten fruit can damage multiple fresh fruits resulting in enormous losses. Hence early detection of rotten fruit plays a crucial role in reducing such losses and enhancing food safety. Manual examination of fruits depends on color, shape, texture, and smell to detect if the fruit is rotten. However, Machine learning and computer vision are useful for automatic detection but can not test smell, it relies only on the changes in the surface of fruitsCao. (2021) turning the task into a challenging venture for researchers. Fruit recognition can be achieved through the application of deep-learning that consists of multi-layer feed-forward networks which is a widely used method for image processing techniques. This study focuses on the problem of fruit defect classification using deep learning techniques transfer learning VGG16 with edge detection.

The focus on these metrics not only proves the efficiency of the VGG16 architecture but also emphasizes the added value brought by including Canny Edge Detection through which the model improves sensitivity to key features, thereby enhancing its ability to distinguish subtle differences between fresh and rotten fruits. This combination of advanced techniques contributes to the overall success of the machine learning-based classification project, aligning with its objective of delivering an accurate and efficient fruit quality assessment system.

1.1 Research Question:

The classification of fresh and rotten fruit has gained increasing importance in research. Many studies are focused on the classification of defective fruits based on their freshness.

[2] https://www.betterhealth.vic.gov.au/health/healthyliving/fruit-and-vegetables

[3] https://www.myplate.gov/ten-tips-build-healthy-meal

However, researchers still face some challenges such as dataset limitations like size and quality constraints, the different nature of fruit types, defect variation, and the effect of environmental factors like lighting condition, variation in background, and image quality on the performance of the model. It is crucial to address these challenges to develop an efficient model suitable for various fields of the agricultural and food industry. Hence, considering these challenges, the following research question was formed.

RQ: How can we develop an accurate and efficient machine learning-based classification model to distinguish between fresh and rotten fruits?

1.2 Research Objectives:

The brief research has motivated to achieve the following objectives during the model development.

1. Diversify the training dataset with a wide range of fruits, and textures and use the data augmentation for improved model adaptability in a real-world context.
2. Improve the accuracy of fresh and rotten fruit classification model by advanced deep learning techniques like VGG16.
3. Combine accurate edge detection with transfer learning to enhance accuracy, loss, and robustness in identifying fruit defects.

This research aims to create an accurate machine learning model, employing VGG16 Transfer learning with edge detection. The multi-tasking framework used in this study involves two methods: one for binary classification of fresh and rotten fruit and another for multi-class labeling to identify the fruit type. The study's innovative approach explains the need for precise fruit quality classification, offering practical applications using advanced deep learning techniques like transfer learning VGG16 with edge detection.

To perform this investigation, a thorough analysis is carried out on the existing literature of fresh and rotten fruit classification in various works that identify current methodologies and gaps in Chapter 2. followed by the research methodology in Chapter 3 explains data collection, preprocessing, and the proposed binary, Multi-class Classification Approach. The design specification is described in Chapter 4, and Implementation in Chapter 5 offers a foundation for setting ideas into practice. Evaluation of results assesses the model performance in Chapter 6. Finally, chapter 7 concludes the document with model results and a projection of future work to further improve the study.

2 Related Work

Fresh and rotten fruit classification is a challenging task. To overcome these challenges it is important to create an effective method for fruit defect classification. this section explains various techniques used by researchers to address these challenges including statistical, machine learning, and deep learning techniques, The following studies show the research conducted from 2014 to the present. However, for each study, a thorough critical assessment is conducted by examining the challenges addressed by researchers, the type of data used, applied methods, practical implications, and future study of findings. the further sections are divided into four parts, each addressing the employed Techniques. 1) Statistical Techniques,2) Machine Learning Techniques,3) Deep Learning Techniques, 4) Critique, Identified Gaps

2.1 Analysis of Statistical Techniques:

The study by Zhang et al. (2015) focuses on the complex challenges of early detection of rottenness in apples for an automatic grading system, considering the similarities of spatial and spectral between rotten and sound tissues. this study implements an algorithm for detecting rottenness in apples by using hyperspectral imaging and spectral analysis. the study achieved a 98% accuracy on 120 apples by combining algorithms like successive projections algorithm (SPA), minimum noise fraction (MNF), and PLS-DA, proving its effectiveness in early rottenness detection and its use in monitoring real-time post-harvest storage of apples.

The study of detection and classification of common pathogenic fungal diseases in post-harvest strawberries in the early stage by employing the electronic nose (E-nose) and gas chromatography-mass spectrometry (GC-MS) conducted by Pan et al. (2014) to analyze the volatile compounds in infected fruits. E-nose effectively identified rottenness on the second day, identified different fungal infection types, and achieved an accuracy of 96.6% through a multilayer perceptron neural network.GC-MS identified specific volatile compounds for each infection. the research concludes that the E-nose helps in early diagnosis and accurate classification of fungal types, providing valuable understanding to reduce losses from fungal infections in strawberries.

The research by Nosseir and Ahmed (2019) for classifying and assessing fruit quality differentiating between fresh and rotten for four types of fruits using fruit image features like color and texture, the RGB values, as well as the first and second statistical order of the Gray Level Co-occurrence Matrix (GLCM) values, are extracted to achieve high accuracy with different classification algorithms. Tested with 46 seasonal fruit pictures, the system accurately detects fruit types. defected fruits are identified effectively from fresh ones by using the Support Vector Machine(SVM) method with accuracy rates of 96% and 98%.

2.2 Analysis of Machine learning Methods:

According to Bhargava and Bansal (2020), "Classification of different fruit types and identification grading of fruit is a challenging task due to the bulk production of fruit products", this study addresses these challenges. The proposed system uses machine learning for automatic detection and grading by utilizing image processing algorithms and extracting characteristics like color, statistics, and geometry. considering fruits like apples, bananas, avocados, and oranges, the study uses different machine-learning methods for classification. Support Vector Machine (SVM) is highly effective in achieving an accuracy of 98.48% for fruit detection. the system's performance with SVM provides the best results for classifying fresh and rotten fruit.

The study by Chandini et al. (2018), suggests a method of combining hardware, machine learning, and image processing for fruit quality detection, outperforming current methods. The system, connecting Raspberry Pi with MATLAB, achieves a classification accuracy of 85.64% using a multiclass SVM classifier. The user-friendly GUI offers an efficient result visualization, highlighting the effectiveness of the proposed methodology. The paper concludes with a discussion of different methods for fruit disease detection, providing significant insights to advance the field of fruit classification.

suggests a method of combining hardware, machine learning, and image processing for fruit quality detection, outperforming current methods. The system, connecting Raspberry Pi with MATLAB, achieves a classification accuracy of 85.64% using a multiclass SVM classifier. The user-friendly GUI offers an efficient result visualization, highlighting the effectiveness of the proposed methodology. The paper concludes with a discussion of different methods for fruit disease detection, providing significant insights to advance the field of fruit classification.

Another study proposes a computer vision-based system for identifying fruit defects caused by nutrient deficiency conducted by Yogesh et al. (2020). By using image analysis techniques system extracts intensity, contour, size, and texture characteristics from fruit images, support vector machine (SVM) technique is employed for accurate categorization and identification of defective fruits. the study further classifies fruit defectives into three distinct defect stages and emphasizes the system's potential in facilitating the early detection of fruit defects and helping in the decision-making of bulk fruit storing. The proposed research offers an efficient strategy for the classification of fruit quality, presenting opportunities for practical use in the fruit industry through further research.

The research by Behera et al. (2020) addresses the potential challenges of fruit grading, identification, and classification. emphasizing the importance of an automated system in the fruit industry, especially in India which is the second-largest fruit producer globally. Approximately 30–35% of harvested fruits are wasted due to a shortage of skilled workers and human personal opinions, the study promotes the integration of machine learning and image processing techniques for intelligent automation. emphasizing the evolution of state-of-the-art methodologies and comparing different methods used for fruit classification based on type, maturity, variety, and intactness, it also explains existing achievements, and constraints suggesting future research in this domain.

Furthermore, Ann and Nosseir (2019) introduced a thorough analysis for fruit classification, particularly targeting four types of fruits (mango, strawberry, apple, and banana) and effectively detecting rotten fruits. the study uses color and texture characters from fruit images. Advanced methods such as RGB value extraction and statistical features from the Gray Level Co-occurrence Matrix (GLCM) are used. different K-Nearest Neighbors algorithms achieve high accuracy in classifying fruit types. linear and quadratic Support Vector Machine (SVM) algorithms classify fresh and rotten fruit effectively based on color and texture features. the two-phase approach results in accuracy ranging from 25% to 98%.the study uses 46 types of seasonal fruits, accurately identifies all the types, and outlines the importance of advanced algorithms for automated classification of fruits.

In the study,Rohit Mamidi et al. (2022) used machine learning and deep learning algorithms to predict fruit freshness from user-provided images. the study evaluates the performance and parameters of the used models through investigation into the applications of machine learning in modern life. significantly, deep learning methods outperformed machine learning models, showing high accuracy and lower false positive rates. the research employs various datasets of three unique fruits, using different vision-based systems and testing multiple layouts, such as binary and multi-class classification tasks. The results emphasize the importance of convolutional neural network elements and deep learning techniques in achieving the highest realization rates. The study provides valuable insights into creating an automated system for classifying fresh and rotten fruit during different consumption and processing stages emphasizing the need for efficient automated technologies to identify fruit degradation.

2.3 Analysis of deep learning Methods:

To address the energy-efficient Convolutional Neural Network (CNN) model for fruit freshness detection, Valentino et al. (2021) introduces an effective solution for identifying fruit freshness while dealing with the issues related to high energy consumption in deep learning, explaining the vital challenges of identifying fruit spoilage in the tropical climate of Indonesia, focusing on sorting and packing processes. The CNN model with six layers achieved an accuracy of 98.64% and reduced electricity consumption in training and testing, emphasizing its practical usefulness for assessing fruit quality.

To detect rotten fruits in the agricultural industry Nerella et al. (2023) proposes an efficient method focusing on the limitations of manual classification in terms of time and effectiveness. the study employs deep learning and computer vision to automate classification to reduce human effort, cost, and time. Various deep-learning models including ResNet50, MobileNetV2, VGG 16, and Inception V3 are trained and tested. Inception V3 proved a robust model achieving 97.1% accuracy in classifying fresh or rotten fruits. This research provides a reliable and automated solution for fruit grading in the agriculture industry.

The research by Jana et al. (2021) focuses on the critical issue of food safety using computer vision and machine learning techniques for the automated detection of rotten fruits and vegetables. the study introduces a specialized convolutional neural network (CNN) with four layers for classifying fresh and rotten. CNN outperforms previous techniques, with classification accuracy ranging from 97.74% to 99.92% and F1 scores from 98.43% to 99.95%. The study is highly effective and provides an advanced solution for automated and accurate detection, addressing global food safety concerns in fruit and vegetable processing.

In a similar study Srinivas and Yadiah (2022) A proposes a deep learning-based approach for fruit quality inspection, targeting the identification of rotten fruits in the agricultural industry. Rottenness not only affects taste and appearance but can also produce harmful mycotoxins. Identifying the limitation of human classification, automated model to detect fruit defects, minimizing production costs, time, and human efforts. The model uses a trained sequential deep learning model, achieving an accuracy of 97% in distinguishing between fresh and rotten fruits (apples, bananas, and oranges) providing valuable solutions for efficient fruit quality assessment and preventing the spread of fruit rottenness in the agricultural.

The research by Roy et al. (2021) introduces automated detection of rotten fruits (fresh apples) using a computer vision framework. Deep learning techniques like semantic segmentation with UNet and Enhanced UNet (En-UNet) are used. the En-UNet performs better compared to UNet with higher training and validation accuracies (97.46% and 97.54% vs. 95.36%) The proposed model is effective for real-time segmentation, identification, and classification of fruit quality, contributing to improved automation in fruit processing industry.

The study by Chakraborty et al. (2021) proposes critical challenges of detecting rotten fruits in the agriculture industry through deep learning methods. Human-based fruit classification is not effective for fruit growers because it is tedious and repetitious. To solve this, the study introduces a model aimed at reducing cost, human efforts, and time by employing automated fruit defect detection in agriculture. A Convolutional Neural Network (CNN) classifies fresh and rotting fruits using methods such as Max pooling, Average pooling, and MobileNetV2. MobileNetV2 performs better and achieves the highest accuracy at 99.46% in training and 99.61% in validation. The proposed CNN model illustrates exceptional capabilities at accurately distinguishing between fresh and rotting fruits, indicating usefulness for the assessment of fruit quality in agriculture.

The study by Kang and Gwak (2022) introduces an ensemble model for fruit freshness classification combining ResNet-50 and ResNet-101 for the classification of fruit freshness and its type. The model has binary and multi-class classifiers to distinguish between fresh or rotten and detect the fruit types. by using transfer learning it achieves high accuracies of 98.50% for fruit freshness and 97.43% for fruit classification. The approach is effective with limited training data and performs better compared to other transfer learning-based ensemble models.

Furthermore, To handle the challenges of precise classification of fresh and rotten fruits in agriculture Sultana et al. (2022) presented a dataset containing sixteen fruit classes() of fresh and rotten fruit.The study serves as a valuable resource for developing efficient algorithms, improving accuracy, and reducing computation time in recognition of fruit freshness. Another study by Mukhiddinov et al. (2022) introduces the crucial need for classifying fruit and vegetable freshness by presenting an improved YOLOv4 model for multiclass categorization, achieving higher average precision than previous versions and providing real-time benefits for the food industry, helps visually impaired individuals in choosing fresh fruit and vegetables.

2.4 Identified Gaps

After critically reviewing the literature, The reviewed research papers significantly contribute to automated fruit quality detection using various methodologies, including KNN, GLCM and SVM, CNN-based frameworks, and ensemble models. However, encounter numerous challenges. These include the limitation of non-diverse and balanced datasets, image variability due to lighting and viewpoints, and complexities in feature extraction like texture and color, affecting model performance. Addressing these challenges involves careful dataset processing, model selection, and better training to improve accuracy in real-world situations. The proposed unique solution, employing transfer learning with VGG16 and edge detection, along with multi-class classification using VGG16 and edge detection on the augmented data, introduces a novel approach to overcome the challenges. The study uses advanced techniques, offers a distinctive solution to bridge these gaps, and contributes to the advancement of fresh and rotten fruit classification methodologies.

3 Methodology

The proposed research work follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology as illustrated in Figure 1. The aim is to systematically perform the development of a fruit freshness detection system for classifying fresh and rotten fruits. The CRISP-DM phases - Project Understanding, Data Acquisition, Data Preparation, Generating Modeling, Model Evaluation, and Deployment - will be adapted and followed to the specific context of fruit classification.

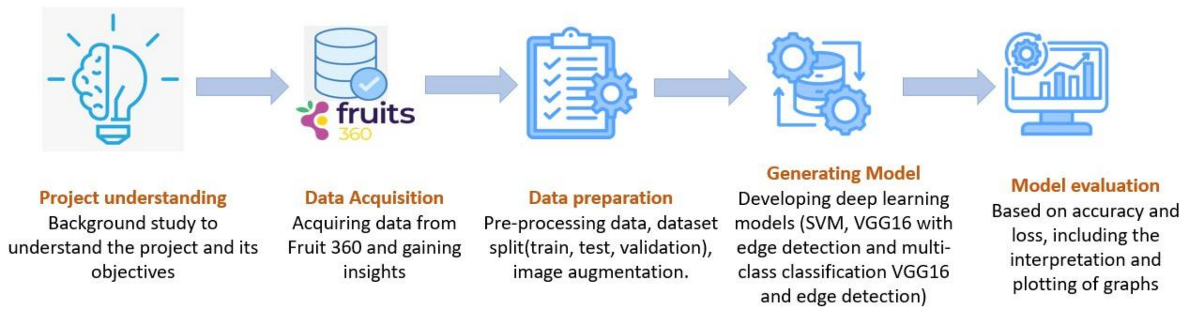


Figure 1: Classification of Fresh and Rotten Fruits Methodology

3.1 Project Understanding:

This phase involves, the Study and analysis of the complexity of the research work. The goal is to extract valuable insights from a business viewpoint, which provides the basis for defining a clear problem statement. To accomplish the project's objectives, an Initial study was done regarding consumer trends and industry dynamics in fruit filtration based on freshness.

3.2 Data Aquisition

This phase involves, obtaining the data from the repository like the Kaggle dataset "Fruit 360" [4]. The acquired data was then loaded for thorough understanding. The focus during this step is to examine the data, understand the distribution, find the patterns, and identify potential quality issues. This step guides subsequent actions for a thorough understanding of the fresh and rotten fruit dataset.

3.3 Data Preprocessing

Following the acquisition of the data, pre-processing was performed based on the knowledge obtained from the previous step the methods used include Data

[4] https://www.kaggle.com/datasets/moltean/fruits/data

Augmentation, Edge Detection. Then the dataset was split into train, test, and validation sets.

3.4 Modelling

This step involves the selection and development of different models for training. The models considered for the study include machine learning models like SVM and Deep Learning models like CNN and VGG16 (for Binary Classification and Multiclass Classification). These models were trained and validated using train and validation sets. After training these models to assess their performance they were employed on test sets.

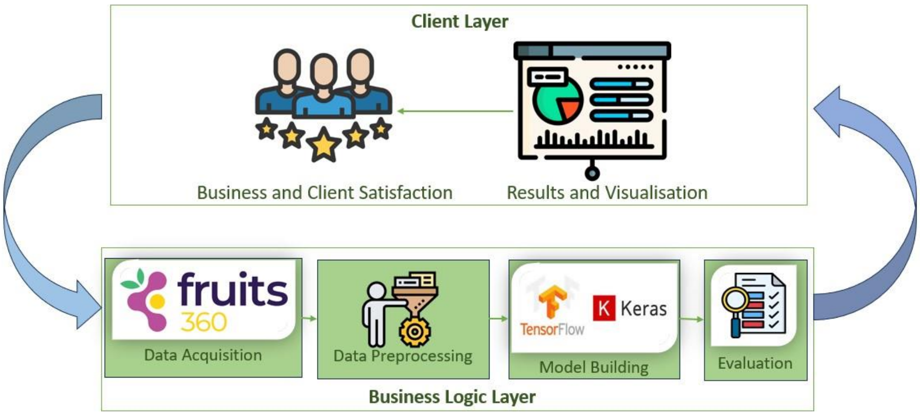


Figure 2: Design Specification

3.5 Evaluation

This step deals with, evaluating the performance of developed models based on parameters such as accuracy and loss. The models were carefully reviewed and evaluated to achieve the desired result of accuracy and make a business impact.

3.6 Results

The result section of the methodology focuses on the interpretation of the gained knowledge and results. Further, the obtained knowledge and results were organized and presented using graphs. Comparison is performed for finding the best-performing model.

4 Design Specification

The design specification includes a visual representation of the client and business layers, illustrating the flow of activities in the fruit classification system as shown in Figure 2. The architecture is divided into two main layers: the Business Layer and the Client Layer. The Business Layer encompasses crucial components of the fruit classification system. It begins with Data Acquisition, where images of fresh and rotten fruits are collected. The acquired data then undergoes the Data Preprocessing stage, including tasks such as resizing, normalization, and augmentation to prepare the dataset for model training. The preprocessed data is used for Model Building, where various classification models, such as SVM, VGG16, and VGG16 with edge detection (for binary and multiclass classification tasks), are trained. These models are designed to classify fruits into distinct categories based on freshness. The Evaluation stage is where the performance of the models is assessed using testing datasets. Evaluation metrics and results are important for assessing models' performance in classifying fresh and rotten fruits. The Business Layer forms a cyclical process, indicating that the system can iterate through data acquisition, preprocessing, model building, and evaluation to continuously improve its performance.

The Client Layer is designed to provide a user-friendly interface and meaningful insights. It consists of components such as Results and Visuals, where the outcomes are presented in a comprehensible format. The results may include accuracy metrics, confusion matrices, and other relevant information. Additionally, the Client Layer includes components related to Client and Business Satisfaction. This involves gathering feedback from both clients and business stakeholders, ensuring that the fruit classification system aligns with their expectations and requirements.

The connection between the Business and Client Layers completes a feedback loop. The evaluation results influence the visuals and results presented to the clients, and client feedback contributes to potential adjustments in the business layer. This visualization underscores the interconnected nature of data processing, model building, and client interaction in the fruit classification system.

5 Implementation

This research work aims to classify fresh and rotten fruit with the help of advanced deep-learning algorithms and real data. The goal is to establish a strong foundation for automated fruit quality assessment, offering practical solutions for agricultural applications.

5.1 Dataset Collection:

The dataset for this implementation was gathered from the "Fruit 360" repository and is publicly available on Kaggle consisting of 18,492 fruit images. It contains a broad collection of images featuring three distinct fruits: Apple, Banana, and Orange based on fresh and rotten classes. as shown in Figure 3

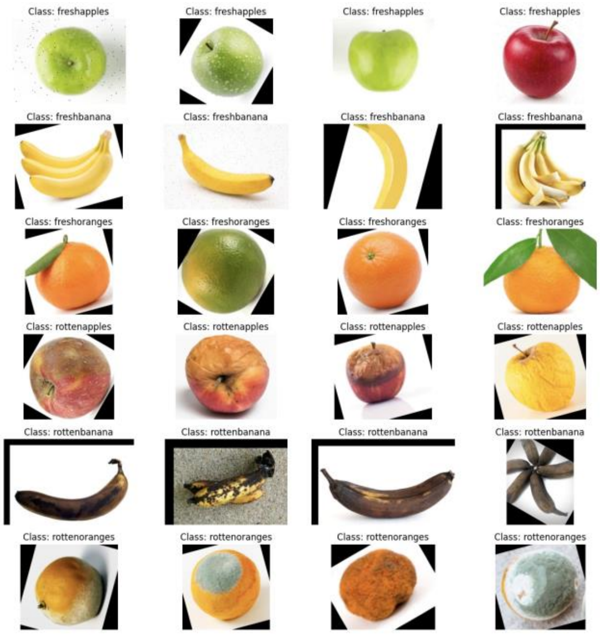
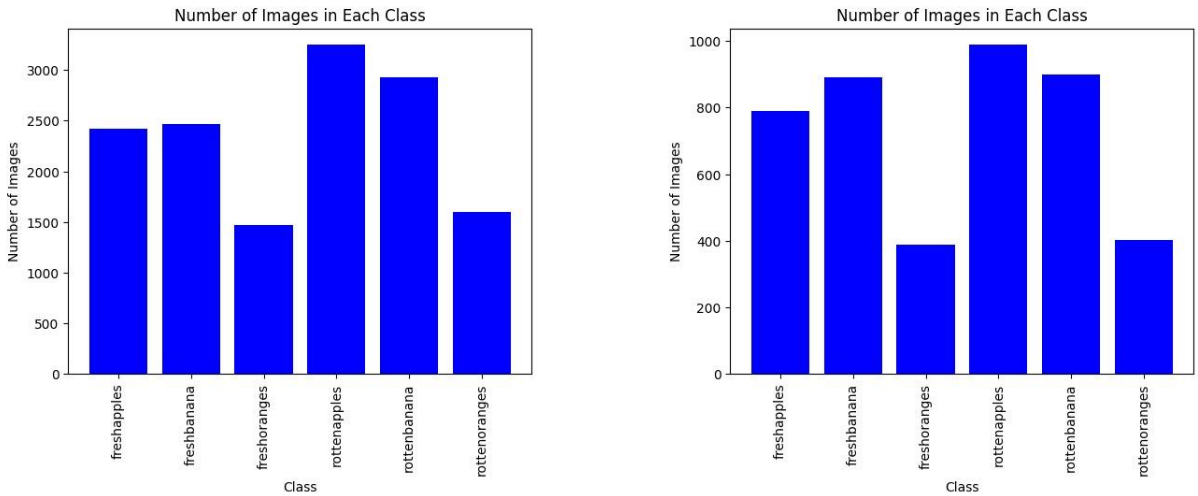


Figure 3: Random Sample Display

The training dataset contains 2424 images of fresh apples, 2468 images of fresh bananas, 1466 images of fresh orange, 3248 images of rotten apples, 2932 images of rotten banana, and 1595 images of rotten orange images. Whereas, the testing dataset consists of 791 fresh apple images, 892 fresh banana images, 388 fresh orange images, 988 rotten apple images, 900 rotten banana images, and 403 rotten orange images as shown in Figure 4. This labelled dataset and segregated dataset help in better training and evaluation processes in the subsequent stages of the implementation.



(a) Distribution of Classes in Train Set.

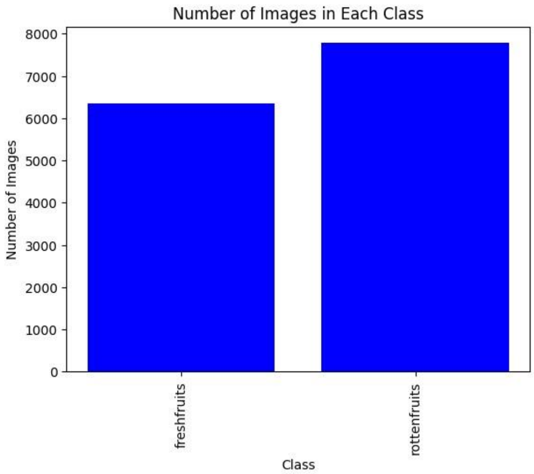
(b) Distribution of Classes in Test Set.

Figure 4: Distribution of Classes in Train and Test Set

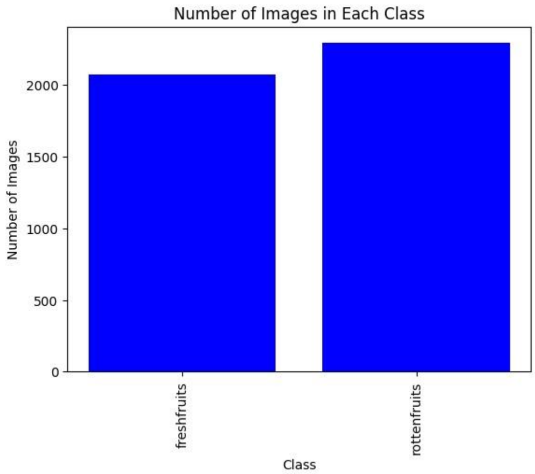
5.2 Data Pre-processing and Data Augmentation:

5.2.1 Data Preparation for Binary and Multi-class classification

For the binary classification model, the dataset was pre-processed by merging images of fresh oranges, fresh bananas, and fresh apples into a unified ”Fresh” category. Similarly, images of rotten oranges, rotten bananas, and rotten apples were merged into the ”Rotten” category as illustrated in Figure 5. This preprocessing step ensures a binary classification setup, where the model distinguishes between fresh and rotten fruits. On the other hand, the multi-class classification model needs a more detailed approach. The dataset was structured into six distinct classes, representing fresh and rotten categories for each type of fruit (apple, banana, and orange).



(a) Distribution of Classes in Train Set



(b) Distribution of Classes in Test Set

Figure 5: Distribution of Classes in Train and Test Set for Binary Classification

5.2.2 Data Augmentation:

To enhance the variability of the training dataset and improve the reliability of the model, data augmentation was employed using the ImageDataGenerator from the Keras library. The augmentation process was configured with different transformations, contributing to a more diverse and extensive dataset for training. The different stages of augmentation are shown in Figure 6 along with the results obtained on each stage.

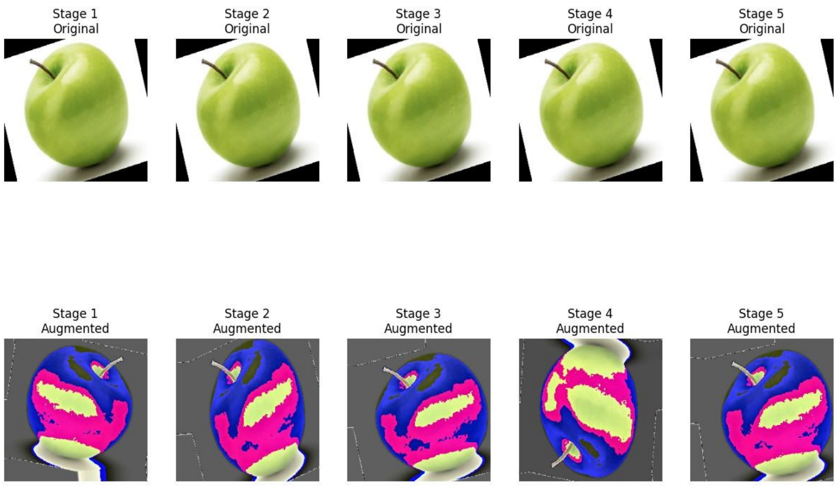


Figure 6: Data Augmentation Stages

The following augmentation techniques were applied:

1. Samplewise Centering: Each sample's mean was set to zero, which helps in standardizing the dataset and centring the distribution of pixel values.
2. Rotation: Images were randomly rotated in the range of 0 to 10 degrees angle, introducing angular diversity and ensuring the model's resilience to variations in fruit orientations.
3. Zooming: A random zoom range of 0.1 was applied, ensuring the model exposure to variations in image scales, contributing to its ability to recognize fruits at different distances.
4. Horizontal and Vertical Shifting: Random horizontal and vertical shifts (up to 10% of the total width and height, respectively) were introduced. This improves the model's adaptability to fruits appearing at different positions within the images.
5. Horizontal and Vertical Flipping: Random horizontal and vertical flips are applied to fruit images, diversifying the dataset by presenting mirror images to the model during training.

These augmentation approaches diversify the training dataset, reducing overfitting and improving the model's ability to generalize to new images of fruit. The configurations align with best practices for training model in image classification tasks.

5.3 Model Building and Selection:

To create an effective fruit classification system, the study developed four different models. These models aim to distinguish between fresh and rotten fruits, with the multi-class deep learning model providing additional granularity to identify specific types within each category.

5.3.1 Support Vector Machine

The initial model employed for the binary classification task, distinguishing between fresh and rotten fruits, is based on the Support Vector Machine (SVM) algorithm. The SVM model aims to find an optimal hyperplane that effectively separates the feature space into two classes—fresh and rotten for binary classification.

The SVM model is initialized using a linear kernel, denoted as $K(x,y) = x^T y$, which indicates that the decision boundary is a linear plane in the input space. The decision boundary is established by maximizing the margin between the two classes, creating a clear separation. The training data, X_{train} and y_{train}, are used to find the optimal hyperplane. The SVM model is trained using the fit method, where the training data is used to determine the coefficients of the hyperplane. Once trained, the model predicts the class labels for the test set (X_{test}) using the predict method, generating the predictions y_{pred}.

5.4 Transfer Learning with VGG16

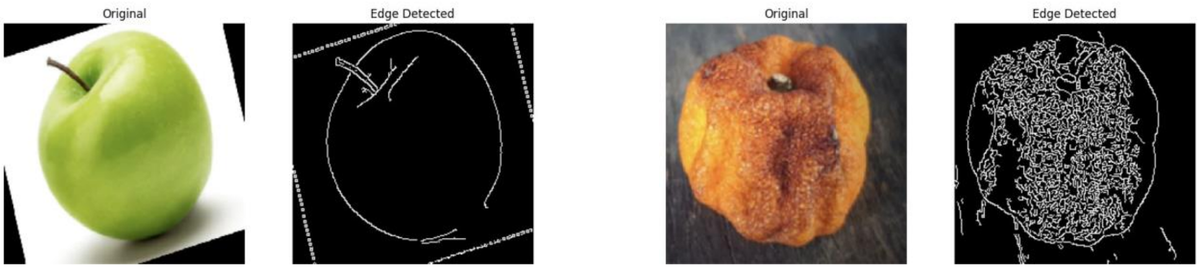
The second model employed for the binary classification task is based on Transfer Learning. VGG16 is a deep convolutional neural network (CNN) which is effective in image classification tasks.

The VGG16 model is initialized with pre-trained weights on the ImageNet dataset, providing a strong foundation for the feature extraction process. The model is configured to accept input images of size 224×224 pixels with three colour channels(RGB). The base VGG16 model is frozen, and its weights are set to non-trainable. This makes sure that the pre-trained weights are maintained during the training of the binary classification head. The architecture of the model is extended by including a Global Average Pooling 2D layer, which reduces the spatial dimensions of the feature map to a single value per feature. The final layer is a Dense layer with a sigmoid activation function, generating a probability score for the binary classification. The model is compiled using the binary cross-entropy loss function, and the binary accuracy metric is used to evaluate the model's performance.

5.5 VGG16 Model with Edge Detection for Fresh and Rotten Fruit Classification

The third model developed for the classification task incorporates edge detection as an additional preprocessing step. This model is based on the VGG16 architecture, similar to the second model (5.4). However, an edge detection function is applied to the input images before providing them to the VGG16 base model.

The VGG16 base model is initialized with pre-trained weights on the ImageNet dataset and configured to accept input images with size 224×224 pixels with three colour channels. As with the previous model, the base VGG16 model's weights are set to non-trainable to retain the pre-trained features. The binary classification head is added on top of the base model, consisting of a Global Average Pooling 2D layer followed by a Dense layer with a sigmoid activation function, which generates a probability score for binary classification. Additionally, an 'apply edge detection' method is defined to perform edge detection on



(a) Edge Detection on Fresh Fruit (b) Edge Detection on Rotten Fruit

Figure 7: Canny Edge Detection

the input images. This function converts the RGB images to grayscale, applies the Canny edge detection algorithm, and then converts the resulting edges back to an RGB format.

5.6 VGG16 Multi-class Classification Model with Edge Detection

The fourth model is designed for multi-class classification, specifically classifying fruits into six different categories: fresh apple, fresh banana, fresh orange, rotten apple, rotten banana, and rotten orange. This model integrates edge detection as a step in preprocessing, similar to the previous binary classification model (5.3.1).

The VGG16 architecture is used as the base model, initialized with pre-trained weights from the ImageNet dataset. The model is configured to accept input images of size 224 × 224 pixels with three colour channels. The weights of the VGG16 base model are set to non-trainable to retain the pre-trained features. The input images were improvised with an additional preprocessing step of edge detection before being fed into the VGG16 base model. The 'apply edge detection' function is defined to convert the RGB images to grayscale, apply the Canny edge detection algorithm from OpenCV, and then convert the resulting edges back to an RGB format. The multiclass classification head is added on top of the base model, consisting of a Global Average Pooling 2D layer followed by a Dense layer with a softmax activation function. The output layer has six nodes corresponding to the different classes, producing probabilities for each class.

6 Evaluation

The evaluation section focuses on the performance assessment of four distinct models developed in the classification of fresh and rotten fruits projects. The models under consideration include Support Vector Machine (SVM), VGG16 Transfer Learning, VGG16 with Edge Detection, and VGG16 Transfer Learning for Multiclass classification with edge detection. Each model is assessed to analyze its effectiveness in classification tasks. The aim of this step is to find the strengths and weaknesses of each model.

6.1 Support Vector Machine (SVM) for Fresh and Rotten Fruit Classification

The evaluation of the first model (SVM), involved a structured training and testing process. The model is trained on a subset of images from the training set, a linear kernel is used for classification.

The accuracy was assessed on a separate set of images and showed an accuracy of 84.17% on the test set. To further validate the model's reliability, an independent validation set was used, resulting in an accuracy of 89.50%. This showcases the model is capable of generalization. The evaluation result shows the good performance of the SVM model, setting a base for comparative assessments with other models in the subsequent sections.

6.2 Transfer Learning with VGG16 for Fresh and Rotten Fruit Classification

The evaluation of the second model, based on Transfer Learning using VGG16 for the binary classification. involved leveraging pre-trained weights from the ImageNet. The model was configured with a binary classification layer for detecting fresh and rotten fruits. To ensure the model's generalization, data augmentation techniques were applied during training, including rotation, zoom, shifts (horizontal and vertical), and horizontal and vertical flips.

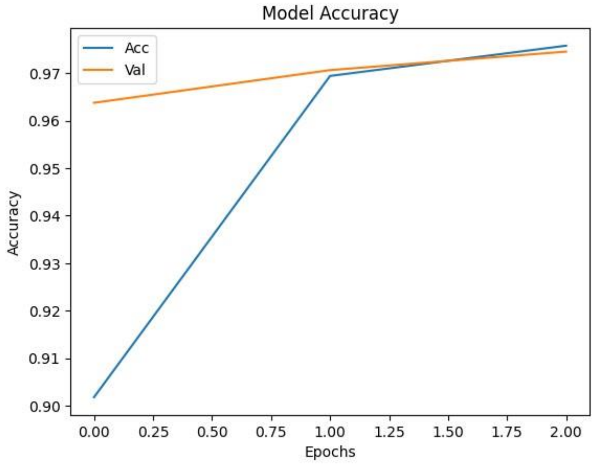


Figure 8: Evaluation of Model2: VGG16

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	0.2767	90.18%	0.1080	96.38%
2	0.0832	96.94%	0.0922	97.07%
3	0.0699	97.58%	0.0856	97.46%

Table 1: Training and Validation Metrics for VGG16 Transfer Learning Model

The model underwent three training epochs, resulting in a gradual reduction in both training and validation losses, reaching 0.0699 on the training set and 0.0856 on the validation set as shown in Figure 8. The accuracy on the validation set was notably high at 97.46%, demonstrating the model's effectiveness in correctly classifying fresh and rotten fruits. The evaluation results illustrated the effectiveness of transfer learning with VGG16 architecture for accurate fruit classification.

6.3 Transfer Learning with VGG16 and Edge Detection for Fresh and Rotten Fruit Classification

The VGG16 Transfer Learning Model with Edge Detection is thoroughly evaluated, and trained for three epochs using binary cross-entropy loss and binary accuracy as metrics. The model achieved a validation accuracy of 97.96%. Performing slightly better than the previous model.

Epoch	**Training Loss**	**Training Accuracy**	**Validation Accuracy**
1	0.2525	90.36%	97.18%
2	0.0884	96.88%	97.80%
3	0.0676	97.71%	97.96%

Table 2: Training and Validation Metrics for VGG16 Transfer Learning Model with Edge Detection

The validation set's loss is 0.0670 which indicates strong performance. The accuracy graph shows steady improvement in both training and validation accuracies (as shown in Table 2). The model achieves a promising accuracy of 97.96% on the validation dataset, which indicates model is effective in distinguishing fresh and rotten fruits for automated classification.

6.4 Multi-class Classification with VGG16 and Edge Detection

The 4^{th} model uses a VGG16-based architecture with edge detection for multi-class fruit classification across six categories. The model is trained on a dataset featuring fresh and rotten apples, bananas, and oranges. Custom generators with edge detection are employed during preprocessing, and the model is compiled using categorical cross-entropy loss and categorical accuracy metrics.

After three training epochs, the model exhibits good improvement, with a loss reduced to 0.0966 and a categorical accuracy of 96.66% (Figure 9). Validation results achieved a categorical accuracy of 97.33%. The model accurately distinguishes between fruit classes, showcasing the efficacy of the VGG16 transfer learning with an edge detection approach for multi-class fruit classification.

6.5 Comparison and Discussion of Developed Models

The results showed consistent improvement in accuracy across models. The SVM achieves 84.17%, while VGG16 and VGG16 with edge detection show substantial gains at 97.43% and 97.96%, respectively as compared in Figure 10.

The multi-class classification model achieves 97.33%, demonstrating a very good ability to identify the fruit category. Integrating edge detection proves effective in refining the models' accuracy. These outcomes highlights the efficacy of deep learning in fruit quality detection, offering diverse options for different requirements.

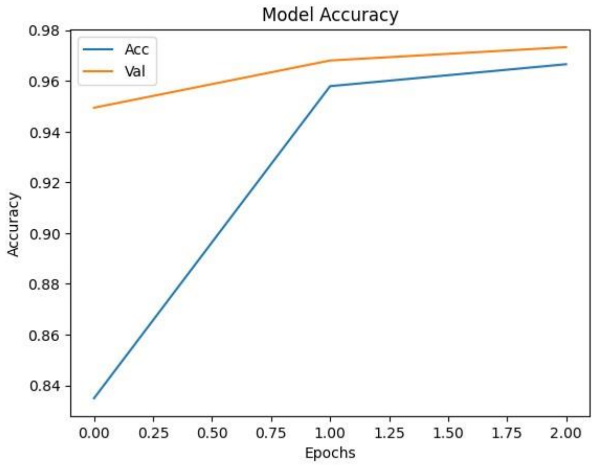
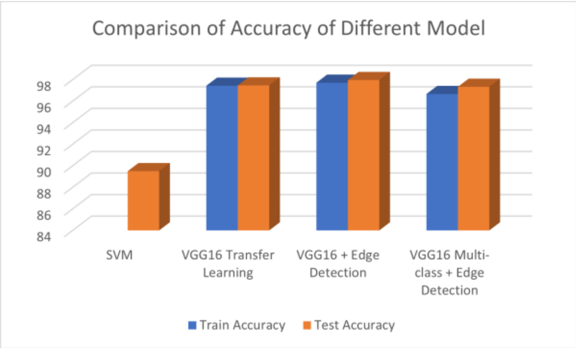
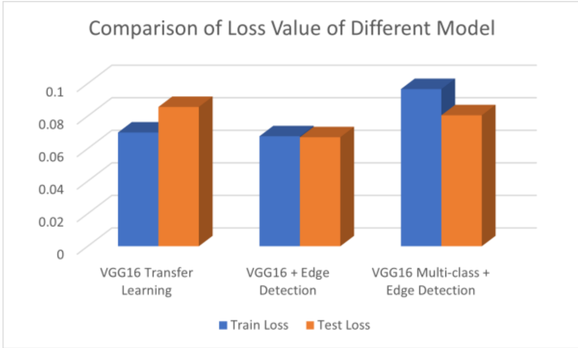


Figure 9: Evaluation of Model4: Transfer Learning with VGG16 and Edge Detection for Multiclass Classification



(a) Accuracy Comparison of Models



(b) Loss Comparison of Models

Figure 10: Accuracy and Loss Comparison of Different Models

Model	**Training Accuracy**	**Validation Accuracy**
SVM	-	89.50
VGG16 Transfer Learning	97.43	97.46
VGG16 + Edge Detection	97.71	97.96
Multi-class VGG16 + Edge Detection	96.66	97.33

Table 3: Comparison of Model Training and Validation Accuracies

7 Conclusion and Future Work

As the project achieved significantly high accuracy, it holds potential for real-world use in the classification of fresh and rotten fruits. this will ensure that only high-quality fruits reach the market, resulting in enhanced financial performance and improving consumer health satisfaction. There are different methods and ideas have been proposed in the field but the focus was only on specific metrics and there are gaps in terms of preprocessing. This project contributes to automated fresh and rotten fruit classification by employing various machine learning techniques such as SVM, VGG16 Transfer learning, VGG16 with Edge Detection, and Multi-class classification using VGG16 with Edge Detection. Leveraging Transfer Learning and introducing two models: a binary classification and multi-class classification for detecting fresh and rotten fruits of six classes. Augmentation techniques improve model generalization. Edge detection before VGG16 application, resulting in better binary accuracy of 97.96% and loss of 0.0670. Whereas, Multi-class classification achieved an accuracy of 97.33% and a loss of 0.0805. This integrated approach improved reliability and performance. In alignment with prior studies, this research employs a unique solution by introducing fruit classification through VGG16 transfer learning with edge detection method improvements to achieve more accurate and reliable fresh and rotten fruit classification.

Future work should focus on the collection of diverse datasets, encompassing a wider array of vegetables alongside fruits. This approach will provide a more comprehensive market perspective, potentially resulting in improved health and financial benefits. Explore advanced deep learning architectures beyond VGG16, and optimize hyperparameters for good accuracy. Employ texture recognition techniques for improving defect identification in fruits. Advancing the field requires a focus on deploying models in edge computing and addressing concerns, including dataset biases. These future directions aim to improve the efficiency and accuracy of advancing fruit quality detection for broader impact.

Acknowledgment

I extend my heartfelt gratitude to Professor Shubham Subhnil for his invaluable guidance during my research. I'm also grateful to the National College of Ireland for providing this research opportunity. Special thanks to my friend Amrit, as well as my family and friends, for their immense support and encouragement throughout my master's degree pursuit.

References

An, X., Li, Z., Zude-Sasse, M., Tchuenbou-Magaia, F. and Yang, Y. (2020). Characterization of textural failure mechanics of strawberry fruit, *Journal of Food Engineering* **282**: 110016.

Ann, A. and Nosseir, S. (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**: 47.

Bargoti, S. and Underwood, J. (2017). Deep fruit detection in orchards, *2017 IEEE international conference on robotics and automation (ICRA)*, IEEE, pp. 3626–3633.

Behera, S., Rath, A., Mahapatra, A. and Sethy, P. (2020). Identification, classification grading of fruits using machine learning computer intelligence: a review, *Journal of Ambient Intelligence and Humanized Computing* .

Bhargava, A. and Bansal, A. (2020). Automatic detection and grading of multiple fruits by machine learning, *Food Analytical Methods* .

Cao. (2021). Detection of fruit surface defects based on machine vision, *Journal of Physics: Conference Series*, IOP Publishing, p. 022048.

Cao, J., Wang, C., Xu, S., Chen, Y., Wang, Y., Li, X. and Sun, C. (2019). The effects of transportation temperature on the decay rate and quality of postharvest ponkan (citrus reticulata blanco) fruit in different storage periods, *Scientia Horticulturae* **247**: 42–48.

Chakraborty, S., Shamrat, F. J. M., Billah, M. M., Al Jubair, M., Alauddin, M. and Ranjan, R. (2021). Implementation of deep learning methods to identify rotten fruits, *2021 5th international conference on trends in electronics and informatics (ICOEI)*, IEEE, pp. 1207–1212.

Chandini, A. A. et al. (2018). Improved quality detection technique for fruits using glcm and multiclass svm, *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, IEEE, pp. 150–155.

Jana, S., Parekh, R. and Sarkar, B. (2021). Detection of rotten fruits and vegetables using deep learning, *Computer Vision and Machine Learning in Agriculture* pp. 31–49.

Kang, J. and Gwak, J. (2022). Ensemble of multi-task deep convolutional neural networks using transfer learning for fruit freshness classification, *Multimedia Tools and Applications* **81**(16): 22355–22377.

Lu Fae, e. a. (2019). Effect of vibration on storage quality and ethylene biosynthesisrelated enzyme genes expression in harvested apple fruit, *Scientia Horticulturae* **249**: 1–6.

Mukhiddinov, M., Muminov, A. and Cho, J. (2022). Improved classification approach for fruits and vegetables freshness based on deep learning, *Sensors* **22**(21): 8192.

Nerella, J. T., Nippulapalli, V. K., Nancharla, S., Vellanki, L. P. and Suhasini, P. S. (2023). Performance comparison of deep learning techniques for classification of fruits as fresh and rotten, *2023 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI)*, IEEE, pp. 1–6.

Nosseir, A. and Ahmed, S. E. A. (2019). Automatic classification for fruits' types and identification of rotten ones using k-nn and svm, **15**(03): 47–61.

Pan, L., Zhang, W., Zhu, N., Mao, S. and Tu, K. (2014). Early detection and classification of pathogenic fungal disease in post-harvest strawberry fruit by electronic nose and gas chromatography–mass spectrometry, *Food Research International* **62**: 162–168.

Rohit Mamidi, S. S., Akhil Munaganuri, C., Gollapalli, T., Aditya, A. T. V. S. and B, R. C. (2022). Implementation of machine learning algorithms to identify freshness of fruits, *2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)*, pp. 1395–1399.

Roy, K., Chaudhuri, S. S. and Pramanik, S. (2021). Deep learning based real-time industrial framework for rotten and fresh fruit detection using semantic segmentation, *Microsystem Technologies* **27**: 3365–3375.

Singh, D. and Sharma, R. (2018). Chapter 1 - postharvest diseases of fruits and vegetables and their management, *in* M. W. Siddiqui (ed.), *Postharvest Disinfection of Fruits and Vegetables*, Academic Press, pp. 1–52.

Srinivas, R. and Yadiah, B. (2022). Deep learning based fruit quality inspection, *Int. J. Res. Appl. Sci. Eng. Technol* **10**: 4535–4539.

Sultana, N., Jahan, M. and Uddin, M. S. (2022). An extensive dataset for successful recognition of fresh and rotten fruits, *Data in Brief* **44**: 108552.

Valentino, F., Cenggoro, T. W. and Pardamean, B. (2021). A design of deep learning experimentation for fruit freshness detection, *IOP Conference Series: Earth and Environmental Science*, Vol. 794, IOP Publishing, p. 012110.

Yogesh, D., Dubey, A., Ratan, R. and Rocha, A. (2020). Computer vision based analysis and detection of defects in fruits causes due to nutrients deficiency, *Cluster Computing* **23**.

Zhang, B., Fan, S., Li, J., Huang, W., Zhao, C., Qian, M. and Zheng, L. (2015). Detection of early rottenness on apples by using hyperspectral imaging combined with spectral analysis and image processing, *Food Analytical Methods* **8**: 2075–2086.