

Article

Not peer-reviewed version

Predicting Apple Quality Using Machine Learning Techniques

Adil Mehmood and [Noor Amin](#) *

Posted Date: 17 November 2025

doi: 10.20944/preprints202511.1154.v1

Keywords: apple quality; machine learning; classification models; agricultural technology; non-destructive evaluation



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Predicting Apple Quality Using Machine Learning Techniques

Adil Mehmood ¹ and Noor Amin ^{2,*}

¹ Department of Software Engineering, University of Sialkot, Sialkot, Pakistan

² School of Computer Science, Taylor's University, Subang Jaya, Malaysia

* Correspondence: nooraminnawab@gmail.com

Abstract

In order to optimize the agricultural supply chain and ensure customer satisfaction, evaluating apple quality is crucial. The use of machine learning (ML) algorithms offers a viable solution to the subjective and time-consuming nature of traditional quality assessment methods. To classify apples as "good" or "bad" quality, the dataset used in this study includes thorough evaluations of relevant characteristics. This study investigates the use of multiple machine learning models to assess apple quality based on characteristics such as size, weight, sweetness, acidity, ripeness, juiciness, and crunchiness. To build predictive models, machine learning techniques such as Naïve Bayes, decision trees, K-nearest neighbors (KNN), and AutoMLP were employed. Among these models, AutoMLP achieved the highest accuracy rate of 88.46%. These results provide a foundation for future advancements in fruit quality assessment and contribute to the progress of agricultural technologies.

Keywords: apple quality; machine learning; classification models; agricultural technology; non-destructive evaluation

1. Introduction

In the food and agriculture sectors, apple quality evaluation is crucial, as it directly impacts supply chain performance, market pricing, and customer satisfaction. Apples' nutritional value and availability make them one of the most popular fruits worldwide. However, assessing apple quality remains challenging due to its dependence on physical, chemical, and sensory characteristics such as size, weight, sweetness, acidity, and maturity.

Traditional methods for evaluating fruit quality often rely on destructive laboratory techniques or subjective human assessment, both of which can be time-consuming and unreliable [1,2]. With technological advancements, machine learning has emerged as a transformative tool across various domains, including agriculture [17]. It is now possible to develop efficient, non-destructive, and highly accurate models for predicting apple quality using machine learning techniques.

These models can preprocess large datasets and detect complex patterns that are often invisible to traditional methods. The implementation of these algorithms is facilitated by the availability of well-structured datasets. The dataset used in this study includes characteristics such as weight [18], size, sweetness, softness, juiciness, crunchiness, ripeness, and acidity, along with harvest time. The target variable is the quality classification of apples as "good" or "bad" [3]. Recent research has demonstrated the effectiveness of various machine learning algorithms in evaluating fruit quality. For example, decision trees and SVM have shown success in identifying sensory attributes, while CNN are used for hyperspectral imaging of fruits.

In this study, we investigate the performance of several machine learning models such as Naïve Bayes, decision trees, KNN, and AutoMLP—in predicting apple quality. Among these, AutoMLP achieved the highest accuracy with a notable prediction rate of 88.46% [4,5].

The significance of this study lies in its ability to provide an efficient, scalable, and non-destructive method for evaluating apple quality [19]. This work helps bridge the gap between

advanced predictive analytics and traditional quality control methods by combining machine learning algorithms[20–23] with structured domain knowledge. The findings have practical implications for farmers, distributors, and retailers, enabling improved decision-making throughout the supply chain[24–26]. This research contributes to the growing body of knowledge in agricultural technology and paves the way for future innovations in the field[27–29].

The structure of this paper is as follows: Section 2 presents the literature review; Section 3 explains the methodology; Section 4 discusses the results; and Section 5 concludes the paper and outlines future work. A detailed description of the dataset and its characteristics is provided in the following sections. Figure 1 shows the flow.

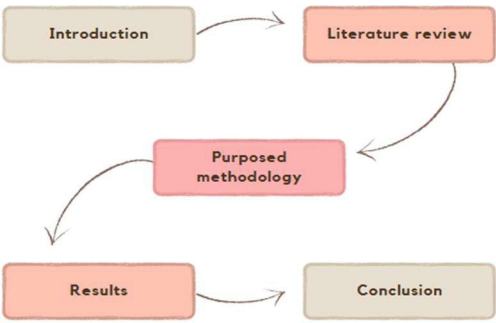


Figure 1. Paper Section Flow.

2. Literature Review

Before starting research into the area of apple quality, the authors find that there is some existing research related to the topic [20]. Rymenants et al. addressed the lack of understanding regarding the perception of sweetness and acidity by evaluating the genetic basis of apple sensory features. Using a 50K SNP array with FlexQTL analysis on full-sib families, the study finds QTLs, including known (Ma, Ma3) and novel loci (LG1, LG6). Ma3 showed partial dominance. The results are consistent with previous studies and improve the accuracy of QTL mapping using sensory data, offering crucial information for apple breeding initiatives.

Liu et al. combine electronic tongues and hyperspectral imaging to non-destructively forecast the sweetness and sourness of an apple. The study used the CARS–PSO–SVR model, which gives lower computational demands and best accuracy ($R^2 = 0.81$ for sourness and 0.887 for sweetness). This study builds on previous research by combining two state-of-the-art technologies and offers a practical, effective method for accurately assessing apple flavor [6].

As they [7] use an instrumental way to solve the problem of predicting the sensory characteristics for texture of apple, the accuracy of predictions may be limited by the need to choose acoustic and mechanical parameters in traditional methods. The paper recommends utilizing the entire mechanical and acoustic dataset from compression experiments along with contrasting machine learning techniques like MLP, SVR, and GPR, using more traditional approaches like PLS.

Byun in 2022 created an NIR spectroscopy module to estimate the non-destructive sugar content in apples. With an adjusted 13-bit ADC resolution and wavelengths (700–1000 nm), the small CMOS-based module achieved a SEC = 0.475 Brix and a correlation coefficient ($R^2 = 0.846$). The study fills the gap in NIR applications by offering a practical, integrated solution that overcomes the drawbacks of conventional large setups for fruit quality. This study closes the gap in applications of NIR [8].

Mignard et al. (2022) examined the combined effects of genetics and environment on apple sugars, using 155 accessions over a five-year period. Although factors like environmental temperature and UV radiation had favorable effects on metabolite concentrations, genetics had the greatest impact. Spanish cultivars were more diverse compared with other nations [9].

In this study, Pissard et al. (2021) finds the effectiveness of tabletop and portable NIR spectrometers for examining the quality of apples. By using techniques like calibration transfer, they

demonstrate that accuracy is similar across devices. In this study's prediction level, it becomes possible to make historical models with a portable spectrometer. This study fills key gaps in the non-destructive work of fruit quality by testing portable spectrometers for field usage and resolving calibration compatibility issues [3].

Analysis and sensory evaluations detect that "Topaz" lacked sweetness and appearance but was superior in flavor and juiciness. This study identifies gaps in previous research and fills the gaps in Bosnia and Herzegovina's regional data [10].

The efficiency of the supply chain and food security are significantly influenced by post-harvest losses in apple production. While NIR technology has been used to monitor apple quality, most research focuses on preservation over time. This study fills the gap by classifying apples according to freshness and variety following 14 days of short storage using NIR spectroscopy and multivariate data analysis. This work emphasizes brief storage, which provides important insights into the role of NIR in short food supply chains [1].

Fathizadeh et al. (2021) investigate a major problem in evaluating and determining fruit quality, ripeness, and shelf life by looking into nondestructive methods for assessing how firm the flesh of apple fruits is. The authors examine many techniques, including resistance-electrical, optical, and acoustic-vibration methods, with a focus on using artificial intelligence and data fusion to improve accuracy [11].

This study [9] explores the function of ROI selection in hyperspectral imaging to forecast the sugar content of apples. The circular ROI with a 25-pixel diameter had the best accuracy using Fuji apples and PLS regression ($R_c = 0.8977$, $R_p = 0.8836$). This work highlights ROI design and size as crucial parameters for enhancing prediction accuracy, advancing hyperspectral imaging applications for fruit quality evaluation [21].

Cao et al. use a rapid, simple, and non-destructive approach. The problem of determining the flavor of apple quality was tackled by conventional procedures, which are complex. This study looked at ethylene release and soluble solid content (SSC) as markers of taste quality in Fuji apples during storage and discovered that these variables were accurate predictors of changes in both flavor and aroma. The authors' approach simplifies flavor evaluation and provides a workable alternative to costlier, time-consuming techniques by using SSC and ethylene [12]

3. Proposed Methodology

The proposed methodology of this research involves using ML techniques to evaluate the quality of apples. Machine learning offers promising advancements across various fields, including industry and healthcare.

In this study, we utilize a dataset containing 9 attributes related to apple quality. Machine learning methods, including data splitting and feature selection, are applied to prepare the dataset for evaluation. Datasets often suffer from inconsistencies and duplications, which can negatively impact algorithm performance.

The dataset was obtained from Kaggle. Several algorithms were applied to the dataset, including KNN, decision tree, AutoMLP, and Naïve Bayes. The overall goal of our methodology is to create an effective and interpretable ML-based apple quality evaluation system that can assist industry professionals and farmers in identifying the quality of apples. The RapidMiner tool was used to implement and evaluate these machine learning algorithms.

3.1. Machine Learning Framework

A machine learning framework is an interface or environment that simplifies and accelerates the development of ML models[13–15]. It typically involves the following stages: data collection and preprocessing, model selection and training, and model evaluation and deployment [16].

In our case, we first collected and cleaned the apple quality dataset by replacing missing values and preparing the data for training and testing. This overall workflow is illustrated in Figure 2.

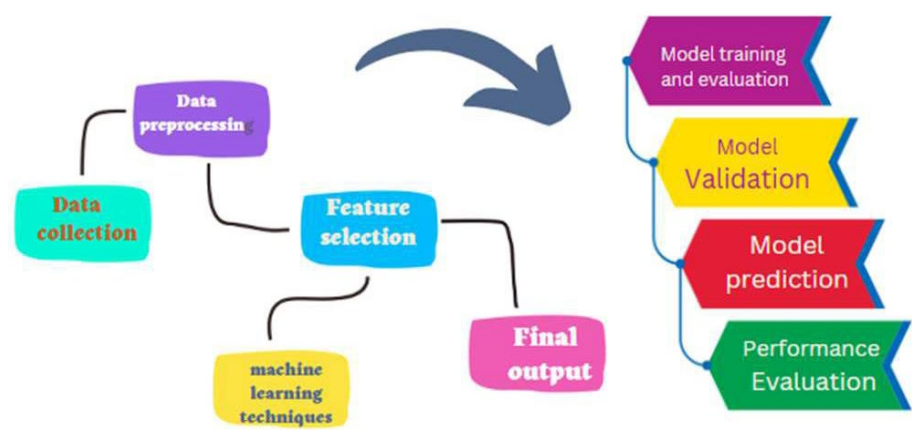


Figure 2. Machine Learning Framework.

3.2. Dataset Selection

The dataset used in this study was downloaded from Kaggle. It consists of **4001 rows** and **9 columns**, with attributes including:

- 1. Size
- 2. Weight
- 3. Sweetness
- 4. Softness
- 5. Harvest Time
- 6. Ripeness
- 7. Acidity
- 8. Quality

3.3. Machine Learning Algorithms

The performance of the selected algorithms was evaluated based on their accuracy using RapidMiner. Among the tested models, AutoMLP achieved the highest prediction accuracy.

3.4. Data Preprocessing

The dataset contained missing values that could negatively impact model accuracy. These missing values were handled using appropriate preprocessing operators in RapidMiner, as illustrated in Figure 3.

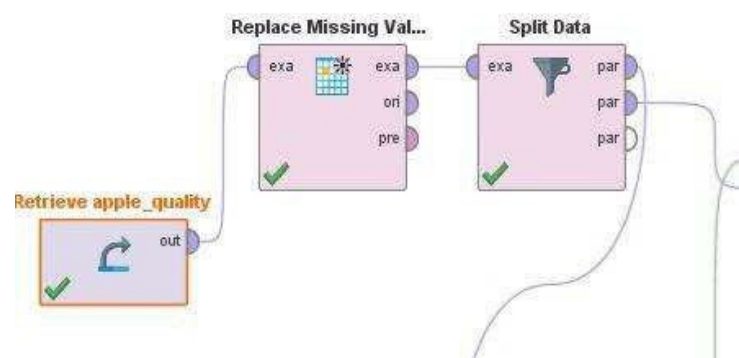


Figure 3. Handling Missing Values in RapidMiner.

3.5. Data Splitting

The dataset was divided into training, testing, and validation sets. Specifically, **70%** of the data was used for training, and **30%** for testing. Validation was performed on a subset of the training data. The data splitting process is shown in Figure 4.

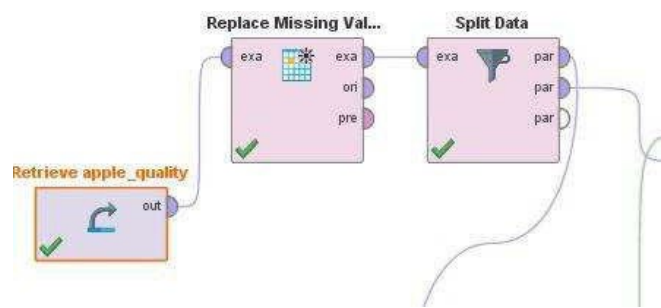


Figure 4. Splitting Data in RapidMiner.

3.6. Model Training and Evaluation

The final operator used in RapidMiner was AutoMLP, which achieved the highest accuracy compared to other algorithms. The full pipeline, including preprocessing, model selection, and evaluation, is summarized in Figure 5.

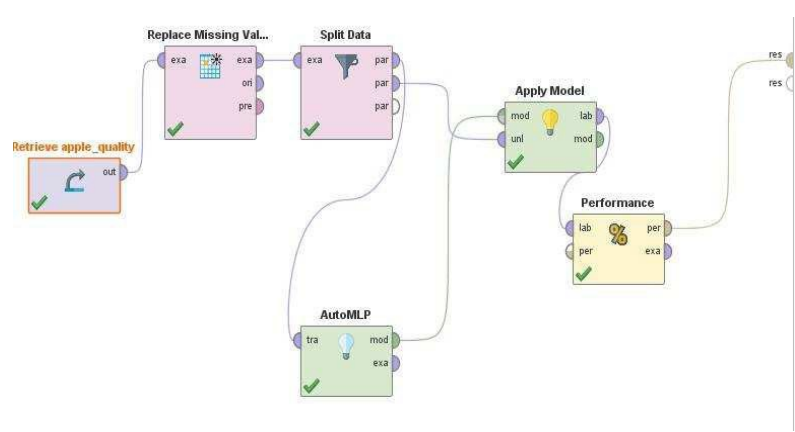


Figure 5. Overall Workflow in RapidMiner.

4. Results

In this part of the research, we elaborate on the performance of our dataset based on the results of applying various machine learning algorithms. To optimize accuracy, we used different algorithms such as KNNs, Decision Tree, AutoMLP, and Naïve Bayes. The confusion matrix is shown in Table 1 and each algorithm produced different accuracy results, as shown in Table 2.

Table 1. Confusion Matrix.

	GOOD	BAD
GOOD	2676	324
BAD	139	862

Table 2. Model Accuracy.

Algorithms	Accuracy (%)
Decision Tree	51.13%
AutoMLP	88.46%

KNNs	57.46%
Naïve Bayes	74.31%

From the results, it is clearly shown that AutoMLP achieved the highest accuracy compared to the other algorithms. The confusion matrix (Table 1) is crucial in evaluating performance by defining classification outcomes, including true positives, false positives, true negatives, and false negatives.

The depicted confusion matrix (Figure 6) visually represents the performance of AutoMLP in our model. In addition, Table 3 presents the class recall and class precision results.

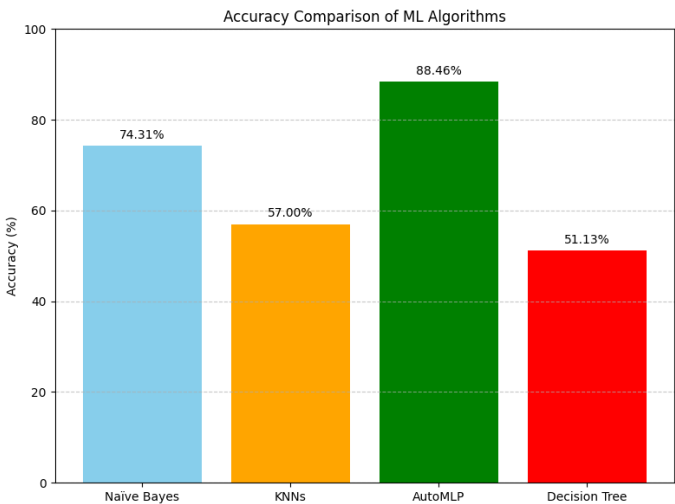


Figure 6. Accuracy Comparison (AutoMLP vs. Others).

Table 3. Comparison of Existing Schemes.

2020	Pedigreed families	-	-
2020	Hyperspectral + Electronic Tongue	Combined model	High
2024	Mechanical and Acoustic Profiles	ML Algorithms	High
2022	NIR spectroscopy data	Integrated NIR Model	High
2022	genetic and climate data	Statically analysis	Variable
2022	Portable NIR Spectroscopy	calibration transfer	High
2022	Consumer survey data	Sensory evaluation	-
2021	Apple firmness data	Non- Destructive method	Effective

5. Conclusions and Future Work

Our research paper focuses on assessing the quality of apple fruit using machine learning algorithms. By selecting different algorithms, applying preprocessing methods, performing feature selection, and using data splitting techniques, we achieved high accuracy on our dataset. The main goal was to identify a model with high accuracy that could effectively assess apple quality. The highest accuracy achieved was 88%, using the AutoMLP algorithm. Future work could focus on expanding the integrated machine learning framework to include additional data sources, such as

the effects of climate change on fruit quality, in order to further improve the accuracy and robustness of the model..

References

1. K. Włodarska, K. Pawlak-Lemańska, and E. Sikorska, "NIR technology for non-destructive monitoring of apple quality during storage," *Logforum*, vol. 20, no. 1, pp. 11–21, 2024, doi: 10.17270/J.LOG.000968.
2. W. Peng et al., "Sweetness classification of apple based on VIS spectroscopy combined with particle swarm optimized BP neural network," *SPIE-Int. Soc. Opt. Eng.*, Mar. 2023, p. 27, doi: 10.1117/12.2663147.
3. A. Pissard et al., "Evaluation of a handheld ultra-compact NIR spectrometer for rapid and non-destructive determination of apple fruit quality," *Postharvest Biol. Technol.*, vol. 172, Feb. 2021, doi: 10.1016/j.postharvbio.2020.111375.
4. K. Fujioka, "Evaluating the non-invasive measurement of apple aroma using electronic nose device through comparison with direct mass spectrometry, sugar content, and ripeness measurements," *Sensors*, vol. 24, no. 10, May 2024, doi: 10.3390/s24103114.
5. P. Mignard, S. Beguería, R. Giménez, C. Font i Forcada, G. Reig, and M. Á. Moreno, "Effect of genetics and climate on apple sugars and organic acids profiles," *Agronomy*, vol. 12, no. 4, Apr. 2022, doi: 10.3390/agronomy12040827.
6. J. Liu et al., "Detection of apple taste information using model based on hyperspectral imaging and electronic tongue data," *Sensors and Materials*, vol. 32, no. 5, pp. 1767–1784, 2020, doi: 10.18494/SAM.2020.2715.
7. R. Ricci, A. Berardinelli, F. Gasperi, I. Endrizzi, F. Melgani, and E. Aprea, "Combining algorithm techniques with mechanical and acoustic profiles for the prediction of apples' sensory attributes," *Chemometrics Intell. Lab. Syst.*, vol. 253, Oct. 2024, doi: 10.1016/j.chemolab.2024.105217.
8. S. Byun, "Design of an integrated near-infrared spectroscopy module for sugar content estimation of apples," *Micromachines*, vol. 13, no. 4, Apr. 2022, doi: 10.3390/mi13040519.
9. X. Ma, H. Luo, F. Zhang, and F. Gao, "Study on the influence of region of interest on the detection of total sugar content in apple using hyperspectral imaging technology," *Food Sci. Technol. (Brazil)*, vol. 42, 2022, doi: 10.1590/FST.87922.
10. P. Drkenda, A. Čulah, N. Spaho, A. Akagić, and M. Hudina, "How do consumers perceive sensory attributes of apple?," *Foods*, vol. 10, no. 11, Nov. 2021, doi: 10.3390/foods10112667.
11. Z. Fathizadeh, M. Aboonajmi, and S. R. Hassan-Beygi, "Nondestructive methods for determining the firmness of apple fruit flesh," *Information Processing in Agriculture*, Dec. 2021, doi: 10.1016/j.inpa.2020.12.002.
12. Y. Cao et al., "Simple and effective characterization of Fuji apple flavor quality by ethylene and sugar content," *Food Anal. Methods*, vol. 14, no. 12, pp. 2576–2584, Dec. 2021, doi: 10.1007/s12161-021-02085-2.
13. A. U. Rehman et al., "A machine learning-based framework for accurate and early diagnosis of liver diseases: A comprehensive study on feature selection, data imbalance, and algorithmic performance," *Int. J. Intell. Syst.*, vol. 2024, no. 1, Jan. 2024, doi: 10.1155/2024/6111312.
14. T. M. Ali et al., "A sequential machine learning-cum-attention mechanism for effective segmentation of brain tumor," *Front. Oncol.*, vol. 12, Jun. 2022, doi: 10.3389/fonc.2022.873268.
15. Almulhim, M., Islam, N., & Zaman, N. (2019). A lightweight and secure authentication scheme for IoT based e-health applications. *International Journal of Computer Science and Network Security*, 19(1), 107-120.
16. Zaman, N., Low, T. J., & Alghamdi, T. (2014, February). Energy efficient routing protocol for wireless sensor network. In *16th international conference on advanced communication technology* (pp. 808-814). IEEE.
17. Azeem, M., Ullah, A., Ashraf, H., Jhanjhi, N. Z., Humayun, M., Aljahdali, S., & Tabbakh, T. A. (2021). Fog-oriented secure and lightweight data aggregation in iomt. *IEEE Access*, 9, 111072-111082.
18. Ahmed, Q. W., Garg, S., Rai, A., Ramachandran, M., Jhanjhi, N. Z., Masud, M., & Baz, M. (2022). Ai-based resource allocation techniques in wireless sensor internet of things networks in energy efficiency with data optimization. *Electronics*, 11(13), 2071.

19. Khan, N. A., Jhanjhi, N. Z., Brohi, S. N., Almazroi, A. A., & Almazroi, A. A. (2022). A secure communication protocol for unmanned aerial vehicles. *CMC-Computers Materials & Continua*, 70(1), 601-618.
20. Muzafar, S., & Jhanjhi, N. Z. (2020). Success stories of ICT implementation in Saudi Arabia. In *Employing Recent Technologies for Improved Digital Governance* (pp. 151-163). IGI Global Scientific Publishing.
21. Jabeen, T., Jabeen, I., Ashraf, H., Jhanjhi, N. Z., Yassine, A., & Hossain, M. S. (2023). An intelligent healthcare system using IoT in wireless sensor network. *Sensors*, 23(11), 5055.
22. Shah, I. A., Jhanjhi, N. Z., & Laraib, A. (2023). Cybersecurity and blockchain usage in contemporary business. In *Handbook of Research on Cybersecurity Issues and Challenges for Business and FinTech Applications* (pp. 49-64). IGI Global.
23. Hanif, M., Ashraf, H., Jalil, Z., Jhanjhi, N. Z., Humayun, M., Saeed, S., & Almuhaideb, A. M. (2022). AI-based wormhole attack detection techniques in wireless sensor networks. *Electronics*, 11(15), 2324.
24. Shah, I. A., Jhanjhi, N. Z., Amsaad, F., & Razaque, A. (2022). The role of cutting-edge technologies in industry 4.0. In *Cyber Security Applications for Industry 4.0* (pp. 97-109). Chapman and Hall/CRC.
25. Humayun, M., Almufareh, M. F., & Jhanjhi, N. Z. (2022). Autonomous traffic system for emergency vehicles. *Electronics*, 11(4), 510.
26. Muzammal, S. M., Murugesan, R. K., Jhanjhi, N. Z., & Jung, L. T. (2020, October). SMTrust: Proposing trust-based secure routing protocol for RPL attacks for IoT applications. In *2020 International Conference on Computational Intelligence (ICCI)* (pp. 305-310). IEEE.
27. Sharma, A., Das, P., Barman, S., Sharma, S. K., Ahmed, B., & Adhikary, T. (2025). A Comparative Study on the Predictive Ability of Machine Learning-and Deep Learning-Based Yield Prediction Model in Horticulture: A Case Study of Apple. *Applied Fruit Science*, 67(4), 191.
28. Han, B., Zhang, J., Almodfer, R., Wang, Y., Sun, W., Bai, T., ... & Hou, W. (2025). Research on Innovative Apple Grading Technology Driven by Intelligent Vision and Machine Learning. *Foods*, 14(2), 258.
29. Feilong, Q., Khan, N. A., Jhanjhi, N. Z., Ashfaq, F., & Hendrawati, T. D. (2025). Improved YOLOv5 Lane Line Real Time Segmentation System Integrating Seg Head Network. *Engineering Proceedings*, 107(1), 49.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.