engineering proceedings

Proceeding Paper

A Prototype to Prevent Fruits from Spoilage: An Approach Using Sensors with ML †

Uttam Narendra Thakur and Angshuman Khan *

Department of Electronics and Communication Engineering, University of Engineering and Management, Jaipur 303807, Rajasthan, India; unthakur08@gmail.com
* Correspondence: angshumankhan2910@gmail.com

† Presented at the 10th International Electronic Conference on Sensors and Applications (ECSA-10), 15–30 November 2023; Available online: https://ecsa-10.sciforum.net/.

Abstract: One of the significant issues facing the world right now is food deterioration. If the freshness or deterioration of a fruit can be determined before it is lost, the fruit waste problem may be mitigated. The goal of this work is to develop a simple model for tracking fruits quality using sensors with machine learning (ML) approach. This model assists from the gases emitted by it in determining the fruits that will ripen and require use earlier. Two gas sensors (MQ3 & MQ7) and an Arduino Uno serve as the main processing components of the suggested system. Principal Component study (PCA) is a widely employed discriminating approach that has been utilised to differentiate between fresh and rotten apples based on sensed data. The study yields a cumulative variance of 99.1% over a span of one week. The data has also been evaluated using a linear Support vector machine (SVM) classifier, which has achieved an accuracy of 99.96%. The distinctive feature of the system is that it evaluates the levels of spoilage based on real-time data and deploys a low-cost, straightforward model that can be used anywhere to preserve any type of fruit.

Keywords: support vector machine; principal component analysis; classification; machine learning

Citation: Thakur, U.N.; Khan, A. A Prototype to Prevent Fruits from Spoilage: An Approach Using Sensors with ML. *Eng. Proc.* **2023**, *56*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s): Name

Published: 15 November 2023

1. Introduction

Food spoilage, which causes 40–50% of all losses of root crops, fruits, and vegetables each year, is one of the biggest problems the world is currently experiencing [1]. The development of original fruits offerings relies heavily on smell [2]. Thus, the primary objective of many studies is to characterize the smell or aroma to identify fruits [3]. Fruits smell differently at different stages of ripeness, such as when they are fresh and when they are not. The release mechanism of volatile organic compounds (VOCs) from fruits has to be investigated [4]. This is also influenced by elements like the climate, temperature, humidity, accumulation of carbohydrates, and fungus infections [5]. It would help in the early identification and monitoring of fruit phases such as raw, ripe, or rotten, protecting sellers and buyers. As an illustration, as a banana is ripening, the flavour chemicals that give it its typical flavour are produced quickly. The aroma's composition also changes at different stages, giving rise to diverse aroma signatures. Several studies have been conducted to investigate the chemical compositions of various volatile organic compounds (VOCs) emitted by bananas during the ripening process. However, the majority of these studies have mostly focused on the compositional changes occurring in fruits as they mature or undergo processing [6]. The fragrance composition changes during fruit ripening, which helps explain aromatic compound production [7]. Varieties of banana cultivars have distinctive flavours due to differences in their predominant volatile components.

Sensors play an important function in detecting these odours given off by the fruits at various phases of life. The sensory analysis with trained panellists precedes to prepare the sensory characteristics of food products [3,4]. An array of non- or partially selective gas sensors, a data processing and pattern recognition system that can identify even complex aromatic profiles. The sensor system responds differently to different aromas, and these responses produce a signal pattern that is unique to that aroma [6,7]. Each chemical, however, has unique attributes and traits of its own. Chemical properties are therefore unique and will be used as the input data to create this kind of system [8]. The computerised markings of explicit odours must be gathered into a library (dataset) in order for the programmed recognition process to be successful [9]. The sensor array finds specific compounds in a medium. A unique sensor is responsible for detecting each targeted aroma, or each sensor is in charge of detecting a certain sort of aroma. The primary aim of this study is to present a sensor array system integrated with a machine-learning technique for the purpose of identifying distinct stages of fruit ripening. The following is the key contribution of this work:

1. The paper recognizes food deterioration as a critical global issue and seeks to contribute to the mitigation of fruit waste by detecting freshness or spoilage before it occurs.
2. It creates a straightforward model that utilizes sensors and machine learning approach to monitor and assess the quality of fruits, providing a practical solution to this problem.
3. The paper demonstrates the successful integration of two gas sensors, MQ3 and MQ7, along with an Arduino Uno, as the primary components of the system. These sensors capture emissions from fruits, enabling the assessment of their ripeness and quality.
4. The research employs PCA to classify sensor data with a good accuracy rate over a one-week period. Additionally, a SVM classifier achieves an even higher accuracy when evaluating the collected data.

2. Literature Study

The identification of fruit odors has been the subject of numerous studies, each having advantages and disadvantages. In this section, a few pertinent recent works are discussed.

The research [10] describes the application of a model for the rapid classification of various varieties of polish honey based on FIGARO semiconductor sensors. It makes use of TGS (FIGARO USA, INC) type semiconductor sensors, PCA, linear discriminant analysis (LDA), cluster analysis (CA), and LDA algorithms. In [11], a reliable and appropriate non-targeted method is created and evaluated for the differentiation of oranges from three different geographic sources. It used principal component analysis and Non-linear Iterative Partial Least Squares (NIPALS) techniques, although sensor details are unclear. A strategy for applying an electronic nose for detecting formaldehyde-containing materials and seafood spoilage is described in [12]. In addition, comparisons and evaluations of static and dynamic properties, as well as sensor stability, are made. MOSFET sensors are combined with PCA and deterministic finite automaton (DFA) algorithms. Based on its volatile composition, Moroccan virgin olive oils are explored in [13] with a model to distinguish between distinct geographic origins. It used six metal oxide semiconductor (MOS) sensors with PCA and LDA algorithms. The bio-inspired neural network, which is based on the olfactory systems of mammals, is given in [14] as a novel data processing method. A set of TGS, MiSC, and SHT75 sensors are used along with a bio-inspired neural network algorithm. In [15], samples were taken from the Pamplona and Toledo DWTPs in the northern Santander region (Colombia) to evaluate a multisensory system. The tests are executed utilising PCA, deterministic finite automaton (DFA), probabilistic neural network (PNN), and TGS sensors. Membrane filtration (MF) was utilised to validate the detection of E. coli in water.

Various sensors, including semiconductor sensors of the FIGARO and TGS types, MOSFET sensors, and metal oxide semiconductor (MOS) sensors, among others, were used in a number of research. The literature does not always include thorough sensor specifications and features. The literature also discusses PCA, LDA, NIPALS, DFA, bio-inspired neural networks, and probabilistic neural networks for odour identification and classification. A straightforward approach to determine the best methods for individual applications is lacking.

In this paper, a straightforward ML-based sensor model for fruit quality tracking is presented. The proposed system uses two gas sensors (MQ3 & MQ7) and an Arduino Uno for processing. The experiments use PCA and an SVM classifier. The system analyses rotting levels using real-time data and uses a low-cost, simple model to preserve any fruit.

3. Proposed Model

Figure 1 illustrates the schematic representation of the proposed setup. The proposed configuration is categorically segmented into three components, namely the sensing chamber, data gathering, and processing unit. The sensing chamber is comprised of a pair of sensors and a sample test platform. The alteration in the environmental conditions within the sensing chamber was measured via sensors and then transmitted to the data acquisition system. The Arduino Uno microcontroller was employed to acquire data from the sensors and thereafter store it in its memory. Subsequently, the response underwent processing by the processing unit in order to facilitate further discrimination and classification.

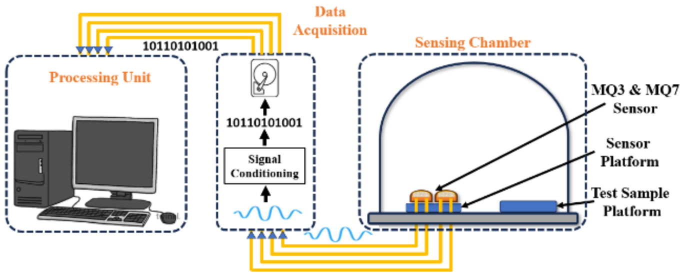


Figure 1. Schematic view of the experimental set-up.

The sensors that detect the quantities of volatile chemicals, MQ3 for alcohol and MQ7 for CO, sense the gases released from the samples. The fresh and ripened stages of fruit ripening are perceived separately from the gas concentrations. The ppm levels and associated change in sensor voltage values obtained when the sensor reacts to each of the many gases that are intended to be monitored are used to calculate the gaseous concentration. The data values are read to the 1 GB secure digital (SD) card in text format with a '.txt' extension. The influence of a change in room temperature on the degree of food rotting is not a concern for the experimental setting, hence the experiment is only conducted at room temperature. The parallel gas sensor array is installed on the breadboard. The sensor array is supplied with +5 V according to the optimal rating of the MQ series gas sensors, with the analogue pin of the MQ7 sensor wired to pin (A5) and the pin of the MQ3 gas sensor wired to the pin (A3) of the Arduino Uno board. The gas sensors are connected to the Arduino Uno board using a breadboard that offers internal connectivity. The ground pin (GND) of the Arduino Uno Board is connected to the gas sensors. The sensors were allowed to read the appropriate gas concentration in the air in order to stabilize the fluctuations of the sensor's voltage level until the sensor is heated up to a particular temperature, ensuring that the sensors detect the particular gas concentration in the aroma of the target food material in this experiment, 'banana.' The data values are transformed from text (.txt) into the comma separated format (.csv), which is the optimum format for using data science applications, to build the dataset in an excel file. Python has been chosen as the main programming language for machine learning algorithms (PCA, SVM) and the data science application. The PCA and SVM are two popular names in the ML paradigm.

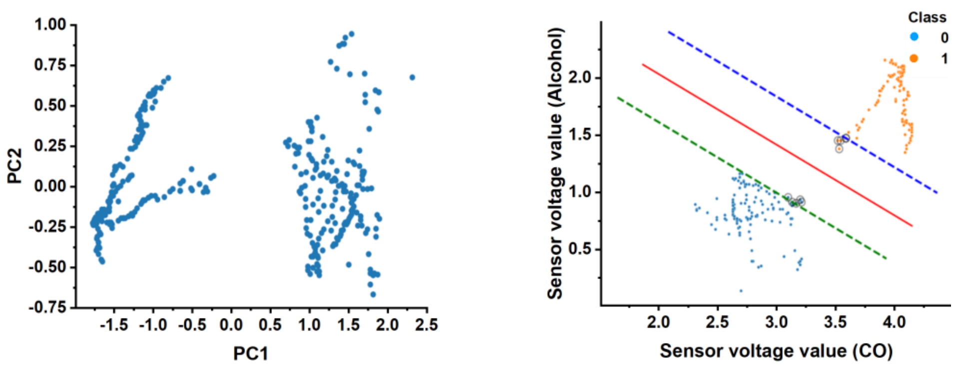
PCA is a dimensionality reduction technique widely used in machine learning and statistics. SVMs are a powerful and versatile class of supervised machine learning algorithms primarily used for classification and regression tasks. The dataset is first examined for missing or null values using the Python 'NumPy' module, and then a standard scaling is applied to the data. The PCA, a feature extractor and unsupervised learning clustering tool, is used to analyse the dataset's primary components that determine categorization between rotten and fresh bananas. The SVM classification, a supervised machine learning approach, is utilised to categorise it in the various groups.

Methodology

A labelled dataset with stipulated accuracy and correctness makes up the current experimental dataset. The specifications are mentioned in the operation of the hardware setup and the environmental circumstances under which the data-gathering technique is carried out. The recorded readings from the MQ7 for carbon monoxide (CO) and the MQ3 for alcohol have been grouped together in a single '.csv' file to represent the values seen in the dataset. The dataset includes the samples of fresh bananas and the examples of damaged bananas. Based on the dataset, it is evident that the MOS gas sensors function as intended because as the sensor voltage progressively rises, the corresponding ppm levels also increase. This is valid for both of the observed gas concentrations. Since the dataset contains huge null values, data cleaning techniques must be used before principal component analysis. The dataset's straightforward overview makes it clear that there is a contrast between the two stages of the banana's life cycle, fresh and spoilt, based on the recorded ppm levels and the matching sensor voltage values for each gas concentration. Thus, the dataset is divided into two classes, Class 1 and Class 0, with Class 1 representing the dataset's spoilt characteristics and Class 0 representing its fresh features.

4. Results and Discussion

The dataset in this experiment is split into fresh and rotten categories. The dataset's "sensor_voltage" value for CO & "sensor_voltage" value for alcohol independent variables served as the basis for the evaluation. A 91.3% percent of the scatter in the data is explained by the first principal component (PC1), while the second PC (PC2) defines the other 7.8%. The principal component analysis plot is displayed in Figure 2a, and it can be seen that there is a clear separation between the two classes of categorization in the two clusters of the eigen values of the two principal components. As a result, a well-fitting hyperplane may be plotted to categorise the these two.



(**a**) (**b**)

Figure 2. Scatter plots: (**a**) PCA analysis; (**b**) SVM analysis.

In the case of SVM, the assessment is carried out to distinguish between the "sensor_voltage" values of CO and alcohol, respectively. The visualization of these results is provided in Figure 2b. Using 'NumPy' library functions, the dataset is cleaned to eliminate any null or missing values, and it is normalized using 'StandardScaler' from the 'Scikitlearn' framework. Cross-validation, which helped the model avoid overfitting, resulted in an accuracy of 99.96%. The linear SVM model fits a hyperplane with a clear distinction between the two classes of fruit: fresh fruit belongs in class 0, whereas ruined fruit belongs in class 1.

Figure 3 displays one of the sample responses obtained from the sensors. While the reaction magnitude of the MQ7 sensor was somewhat higher than that of the MQ2 sensor, the transient responses of both MQ sensors had nearly comparable response and recovery times. The comparison between our model and earlier research in Table 1 shows that our model has a relatively good accuracy rate.

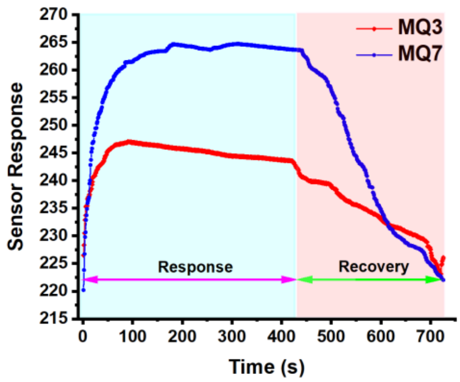


Figure 3. Sensor response of MQ3 & MQ7.

Table 1. Comparison of suggested and existing schemes.

Works	Application	Algorithm Used	No. of Sensors	Type of Sensor	Accuracy
[10]	Classification of polish honey	PCA, LDA, CA	5	TGS	96.00%
[11]	Discrimination of oranges	PCA, NIPALS (Non-linear Iterative Partial Least Squares)	NR	NR	95.70%
[12]	Spoiling detections in octopus	PCA and DFA	6	MOSFET	93.80%
[13]	Discrimination of geographical origin virgin olive oil	PCA, LDA	6	MOS	NR
[14]	Chinese liquor classification	Bio-inspired neural network	9	TGS, MiSC and SHT75	NR
[15]	Detection of the E. coli Bacteria in Drinking Water	PCA, Deterministic Finite Automaton (DFA), PNN (Probabilistic Neural Network)	16	TGS	80.00%
This work	Discrimination and identification of fresh and spoiled fruits.	PCA, SVM	2	MQ	99.96%

NR: Not reported.

5. Conclusions

The proposed system uses two gas sensors (MQ3 & MQ7) and an Arduino Uno for processing. Principal Component Analysis (PCA) is a popular method for distinguishing fresh from bad apples using sensory data. The study generates 99.1% cumulative variance in one week. A linear Support vector machine (SVM) classifier evaluated the data with 99.96% accuracy. The technology uses real-time data to assess deterioration and deploy a low-cost, simple model that can be used anywhere to preserve any fruit. Based on the obtained findings, a low-cost system with the used sensor array and pattern recognition algorithm in this investigation exhibit potential for the detection and classification of between the under and over riped fruits. Further, with the reported approach, one can distinguish between the different stages of the ripping process.

Author Contributions: Conceptualization, U.N.T.; methodology, U.N.T.; software, U.N.T. and A.K.; validation, U.N.T.; formal analysis, U.N.T.; investigation, U.N.T. and A.K.; resources, A.K.; data curation, U.N.T.; writing—original draft preparation, A.K.; writing—review and editing, A.K.; visualization, U.N.T. and A.K.; supervision, A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xue, L.; Liu, G.; Parfitt, J.; Liu, X.; Van Herpen, E.; Stenmarck, Å.; O'Connor, C.; Östergren, K.; Cheng, S. Missing Food, Missing Data? A Critical Review of Global Food Losses and Food Waste Data. *Environ. Sci. Technol*. **2017**, *51*, 6618–6633. https://doi.org/10.1021/acs.est.7b00401.
2. Harvey, J.M. Reduction of Losses in Fresh Market Fruits and Vegetables. *Annu. Rev. Phytopathol.* **1978**, *16*, 321–341. https://doi.org/10.1146/annurev.py.16.090178.001541.
3. Stitzel, S.E.; Aernecke, M.J.; Walt, D.R. Artificial Noses. *Annu. Rev. Biomed. Eng.* **2011**, *13*, 607026. https://doi.org/10.1146/annurev-bioeng-071910-124633.
4. Green, G.C.; Chan, A.D.; Goubran, R.A. Monitoring of food spoilage with electronic nose: Potential applications for smart homes. In Proceedings of the 2009 3rd International Conference on Pervasive Computing Technologies for Healthcare, London, UK, 1–3 April 2009; pp. 1–7. https://doi.org/10.1109/PCTHEALTH.2009.5291419.
5. Aung, M.M.; Chang, Y.S. Temperature management for the quality assurance of a perishable food supply chain. *Food Control.* **2014**, *40*, 198–207. https://doi.org/10.1016/j.foodcont.2013.11.016.
6. Loutfi, A.; Coradeschi, S.; Mani, G.K.; Shankar, P.; Rayappan, J.B.B. Electronic noses for food quality: A review. *J. Food Eng.* **2015**, *144*, 103–111. https://doi.org/10.1016/j.jfoodeng.2014.07.019.
7. Berna, A. Metal Oxide Sensors for Electronic Noses and Their Application to Food Analysis. *Sensors* **2010**, *10*, 3882–3910. https://doi.org/10.3390/s100403882.
8. Mohamed, R.R.; Yaacob, R.; A Mohamed, M.; Dir, T.A.T.; A Rahim, F. Food Freshness Using Electronic Nose and Its Classification Method: A Review. *Int. J. Eng. Technol.* **2018**, *7*, 49. https://doi.org/10.14419/ijet.v7i3.28.20964.
9. Sarno, R.; Wijaya, D.R. Recent development in electronic nose data processing for beef quality assessment. *Telecommun. Comput. Electron. Control* **2019**, *17*, 337–348. https://doi.org/10.12928/telkomnika.v17i1.10565.
10. Dymerski, T.; Gębicki, J.; Wardencki, W.; Namieśnik, J. Application of an electronic nose instrument to fast classification of Polish honey types. *Sensors* **2014**, *14*, 10709–10724. https://doi.org/10.3390/s140610709.
11. Centonze, V.; Lippolis, V.; Cervellieri, S.; Damascelli, A.; Casiello, G.; Pascale, M.; Logrieco, A.F.; Longobardi, F. Discrimination of geographical origin of oranges (Citrus sinensis L. Osbeck) by mass spectrometry-based electronic nose and characterization of volatile compounds. *Food Chem.* **2018**, *277*, 25–30. https://doi.org/10.1016/j.foodchem.2018.10.105.
12. Zhang, S.; Xie, C.; Bai, Z.; Hu, M.; Li, H.; Zeng, D. Spoiling and formaldehyde-containing detections in octopus with an E-nose. *Food Chem.* **2009**, *113*, 1346–1350. https://doi.org/10.1016/j.foodchem.2008.08.090.
13. Haddi, Z.; Amari, A.; Ali, A.O.; El Bari, N.; Barhoumi, H.; Maaref, A.; Jaffrezic-Renault, N.; Bouchikhi, B. Discrimination and identification of geographical origin virgin olive oil by an e-nose based on MOS sensors and pattern recognition techniques. *Procedia Eng.* **2011**, *25*, 1137–1140. https://doi.org/10.1016/j.proeng.2011.12.280.
14. Liu, Y.-J.; Zeng, M.; Meng, Q.-H. Electronic nose using a bio-inspired neural network modeled on mammalian olfactory system for Chinese liquor classification. *Rev. Sci. Instrum.* **2019**, *90*, 025001. https://doi.org/10.1063/1.5064540.
15. Carrillo-Gómez, J.; Durán-Acevedo, C.; García-Rico, R. Concentration detection of the E. coli bacteria in drinkingwater treatment plants through an E-nose and a volatiles extraction system (VES). *Water* **2019**, *11*, 774. https://doi.org/10.3390/w11040774.