Leveraging IoT and Machine Learning for Automated Fruit Quality Monitoring: A Scalable Approach for Supply Chain Optimization

S. Rama Subbanna
Eluru College of Engg. and Technology
Andhra Pradesh, India

S. Suresh
Professor, Dept. of CSE
Eluru College of Engg. and Technology
Andhra Pradesh, India

ABSTRACT

Food wastage is a pervasive issue in day-to-day life, contributing significantly to environmental and economic challenges. This research introduces a novel Internet of Things (IoT) and Machine Learning (ML)-driven framework aimed at addressing food loss in supply chains and everyday settings through automated fruit quality monitoring and spoilage detection. Utilizing IoT sensors and an ESP32 microcontroller, the system collects real-time environmental data such as temperature, humidity, and gas emissions to classify fruit ripeness stages. By providing timely and accurate predictions, this framework enables individuals, retailers, and supply chain operators to take proactive measures to reduce wastage. Advanced ML models, including Random Forest and CatBoost, ensure exceptional accuracy in identifying ripeness and spoilage. This system not only minimizes human error but also enhances supply chain efficiency and promotes sustainable practices. By automating the monitoring process, this research offers a scalable and practical solution to prevent food waste, ensuring better resource utilization and contributing to global food security. Furthermore, it outlines future applications, including blockchain integration for end-to-end transparency in the food industry.

Keywords

IoT, Machine Learning, Fruit Quality Monitoring, Supply Chain Optimization, Food Wastage Reduction, Spoilage Detection

1. INTRODUCTION

Food waste and loss represent critical global challenges, significantly affecting environmental sustainability, economic efficiency, and food security, particularly in developing nations such as India. Reports indicate that approximately 40% of fruits and vegetables are wasted annually due to inefficiencies in storage, transportation, and ripeness monitoring [1, 3]. This substantial wastage not only contributes to resource scarcity but also results in extensive greenhouse gas emissions, particularly from perishable agricultural produce like fruits and vegetables [2].

Traditional approaches to assessing fruit ripeness and spoilage, such as manual visual inspection, are labor intensive, subjective, and error-prone. These methods lead to increased spoilage rates, inefficiencies in supply chain operations, and diminished consumer satisfaction. Bananas, as an example, are particularly susceptible due to their rapid ripening process, necessitating timely interventions to minimize waste [4, 8].

The integration of Internet of Things (IoT) and Machine Learning (ML) technologies provides new opportunities to address these challenges. IoT enables real-time data collection, such as temperature, humidity, and gas emissions, while ML models facilitate automated and accurate quality monitoring [6, 9]. Studies have demonstrated the effectiveness of ML-based image classification techniques for detecting fruit ripeness and freshness [5, 7]. Despite these advancements, existing solutions often face barriers, including high implementation costs and limited scalability, particularly in small-scale and resource-constrained settings [1, 3].

This research aims to overcome these challenges by proposing a novel IoT-enabled, ML-driven framework for real-time fruit quality monitoring, focusing on bananas as a case study. The framework integrates an ESP32 microcontroller with cost-effective sensors, including the SHT40 (to measure temperature and humidity) and the SGP30 (to detect gas emissions), for environmental data collection. Advanced ML models, such as Random Forest and CatBoost, are employed to classify ripeness stages and detect spoilage, automating the entire process to reduce human error, optimize supply chain efficiency, and promote sustainability.

By bridging the gap between affordability and technological innovation, the proposed system offers a practical and scalable solution to reduce post-harvest losses and ensure better resource utilization. Furthermore, it lays the groundwork for potential blockchain integration to enhance traceability and accountability across the food supply chain. This study contributes to addressing global food waste challenges by offering a framework applicable to diverse agricultural contexts, from small local markets to industrial supply chains.

The aim of this research is to develop a cost-effective, IoT-enabled, and machine-learning-driven framework for real-time fruit quality monitoring, with a focus on classifying ripeness stages and detecting spoilage. By leveraging affordable hardware and advanced ML algorithms, the proposed system seeks to address inefficiencies in the agricultural supply chain, reduce post-harvest losses, and promote sustainable practices in food production and distribution.

The following are the key contributions of this work:

1. **Development of an IoT-Enabled Framework for Real-Time Monitoring:**The research introduces an innovative system combining the ESP32 microcontroller with SHT40 (temperature and humidity) and SGP30 (gas emissions) sensors to collect critical environmental data in real time, enabling accurate monitoring of fruit quality.

2. **Integration of Advanced Machine Learning Models:**The study leverages state-of-the-art machine learning models, including Random Forest and CatBoost, to classify fruit ripeness stages and detect spoilage with high accuracy,

minimizing the reliance on subjective and error-prone manual inspections.

3. **Focus on Cost-Effectiveness and Scalability:**The proposed solution prioritizes affordability and ease of deployment, making it suitable for both small-scale and large-scale agricultural operations, particularly in resource-constrained settings.

4. **Targeted Application on High-Wastage Crops:**Bananas are used as the focal crop for testing the system's capabilities, illustrating its potential impact on mitigating waste for highly perishable fruits that are integral to global agriculture.

5. **Potential for Supply Chain Optimization and Sustainability:**By automating ripeness monitoring and enabling timely interventions, the framework reduces post-harvest losses, optimizes supply chain efficiency, and supports sustainable agricultural practices.

2. LITERATURE SURVEY

The integration of Internet of Things (IoT) and Machine Learning (ML) technologies has revolutionized the assessment of fruit ripeness and quality. IoT enables real-time data collection, while ML provides efficient and accurate analysis of complex datasets. Various studies have explored different methodologies and approaches to enhance the monitoring of fruit ripeness and spoilage. The following sections summarize the findings, methodologies, and outcomes of significant investigations.

Table 1: IoT-Based Investigations for Fruit Ripeness and Quality

Case Study	Objectives & Outcomes	References
IoT-Based Fruit Quality Inspection and Lifespan Detection System	Proposed a system for monitoring fruit quality and estimating lifespan using IoT-enabled sensors.	[15]
E-nose: A Low-Cost Fruit Ripeness Monitoring System	Developed an electronic nose using low-cost sensors to monitor fruit ripeness with 97.05% accuracy.	[16]
Banana ripeness stage identification: a deep learning approach	Develop a Deep Learning Model for Banana Ripeness Detection.	[14]
Developing an IoT and ML-Driven Platform for Fruit Ripeness Evaluation and Spoilage Detection	Focused on bananas to monitor ripeness stages and predict spoilage using IoT sensors and ML algorithms.	[13]

Table 1 represents various IoT-based investigations focused on assessing fruit ripeness and quality. It outlines the technologies, methods, and parameters utilized for real-time monitoring and analysis of fruit characteristics, ensuring optimal harvest timing and product quality.

Table 2 presents IoT and machine learning (ML)-based investigations for assessing fruit ripeness and quality. It details the integration of IoT devices with ML algorithms to enhance the accuracy and efficiency of fruit quality assessments and ripeness predictions.

To address these limitations, this study proposes a comprehensive IoT and ML-based framework for realtime monitoring of fruit ripeness and spoilage, focusing on bananas as a case study. The framework integrates IoT sensors (e.g., SHT40, MQ gas sensors) with advanced ML algorithms (e.g., Random Forest, CatBoost)

- **Enhance Scalability:** Implement solutions suitable for retail trays, cold storage, and supply chain logistics.

- **Improve Accuracy:** Utilize advanced ML models for precise ripeness classification based on gas emissions, temperature, and humidity data.

- **Reduce Food Waste:** Provide actionable insights for supply chain optimization and reduce spoilage losses.

Table 2: IoT and ML-Based Investigations for Fruit Ripeness and Quality

Case Study	Objectives & Outcomes	References
Detection of Freshness of Fruits Using Machine Learning Techniques	Proposed ML techniques for identifying freshness in fruits to improve food quality management.	[7]
Classification of Cape Gooseberry Fruit According to Ripeness Levels	Employed machine learning and color spaces to classify cape gooseberries by ripeness stages.	[6]
Tomato Classification Using K-NN, MLP, and K-Means Clustering	Utilized machine learning to classify tomatoes based on organoleptic maturity using color analysis.	[8]
Pixel-Based Color Image Classification for Tomato Ripeness	Developed a machine learning-based system to detect tomato ripeness stages using pixel-based image data.	[9]
Fruit Ripeness Detection Using Convolutional Neural Networks	Employed CNN models for ripeness detection to improve accuracy and reduce human error in manual grading.	[12]
Fruit Ripeness Detection Method Using Deep Learning	Adapted deep learning models to enhance ripeness stage classification for multiple fruits.	[11]
Ripe Fruit Detection and Classification using Machine Learning.	Applied ML algorithms like K-NN, SVM, and Decision Trees for ripeness classification.	[10]

3. METHODOLOGY

The proposed framework integrates IoT hardware, cloud-based data storage, machine learning (ML) algorithms, and supply chain management strategies to provide an automated solution for monitoring fruit quality. This scalable and modular system enhances storage optimization, reduces fruit wastage, and facilitates better decision-making across the supply chain. The framework is inspired by advancements in IoT and ML latforms for fruit ripeness evaluation and spoilage etection, such as the study conducted by [14], focusing on bananas as a case study.

Figure 1 demonstrates the IoT-based fruit sensing model designed for optimizing supply chain processes. It highlights the workflow, including sensor-based data collection, cloud integration for data storage and analysis, an machine learning-driven predictions for fruit quality assessment.

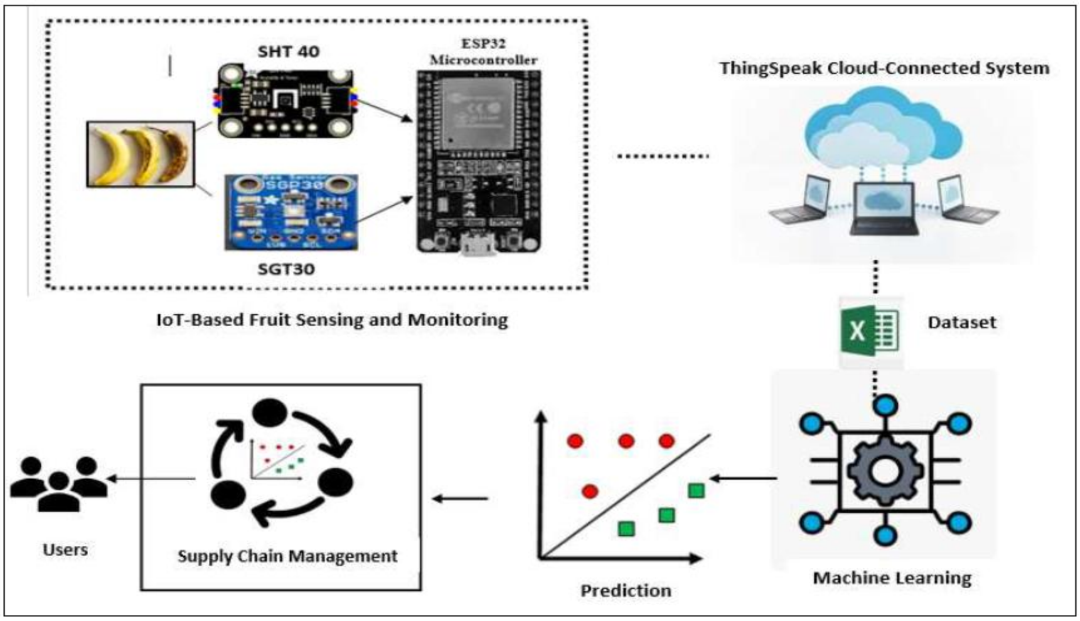

Figure 1: IoT-Based Fruit Sensing Model for Supply Chain Optimization

3.1 Sensor Node Design*:* The sensor node is designed to monitor environmental parameters that influence fruit ripeness and spoilage .It integrates the following components for efficient data collection and wireless communication:

Hardware Components:

• **ESP32 Microcontroller**: Functions as the core processing and communication unit, offering builtin Wi-Fi capabilities for seamless data transmission. It serves as the foundation for managing and processing data collected from the connected sensors.

• **SHT40 Sensor**: Measures temperature and humidity with high precision using capacitive sensing technology. The SHT40 is interfaced with the ESP32 microcontroller via the I2C protocol, enabling efficient retrieval of data. Temperature and humidity readings are processed using mathematical formulas to derive relative humidity and temperature values.

• **SGP30 Sensor**: A compact and advanced gas sensor designed for detecting Total Volatile Organic Compounds (TVOC) and Carbon Dioxide Equivalent (CO2eq). These measurements are crucial indicators of fruit ripeness and spoilage. The SGP30 is also connected to the ESP32 using the I2C protocol, ensuring accurate and efficient gas monitoring.

3.2 Data Collection:

The sensor node collects real-time environmental parameters such as temperature, humidity, and gas concentrations emitted during the ripening process. Using the ESP32's built-in Wi-Fi capabilities, the captured data is transmitted wirelessly to the cloud platform for further analysis and storage.

Connectivity and Functionality:

The ESP32 microcontroller not only processes the data from the sensors but also handles wireless communication with cloud platforms, enabling real-time monitoring and storage. The firmware for the sensor node is developed and deployed using the Arduino IDE, providing a robust system for managing and transmitting the collected data.

Experimental Setup:

The experimental setup involved placing the sensor node within a controlled environment, such as a plastic container, to monitor the gases emitted by fruits like bananas during ripening. This design highlights the versatility and efficiency of the sensor node, making it suitable for IoT-based applications in fruit quality monitoring and supply chain optimization.

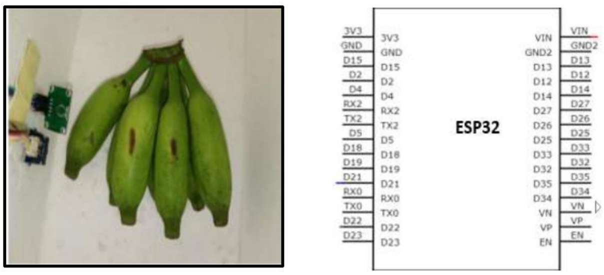

Figure 2: Green Banana with Gas Sensor Setup, ESP32 Pinout Diagram

Figure 2 shows a schematic representation of the sensor node setup, highlighting the SGP30 gas sensor and the SHT40 temperature/humidity sensor connected to the ESP32 microcontroller. This configuration enables real-time data acquisition for monitoring environmental parameters critical to fruit ripeness detection.

Please use a 9-point Times Roman font, or other Roman font with serifs, as close as possible in appearance to Times Roman in which these guidelines have been set. The goal is to have a 9-point text, as you see here. Please use sans-serif or non-proportional fonts only for special purposes, such as distinguishing source code text. If Times Roman is not available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times. Right margins should be justified, not ragged.

3.3 ThingSpeak Cloud Platform

The collected data is transmitted wirelessly to a cloud-based platform, ensuring secure and efficient storage and real-time accessibility. ThingSpeak acts as the primary data storage and management system, enabling seamless integration and analysis of both real-time and historical data from the sensor nodes. This allows stakeholders to track trends, identify patterns, and make data-driven decisions regarding fruit quality and storage conditions. The platform provides interactive and user-friendly dashboards, equipped with visually appealing graphs and charts, to monitor key parameters such as temperature, humidity, and gas concentrations in real time.

Customizable alerts and notifications can be configured to inform users when critical thresholds are breached, ensuring timely action to prevent spoilage and optimize the supply chain. Moreover, ThingSpeak supports API integration, enabling compatibility with machine learning models and other analytical tools for advanced insights and automation. This centralized approach facilitates efficient data handling and contributes to a scalable and robust monitoring system for agricultural applications.

3.4 Data Pre-processing and Analysis

This research employed comprehensive data pre-processing and analysis methods to ensure the quality and reliability of the dataset, enabling an in-depth exploration of the banana ripening process and the utility of sensor-based monitoring systems in agriculture. The raw sensor data was meticulously cleaned, standardized, and transformed into a well-structured format suitable for both statistical analysis and machine learning model development.

• **Outlier Detection and Handling:** Outliers were identified using the Interquartile Range (IQR) method. This involved computing the IQR and removing data points lying outside the acceptable range, thus preventing skewed analysis.

• **Date Standardization:** Timestamps were reformatted into a uniform structure, facilitating accurate time-series analysis and synchronization across multiple sensors.

• **Handling Missing Values:** Missing entries were systematically identified and addressed, either through interpolation or removal, ensuring that the resulting dataset was both complete and reliable for subsequent analyses.

• **Labeling Ripeness Stages:** Based on sensor readings (e.g., TVOC, CO2) and visual inspections (e.g., peel color, texture), bananas were assigned to one of three ripeness stages:

o **Not Ripe**:

▪ Visual Indicators: Greenish peel, firm texture.

▪ Sensor Readings: Low TVOC levels and minimal CO2.

o **Ripe**:

▪ Visual Indicators: Bright yellow peel, softer yet still relatively firm texture.

▪ Sensor Readings: Moderate increases in TVOC and CO2.

o **Spoiled**:

▪ Visual Indicators: Brown spots or fully brown peel, very soft and mushy texture.

▪ Sensor Readings: Significantly elevated TVOC and CO2 levels.

3.5 Machine Learning Module: The machine learning module is designed to classify fruit ripeness into

three categories—**Not Ripe**, **Ripe**, and **Spoiled**—using sensor data (e.g., temperature, humidity, gas emissions). This section outlines each phase of the process, from data preparation to final evaluation.

3.5.1. Data Preparation and Labeling

Relevant variables (temperature, humidity, CO_2, TVOC, etc.) were chosen based on their correlation with the ripening process. Each data instance was assigned one of three labels—Not Ripe, Ripe, or Spoiled— using both sensor readings and visual inspections.

5.3.2. Splitting the Dataset

• **Train-Test Split**: The labeled dataset was divided into training (e.g., 80%) and testing (e.g., 20%) subsets.

• **Cross-Validation**: To enhance reliability, k-fold cross-validation was employed. This technique systematically partitions the training data into k subsets (folds), ensuring every data point is used for both training and validation at least once.

5.3.3. Model Selection and Hyperparameter Tuning:

To identify the most accurate and efficient approach for ripeness classification, multiple algorithms were explored, including Random Forest, Support Vector Machines (SVM), CatBoost, K-Nearest Neighbors, (KNN), XGBoost, Decision Tree, Logistic Regression, and Naive Bayes.

• **Hyperparameter Tuning:** Grid search and randomized search methods were employed to determine the optimal combination of parameters for each algorithm. For instance, the number of trees in Random Forest, kernel type in SVM, and learning rate in boosting algorithms were finetuned to enhance performance.

• **Scoring Metrics:** The models were evaluated using multiple metrics such as accuracy, precision, recall, and F1-score. This multi-dimensional evaluation ensured a comprehensive and balanced assessment of each algorithm's performance, providing insights into their strengths and limitations

5.3.4. Training and Evaluation Process

Table 3: Model Performance Overview

Case Study	Objectives & Outcomes	References
Random Forest	93.4	Ensemble method that handles non-linear relationships well; good at managing diverse feature sets.
SVM	90.1	Effective for well-defined class boundaries; requires careful kernel choice and parameter tuning.

CatBoost	94.7	Handles categorical data effectively; often achieves high accuracy with minimal tuning.
KNN	88.6	Easy to implement; performance depends heavily on choice of k and distance metric.
XGBoost	92.8	Robust gradient-boosting method; efficient for larger datasets; can require detailed hyperparameter tuning.
Random Forest	93.4	Ensemble method that handles non-linear relationships well; good at managing diverse feature sets.
SVM	90.1	Effective for well-defined class boundaries; requires careful kernel choice and parameter tuning.

During this phase, each algorithm was trained on the labeled dataset and validated using either a hold-out test set or cross-validation. The overall accuracy in classifying the three ripeness stages, along with key observations. Simpler models like Logistic Regression and Naive Bayes were easier to train but lagged in accuracy for this dataset.

5.3.5. Supply Chain Management: The system incorporates features to optimize supply chain operations:

• Real-Time Monitoring:
Tracks fruit quality at every stage of the supply chain, from storage to delivery.
• Optimized Storage:
Dynamic adjustments to storage parameters such as temperature and humidity are made to preserve fruit freshness.
• Inventory Management:
Fruits are categorized based on ripeness to prioritize distribution and avoid spoilage.
• Logistics Optimization:

Predictive analytics are used to determine the optimal time for transportation and distribution.

5.3.6. User Interface: The user interface ensures seamless interaction with the system:

• Applications:

Both mobile and web-based apps are provided for stakeholders, offering insights into fruit quality and trends.
• Notifications:
Alerts (via SMS and email) are sent when specific thresholds, such as high CO2 levels indicating spoilage, are breached.

This framework serves as an end-to-end solution for fruit quality monitoring, paving the way for enhanced decision-making and sustainable supply chain practices.

4. RESULTS

Figure 3 presents the confusion matrix for the best-performing model, CatBoost. The diagonal elements represent correct predictions for each ripeness class (Not Ripe, Ripe, Spoiled). Notably, the model correctly identified a significant portion of "Spoiled" fruits (51 correct classifications), indicating strong sensitivity to higher CO_2 and TVOC levels associated with spoilage.

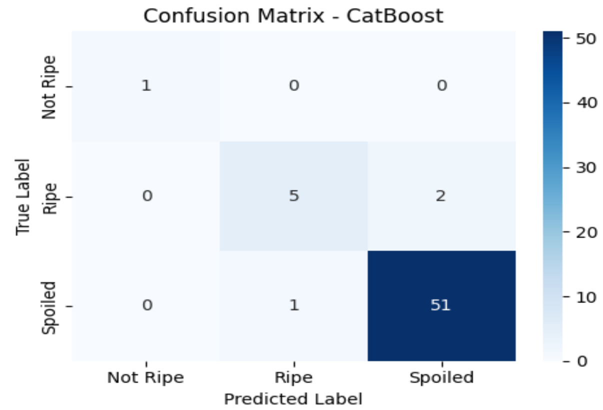

Figure.3:
However, a smaller number of misclassifications occurred in the "Not Ripe" and "Ripe" classes, suggesting these categories may share overlapping features in certanges of gas sensor readings.

As shown in **Figure 4**, multiple classification algorithms (CatBoost, Random Forest, KNN, XGBoost, Decision Tree, Naive Bayes, Logistic Regression, and SVM) were evaluated on the fruit ripeness dataset. CatBoost achieved the highest accuracy, surpassing 90%, while SVM had the lowest performance among the tested models. These findings demonstrate that ensemble-based methods (CatBoost and Random Forest) generally perform better for this particular dataset, likely due to their robustness against diverse feature distributions.

Figure 5 depicts synthetic temperature data over 300 samples, illustrating random fluctuations between 20°C and 30°C. Although temperature was not directly used for classification in this synthetic scenario, real-world systems often rely on temperature as a key variable influencing the rate of fruit ripening. Monitoring temperature trends can help stakeholders intervene proactively to maintain optimal storage conditions.

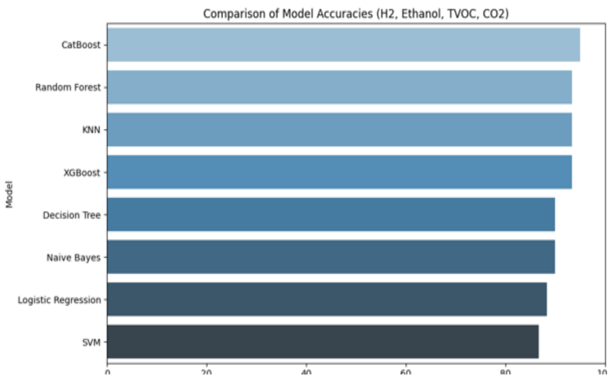

Figure 4: Comparison of Model Accuracies H2, Ethanol, TVOC, CO2)

Figure 5 provides a summary of how the best-performing model's predictions can guide routing and storage choices. In this example, a large number of batches are classified as "Spoiled," leading to a decision of "Discard/Compost." Only a small portion are sent to "Short-Term Storage," while some are routed to the "Nearest Market" based on the "Ripe" classification. Although this is a simplified illustration, it demonstrates how predictive analytics can be integrated into supply chain workflows to minimize waste and optimize resource allocation.

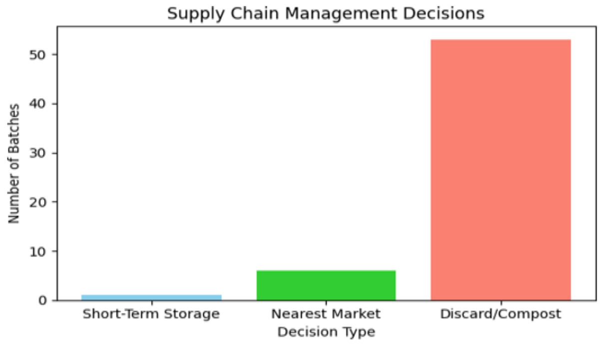

Figure 4: Supply Chain Management Decisions

5.CONCLUSION

In conclusion, this research presents an innovative and scalable approach to reducing food waste through the integration of IoT and Machine Learning technologies. By leveraging real-time environmental data and advanced ML algorithms, the proposed framework efficiently monitors fruit quality, detects spoilage, and predicts ripeness, ultimately enhancing supply chain operations and minimizing food wastage. The system's automation reduces human error and ensures proactive decision-making, offering a practical solution for retailers, consumers, and supply chain operators alike. Furthermore, the potential for future advancements, such as blockchain integration, promises even greater transparency and traceability within the food industry. Overall, this research contributes to the optimization of resource utilization, promotes sustainable practices, and supports the broader goal of global food security.

6. REFERENCES

[1] Sahoo, A., Dwivedi, P., Madheshiya, U., Kumar, R.K., Sharma, S., Tiwari, Insights into the management of food waste in developing countries: with special reference to India, *Environ. Sci. Pollut. Res.*, 31(12), 17887–17913 (2024). https://doi.org/10.1007/S11356-023-27901-6.

[2] Guo, X., Broeze, J., Groot, J.J., Axmann, H., Vollebregt, M., A worldwide hotspot analysis on food loss and waste, associated greenhouse gas emissions, and protein losses, *Sustainability*, 12(2020), 7488. https://doi.org/10.3390/SU12187488.

[3] Gustavsson, J., Cederberg, C., Sonesson, U., Global food losses and food waste (2011).

[4] Onwude, D.I., Chen, G., Eke-Emezie, N., Kabutey, A., Khaled, A.Y., Sturm, B., Recent advances in reducing food losses in the supply chain of fresh agricultural produce, *Processes*, 8(11), 1431 (2020). https://doi.org/10.3390/PR8111431.

[5] Kumar, V., Sharma, K., Kedam, N., Patel, A., Kate, T.R., Rathnayake, U., A comprehensive review on smart and sustainable agriculture using IoT technologies, https://doi.org/10.1016/j.atech.2024.100487.

[6] Castro, W., Oblitas, J., De-La-Torre, M., Cotrina, C., Bazan, K., Avila-George, H., Classification of cape gooseberry fruit according to its level of ripeness using machine learning techniques and different color spaces, *IEEE Access*, 7, 27389–27400 (2019). https://doi.org/10.1109/ACCESS.2019.2898223.

[7] Jayasinghe, P.K.S.C., Sammani, S., Detection of freshness of the fruits using machine learning techniques, (2022), Accessed: Aug. 07, 2024. [Online]. Available: http://ir.lib.seu.ac.lk/handle/123456789/6413.

[8] Pacheco, W.D.N., López, F.R.J., Tomato classification according to organoleptic maturity (coloration) using machine learning algorithms K-NN, MLP, and K-Means clustering, 22nd Symposium on Image, Signal Processing and Artificial Vision, STSIVA 2019 - Conference Proceedings (Apr. 2019). https://doi.org/10.1109/STSIVA.2019.8730232.

[9] Garcia, M.B., Ambat, S., Adao, R.T., "Tomayto, tomahto": a machine learning approach for tomato ripening stage identification using pixel-based color image classification.

[10] Munsayac, A.D., AfricaDe, Ripe Fruit Detection and Classification using Machine Learning, https://doi.org/10.30534/ijeter/2020/60852020.

[11] Zhang, W., A Fruit Ripeness Detection Method using Adapted Deep Learning-based Approach, https://doi.org/10.14569/IJACSA.2023.01409121.

[12] Setiawan, F.B., Adipradana, C.B., Fruit Ripeness Detection Using Convolutional Neural Network, https://doi.org/10.33387/protk.v10i1.5549.

[13] Rajini, M., Voola, P., Developing an IoT and ML-driven platform for fruit ripeness evaluation and spoilage detection: A case study on bananas, https://doi.org/10.1016/j.prime.2025.100896.

[14] Saranya, N., Srinivasan, K., Banana ripeness stage identification: a deep learning approach, https://doi.org/10.1007/s12652-021-03267-w.

[15] Saha, A., Ali, L., IoT Based Fruit Quality Inspection and Lifespan Detection System, https://doi.org/10.1109/ICCUBEA58933.2023.10392254.

[16] Tyagi, P., Semwa, R., Tiwary, U.S., E-nose: a low-cost fruit ripeness monitoring system, https://doi.org/10.4081/jae.2022.1389.

[17] Fernandes, D.L.A., Oliveira, J.A.B.P., Gomes, M.T.S.R., Detecting spoiled fruit in the house of the future, https://doi.org/10.1016/j.aca.2008.01.068.