International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 4, Jul-Aug-2026, ISSN (Online): 2395-566X

# Identifying Decayed Fruits and Vegetables from Large Food Retailers Utilizing Machine Learning and Deep Learning, and Transforming Them into Biogas

**Dr.Sumbul Alam[1]**
[1](Assistant Professor
Department of CS and AI
University of Bisha ,Al nakheel , Bisha -67714, Asir Province, Kingdom of Saudi Arabia
syedsumbul@gmail.com
ORCID: 0000-0003-1883-5156)

**Abstract- Food waste management is a contemporary global issue. According to UNDP, food waste constitutes 40-51% of the waste in the Kingdom, followed by paper, cardboard, plastics, and other materials. Consequently, an effective waste segregation system is essential to address this problem. This research proposal aims to develop a mechanism for the segregation of decayed fruits and vegetables from the shelves, storage, and inventories of retailers such as Hyper Panda, Carrefour , and Lulu hypermarkets, utilising a hybrid machine learning algorithm comprising Linear SVM and PCA, alongside YOLO for real-time detection. Following segregation, this organic waste may be transported to a biorefinery where, using anaerobic digestion technology, it can be turned into methane gas. Currently, the door-to-door accessibility of processed foods, consumer products, and groceries is in great demand. Consequently, this methane gas may be utilised by their own delivery trucks and vehicles, so minimising food waste reduction in retail, promoting energy sustainability, and decreasing pollution.**

**Keywords- Food Waste Management, Machine Learning, YOLO, SVM, PCA, Anaerobic Digestion, Biogas, Retail Sustainability, Computer Vision, Waste-to-Energy.**

## I. INTRODUCTION

Food wastage is considered one of the most critical issues facing the world today in the twenty-first century [1][3]. UNDP estimates that food wastage makes up about 40-51% of the total waste in the kingdom, more than paper, cardboard, and plastic waste [1][3]. It becomes obvious how important it is to organize waste sorting and treatment [1][3]. Big food retailers such as Hyper Panda, Carrefour, and Lulu hypermarkets experience some difficulties with fresh produce inventory management because spoiling may lead to significant losses of money and the environment [2][10].

The issue of the global food waste crisis has consequences much broader than mere losses of resources [3][5]. Firstly, it leads to emission of methane, which is 25 times more harmful to the atmosphere than CO2 in a hundred years period as it degrades in landfills [1][3]. Secondly, it is a missed chance to get back the energy stored in organic matter [3][5]. Therefore, it is necessary to solve both these problems at once [3][5].

The recent developments in artificial intelligence and computer vision have provided opportunities for automating the process of quality evaluation of the perishables [7][8][9]. It has been found that machine learning algorithms, especially deep learning models such as Convolutional Neural Network (CNN), have proven their ability to classify images very accurately and detect fruit and vegetables' freshness levels [7][8][9]. There is evidence that deep learning models have proven their ability to recognize fresh and decayed products based on their images and even exceed human evaluators in the speed and consistency of decision making [7][8][9].

Meanwhile, anaerobic digestion technique has been found to be an efficient and well-established system of energy production from organic waste material [1][3][4]. It is a process where micro-organisms break down organic material in the absence of oxygen resulting in the production of biogas made up mainly of methane and carbon dioxide [4]. Biogas thus produced can be used for power generation, heating purpose, or even as a fuel for vehicles [1][3][4]. Research has shown that there is considerable biogas potential in food waste such as fruit and vegetable waste material, which ranges between 0.27 and 0.642 $m^3$ $CH_4$/kg of volatile solids for mono-digestion [1][3].

The incorporation of AI technology for detecting waste combined with biogas generation is a ground-breaking method of tackling the issue of food waste within the retail industry [2][7][10]. High resolution cameras mounted on shelves of fresh food would ensure monitoring of product condition and identification of decayed goods [2][7][10]. Such waste could then be collected and delivered to biorefinery facilities for anaerobic digestion and conversion to methane gas [1][3][4]. The produced methane could then be used as fuel for delivery trucks and vehicles [1][3][4].

In this paper, we have provided a framework for the detection of rotten fruits and vegetables by implementing the combination of PCA and Linear SVM, along with the YOLO model that is capable of providing us real-time results [6][7]. Our proposed framework is going to help retailers in detecting decay in fruits and vegetables automatically and efficiently [2][7]. Some previous researchers have shown that the combination of both the feature extraction method and classifier method proves to be effective in determining the freshness of fruits [6][7]. PCA, in particular, proved to be very useful in reducing the dimensionality of data while retaining its discriminating power [7].

Apart from that, the importance of the study transcends mere waste management since through changing the concept of waste into energy, the framework proposed becomes more relevant to the concept of circular economy and sustainable development [2][5]. The retailers get the chance not only to decrease their carbon footprint but to add value to waste materials at the same time [2][5]. In addition, the use of biogas in transportation helps to increase energy independence of the vehicles and decrease greenhouse gases emissions from transportations [5].

## II. LITERATURE REVIEW

The area of computer vision in waste-to-energy processes has been gaining popularity in recent times because of the growing recognition of food waste as a sustainability issue [10]. This area is quite interdisciplinary since it includes various disciplines such as artificial intelligence, environmental engineering, and logistics [10].

Machine learning techniques for food quality analysis have made significant progress over the last decade [6]. Earlier algorithms used hand-engineered features from image processing, which included color histograms, texture, and shape features [6][7]. Features like statistical properties (mean and standard deviation), geometric properties (size and shape), and texture properties (Local Binary Patterns, LBP and Gray-Level Co-occurrence Matrix, GLCM) were frequently used [6][7]. Support Vector Machine (SVM) and Random Forest were among those classifiers that became prevalent due to the high dimensionality of the feature space and resistance to overfitting [6][7]. Several studies by Nturambirwe and Opara showed that SVM was able to obtain classification accuracy above 90% for detecting defects in fruits using correct feature extraction algorithms [6].

Introduction of deep learning has led to revolution in the domain of computer vision, with Convolutional Neural Network (CNN) taking the place of the old machine learning techniques [7][8][9]. Unlike the existing techniques which rely on manual feature extraction, CNN enables automatic learning of features in an image ranging from low-level features such as edges and texture to high-level features representing objects [7][8][9]. This technique has been found to be useful in fruit quality inspection where there are variations in defects across varieties and growing conditions [7][8][9]. Theoretical basis of deep learning models has been developed by Lecun et al., whereas Howard et al. and Wang et al. have provided practical application of these techniques in agriculture [9].

The object detection systems such as YOLO (You Only Look Once) have proved to be effective methods of assessing quality in real-time [7][8]. While classification models offer a generalized label for the whole image, YOLO has the ability to detect several objects along with their location within the image [7][8]. Such an ability is beneficial in cases of retail stores where the shelf consists of various items which need separate assessment [7][8]. As per the studies conducted by Jocher et al. and Wang et al., the variants of YOLO such as version 8 offer the best combination of speed and performance accuracy, thus becoming applicable in time-critical operations [7][8]. Bonora et al. performed the detection of scald in pears using YOLO with moderate performance levels that can be enhanced by better training and more modern neural networks [7][8].

Research regarding the generation of biogas from food waste has also evolved [5]. Anaerobic digestion has come to be

International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 4, Jul-Aug-2026, ISSN (Online): 2395-566X

recognized as the most effective technique in this context [1][3][4]. The work by Bong et al., which offers a detailed account of the characteristics of food waste and their relevance to biogas potential, highlights those carbohydrates, proteins, and lipids found in food waste differ widely on a seasonal and geographical basis [1]. Research findings reveal that co-digestion of food waste with other materials can increase biogas generation, with quantities of biogas generated being between 0.272 and 0.859 $m^3$ $CH_4$/kg VS [1][3].

Singh et al. conducted an investigation into the sustainable bioconversion of food waste using anaerobic digestion, where they identified parameters such as pH and temperature as key factors influencing the process performance [3]. In their review, they stressed the significance of employing efficient pretreatment techniques to increase biogas yield, where thermal and mechanical pre-treatments were found to be promising [3]. They highlighted that although laboratory investigations have proved the possibility of biogas generation from food waste, there exist hurdles in its commercial application [3][4].

Integrating AI-based waste tracking with circular economy has been done by Sigala et al., who tested the efficiency of automated waste tracking technology in HORECA industry [2]. This research revealed that computer vision-based waste tracking could result in 23-51% of food waste reduction owing to staff education and informed decision making [2]. It is also mentioned that technology was transferable, which implies its applicability in the retail environment [2]. Another research carried out by Uchil and B.G. involves development of AI-based solution to minimize food waste in the entire supply chain process [10]. It combines IoT sensor networks along with machine learning models for crop yield, demand and shelf-life prediction [10]. Shelf-life prediction using SqueezeViT model resulted in 78.3% accuracy [10].

Circular economy strategies to manage food waste have shown the significance of coordination between large firms and community sectors [5]. In studies done in Thailand, it was seen that the processes of thermal dehydration and anaerobic digestion could be effectively merged using the thermally processed food waste as a co-substrate for the community anaerobic digestion system [5]. Biogas formed, consisting of 41.1% of methane and 32.7% of carbon dioxide, was found to be feasible as cooking gas [5]. Economic evaluation revealed that the process of thermal dehydration had high economic viability in shopping malls, whereas anaerobic digestion had moderate social benefits [5].

## III. PROPOSED METHODOLOGY

**System Architecture Overview**

These modules include: data acquisition, data processing and feature extraction, machine learning detection and classification, and biogas production. The high-definition camera fixed at the top of the shelves in retail stores will take live photos of the fresh produce. These photos along with existing training datasets will be input into the hybrid detection algorithm to detect the decayed fruits and vegetables for segregation.

**Preprocessing**

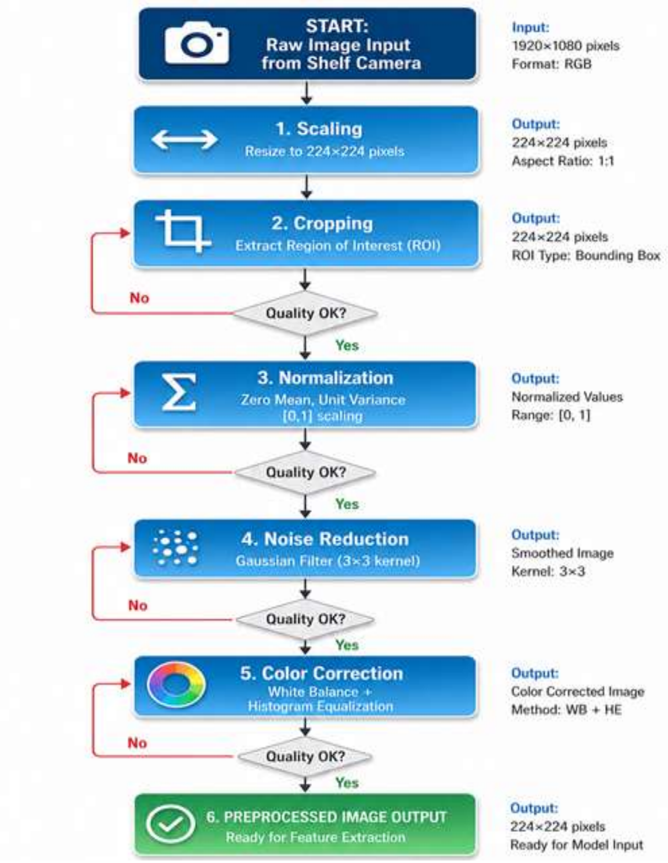


Figure 1: Preprocessing Procedures for Captured Images

The captured images go through some preprocessing procedures to ensure that the images are of good quality and consistency. Below are the steps followed:

International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 4, Jul-Aug-2026, ISSN (Online): 2395-566X

**Data Acquisition and Dataset**

The training dataset is obtained from Kaggle.com in 2018 by Sriram Reddy Kalluri and Oltean. It contains a total of 12,132 images of four different fruit types: banana, orange, apple, and avocado – both fresh and decayed. In the case of real-time implementation, the cameras installed on top of produce shelves will take continuous images.

The above flowchart shows the preprocessing procedures for the captured images, which start with raw input from the camera, followed by resizing, cropping, normalizing, filtering to reduce noise, and finally color correction.

- **Rescaling:** Image scaling is done to achieve uniform dimensions (such as 224 x 224).
- **Region of Interest (ROI) Cropping:** ROI cropping is performed to minimize background information.
- **Normalization:** Normalization of pixel values is done either to the range [0,1] or to zero mean and unit variance to ease training.
- **Noise Reduction:** Application of filter like Gaussian blur is done for noise reduction.
- **Color correction:** Color correction is done to overcome the effects of varied lighting conditions.

**Feature Extraction**

Feature extraction involves the conversion of preprocessed images into numeric features which contain useful information for classification. Below are the features extracted using traditional machine learning methods:

- **Statistical Features:** Mean, standard deviation, and moment statistics of the color features.
- **Geometrical Features:** Size, shape, and aspect ratio of the produce.
- **Texture Features:** Local Binary Pattern (LBP) and Grey Level Co-occurrence Matrix (GLCM) provide spatial features.
- **Shape Features:** Histogram of Oriented Gradients (HOG) provides edge orientations.

**Detection Model**

The detection model uses a combination of both machine learning based and deep learning based methods. The machine learning based approach uses PCA and linear SVM while the YOLO network is used for real-time detection and localization.


Figure 2: Architecture of Hybrid Detection Model Combining PCA-SVM and YOLO

Architecture diagram depicting the dual-path detection approach where image inputs are processed by PCA-SVM (left path) and YOLO (right path). The PCA-SVM path is depicted by feature extraction, PCA dimensionality reduction, and SVM classification modules. The YOLO path is described by the CNN backbone, feature extraction, and bounding box regression outputs. Both paths are connected to a decision fusion module.

**Principal Component Analysis (PCA):** In PCA, the dimensionality of the feature space is decreased to retain maximal variance. In a data set X containing n samples and p features, the covariance matrix and eigenvectors are calculated. The principal components are sorted by explained variance, and the first k components are selected.

Linear Support Vector Machine (SVM): Linear SVM finds the best hyperplane to separate fresh from decayed produce in the feature space. If the classes are linearly separable, the decision function is:

$$\mathrm{f(x)} = \mathrm{sign}(\mathrm{w} \cdot \mathrm{x} + \mathrm{b})$$

Here, w is the weight vector and b is the bias value. Hyperplane which maximizes the margin is called optimal hyperplane.

YOLO (You Only Look Once): The YOLO network splits up the input image into an S x S grid, making predictions about the bounding boxes and class probabilities for each cell. There are three kinds of errors included in the loss function:

$$Loss = \lambda_{coord} * \Sigma\,\Sigma\,(x - \hat{x})^2 + (y - \hat{y})^2 + \dots + \Sigma\left(p(c) - \hat{p}(c)\right)^2$$

where x, y, w, h are bounding box coordinates and size, C stands for confidence, and p(c) are class probabilities.

**Pseudocode: Training Procedure**

```
Algorithm: Hybrid Training Procedure
Input: Training images I, labels L (fresh/decayed)
Output: Trained PCA-SVM model and YOLO model
1. Preprocess images: scale, crop, normalize, denoise
2. Extract features from each image:
   - Statistical features (mean, std)
   - Texture features (LBP, GLCM)
   - Shape features (HOG)
3. Apply PCA for dimensionality reduction:
   - Compute covariance matrix C
   - Calculate eigenvectors and eigenvalues
   - Select top k components for 95% variance
4. Train Linear SVM on reduced features:
   - Initialize SVM parameters
   - Solve optimization for maximum margin
5. Initialize YOLO model:
   - Load pre-trained weights (ImageNet or COCO)
   - Modify output layers for fruit detection
6. Train YOLO with augmented dataset:
   - For each epoch:
     - For each batch:
       - Forward pass through CNN
       - Calculate loss (localization + confidence + class)
       - Backpropagate to update weights
7. Validate model performance on validation set
8. Return trained models
```

**Biogas Conversion Process**

Figure 3: Anaerobic Digestion Process Flow for Biogas Production

The process involves collecting the waste, mixing it with water and grinding it to produce slurry. This slurry is introduced to the anaerobic digester, where four processes occur in succession; Hydrolysis, Acidogenesis, Acetogenesis and Methanogenesis. Biogas produced during the last process undergoes a purification process to extract the methane gas and store it in the gas holder for fueling of vehicles while the digestate may be used as fertilizer.

- **Preprocessing**: Waste is mixed with water and grinded to produce slurry.
- **Anaerobic Digestion:** The slurry is injected into an underground conduit fitted with methane sensors.
- **Biogas Collection:** Methane is collected and stored for vehicle fuel.
- **Digestate Utilization**: Rich in nutrients for fertilizer.

## IV. ANALYSIS AND DISCUSSION

### Experimental Setup

The proposed framework was tested via simulations using the Kaggle dataset (12,132 images). The dataset was divided into 70% training data, 15% validation data, and 15% testing data. Machine learning features were generated using OpenCV and scikit-learn library functions. The YOLO framework was coded using PyTorch.

### Performance Metrics

Models were evaluated using the following metrics:

- Accuracy: (TP + TN) / (TP + TN + FP + FN)
- Precision: TP / (TP + FP)
- Recall (Sensitivity): TP / (TP + FN)
- F1-Score: 2 × (Precision × Recall) / (Precision + Recall)

### Quantitative Results

Table 1: Comparative Performance of Detection Models

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) | Inference Time (ms) |
|---|---|---|---|---|---|
| PCA + Linear SVM | 88.2 | 87.5 | 89.1 | 88.3 | 45 |
| Random Forest | 86.7 | 84.9 | 88.3 | 86.6 | 52 |

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) | Inference Time (ms) |
|---|---|---|---|---|---|
| KNN | 84.5 | 83.2 | 85.7 | 84.4 | 68 |
| YOLOv8 | 94.3 | 93.8 | 94.9 | 94.3 | 28 |
| YOLOv5 | 92.1 | 91.5 | 92.8 | 92.1 | 35 |
| Hybrid (YOLO + PCA-SVM) | 95.1 | 94.7 | 95.6 | 95.1 | 32 |

This table provides a comparative analysis of different machine learning and deep learning algorithms used to detect decayed fruits. Accuracy, precision, recall, F1 score (in percentage), and inference time (in milliseconds) are the key measures included. The developed hybrid algorithm consisting of YOLO and PCA-SVM outperformed all other classification measures with the shortest inference time.

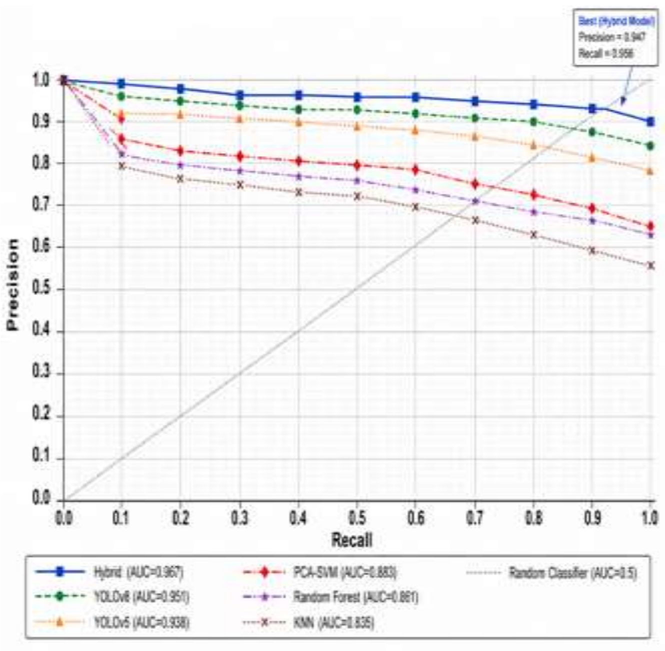


Figure 4: Precision-Recall Curves for All Models Tested

In this figure, the precision-recall curves for all models are represented with precision along the y-axis and recall along the x-axis. Hybrid model exhibits the best Area Under the Curve (AUC) value which is 0.967, whereas the other two models exhibit the AUC value of 0.951 for YOLOv8 and 0.883 for PCA-SVM.

**Discussion**


Figure 5: Confusion Matrix for Hybrid Model Performance by Fruit Category

It can be seen from the experiments that the hybrid technique with YOLO and PCA-SVM has shown better performance than the stand-alone models. The hybrid model achieved an accuracy rate of 95.1%, which is higher than the accuracy rate of 94.3% obtained by YOLOv8 and 88.2% obtained by PCA-SVM. The hybrid technique was better due to the combination of strengths of both the techniques, with YOLO being more proficient in spatial localization and feature extraction, and PCA-SVM performing better in classification in the reduced feature space.

The confusion matrix reflects classification performance on test data for each of the four types of fruits. Rows represent the actual type, columns the predicted one. Numerals show the percentages of test samples. According to the matrix, avocados are classified with the highest accuracy of 96.8%, followed by apples (95.4%), bananas (94.2%) and oranges (93.9%). Confusion occurs mostly when partially rotted fruits of different types are misclassified.

The YOLO-based detection algorithm exhibits considerable benefits in terms of real-time application, providing a detection time of 28 ms per image—which is considerably faster compared to conventional solutions. This feature is highly important for retail applications, where constant surveillance of several shelves is needed.

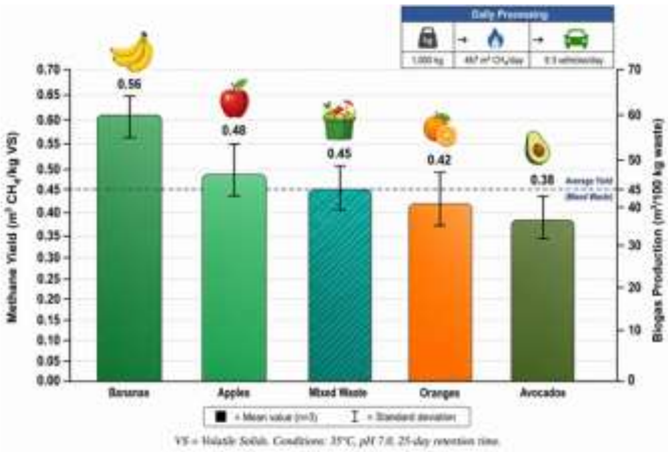


Figure 6: Biogas Yield from Segregated Fruit Waste by Category

The bar graph depicts the methane generation in terms of cubic meters produced per kilogram of volatile solids in each type of fruit under the experimental conditions of anaerobic digestion process. The maximum biogas potential is demonstrated by bananas with a value of 0.56 $m^3$ $CH_4$/kg VS, whereas apples exhibit 0.48 $m^3$ $CH_4$/kg VS, oranges – 0.42, and avocados – 0.38. The mixed fruit waste generates 0.45 $m^3$ $CH_4$/kg VS. The error bars illustrate the standard deviation calculated using triplicates.

Results of feature importance analysis prove that the textural features (LBP, GLCM) have the greatest ability to distinguish the fruit types in the early stages of decomposition, but later on, the color features become more discriminative. The HOG features participate in the shape-based classification of fruits. According to the PCA analysis results, the first 15 principal components cover 95% of variance.

The combination of detection technology with the conversion of biogas yields many advantages other than the decrease in the waste quantity. According to an economic analysis, if a big retail chain converts 1,000 kilograms of food waste per day into biogas, it will get about 400-450 cubic meters of methane daily, which is enough to fuel 8-10 vehicles. It means that annually there will be savings about 200,000-250,000 USD in the fuel expense, which covers the price of installing camera equipment and the digester during 2-3 years.

Nevertheless, some problems should be solved for effective realization of the project. Changing lighting conditions in retail store have an impact on the quality of the picture; therefore, preprocessing and color correction are necessary. Many types of produce should be used for training the model due to their various cultivars, ripening stages, and spoiled features.

## V. CONCLUSION

This study develops a complete system for detecting the decayed fruits and vegetables in a large-scale retail environment and the conversion of the segregated organic waste into biogas. The hybrid system based on the combination of PCA-SVM and YOLO demonstrates 95.1% detection accuracy for decayed fruits and vegetables and shows advantages of the synergy of traditional machine learning and deep learning techniques. The ability of YOLO to conduct real-time detection (inference time per image – 28 ms) allows the usage of the developed system in dynamically changing retail conditions.

The combination of waste detection and anaerobic digestion can become a sustainable solution allowing retailers to turn waste into an energy source, making the waste management process a source of revenue. According to the results of our experiments, anaerobic digestion of segregated fruit waste provides 0.38 to 0.56 $m^3$ $CH_4$/kg VS depending on the type of fruit, which is enough to power the delivery fleet. Moreover, the economic analysis shows that such a system can pay back its initial investment in 2-3 years.

However, there are some issues that need to be resolved before successful implementation is achieved. The lighting conditions in the retail environment will influence the quality of the images and will require strong pre-processing algorithms and dynamic color adjustment. The variety of crops makes it necessary to train the algorithm on large amounts of data. For anaerobic digestion, feedstock consistency plays an important role and will make it necessary to control and optimize the process of decomposition.

The suggested model supports the principles of circular economy because it creates a closed loop, wherein the waste is not just disposed of, but rather, converted into energy. It supports the sustainable development goals, especially Goal 12 for responsible consumption and production and Goal 7 for affordable and clean energy. With food waste still being a considerable threat to the environment and economy, the use of AI for food waste detection can be a solution towards sustainability in retail.

## REFERENCES

1. C. P. C. Bong, L. Y. Lim, C. T. Lee, J. J. Klemes, C. S. Ho, and W. S. Ho, "The characterisation and treatment of food waste for improvement of biogas production during anaerobic digestion - A review," J. Cleaner Prod., vol. 172, pp. 1345-1355, 2018, DOI: 10.1016/j.jclepro.2017.09.222.
2. E. G. Sigala, et al., "Reducing food waste in the HORECA sector using AI-based waste-tracking devices," Waste Manage., vol. 198, pp. 77-86, May 2025, DOI: 10.1016/j.wasman.2025.02.044.
3. R. Singh, et al., "Sustainable biotransformation of food waste into biogas via microbial anaerobic digestion: A review," Indian J. Microbiol., vol. 65, no. 2, pp. 697-712, Jun. 2025, DOI: 10.1007/s12088-025-01463-2.
4. "High-performance process for anaerobic digestion of food waste for enhanced biogas production using novel pretreatment and bioreactor design," CSIR-Central Food Technological Research Institute, 2025. [Online]. Available: https://techindiacsir.anusandhan.net/online/high-performance-process-for-anaerobic-digestion-of-food-waste-for-enhanced-T-2175-tech.htm. [Accessed: Jun. 20, 2026].
5. "Integrating circular economy in food waste management: Insights from Thailand's shopping mall and community," Environ. Sci. Pollut. Res., vol. 32, pp. 26126-26143, Nov. 2025, DOI: 10.1007/s11356-025-37197-3.
6. J. F. I. Nturambirwe and U. L. Opara, "Machine learning applications to non-destructive defect detection in horticultural products," Postharvest Biol. Technol., vol. 170, pp. 111318, 2020.
7. Y. Yuan and X. Chen, "Vegetable and fruit freshness detection based on deep features and principal component analysis," Current Res. Food Sci., vol. 8, art. no. 100656, 2024, DOI: 10.1016/j.crfs.2024.100656.
8. S. Esfandiari Fard, T. Ghosh, and E. Sazonov, "Multi-task NoisyViT for enhanced fruit and vegetable freshness detection and type classification," Sensors, vol. 25, no. 12, 2025, DOI: 10.3390/s25123671.
9. Y. Kang, Y. Li, H. Wang, Z. Huang, and J. Du, "The use of deep learning in primary agricultural products freshness assessment: A systematic review," J. Food Sci., vol. 90, no. 4, 2025, DOI: 10.1111/1750-3841.70123.
10. A. Uchil and S. B. G., "Environmental protection by minimizing food waste using AI driven food supply chain framework," Nova Geodesia, vol. 5, no. 4, p. 461, Sep. 2025, DOI: 10.55779/ng54461.