Journal of ICT, 23, No. 2 (April) 2024, pp:253-292

JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY
https://e-journal.uum.edu.my/index.php/jict

How to cite this article:
Wan Azman, W. N. F. S., Ku Azir, K. N. F., & Mohd Khairuddin, A. (2024). An embedded machine learning-based spoiled leftover food detection device for multiclass classification. *Journal of Information and Communication Technology, 23*(2), 253-292. https://doi.org/10.32890/jict2024.23.2.4

An Embedded Machine Learning-Based Spoiled Leftover Food Detection Device for Multiclass Classification

[1]Wan Nur Fadhlina Syamimi Wan Azman,
[2]Ku Nurul Fazira Ku Azir & [3]Adam Mohd Khairuddin
[1,2&3]Faculty of Electronic Engineering & Technology
Universiti Malaysia Perlis, Malaysia.
[1,2&3]Centre of Excellence for Advanced Computing,
Universiti Malaysia Perlis, Malaysia.

*[1]syamimifadhlina@gmail.com
[2]fazira@unimap.edu.my
[3]adamkhairuddin@unimap.edu.my
*Corresponding author

Received: 24/3/2024 Revised:14/4/2024 Accepted: 15/4/2024 Published: 30/4/2024

ABSTRACT

Food waste's negative environmental repercussions are causing it to become a global concern. Several studies have examined the factors influencing food waste behaviour and management. This work was motivated by the lack of previous research on machine learning and electronic noses to detect contamination from leftover cooked food. This work proposes using machine learning algorithms and electronic nose technology to recognise and forecast the contamination in leftover cooked food. After five days of storage, the freshness of cooked leftovers was evaluated using an electronic nose combined with machine learning algorithms. Most food samples used in this work were from Malaysian's leftover lunch and dinner dishes. Four (4) gas sensors—MQ-2, MQ-136, MQ-137, and MQ-138—are used in developing the electronic nose to identify the presence of gas in the food sample. The data from the gas sensors was analysed using machine learning methods, namely Random Forest, k-nearest Neighbors, Support Vector Machine, and Linear Discriminant Analysis. Based on the results, a multi-classification technique yielded a greater accuracy rate in classifying and identifying the level of contamination in the cooked food leftovers, with average accuracy ranging from 90 percent to 100 percent. In conclusion, the work demonstrates a novel method for using machine learning algorithms to classify, identify, and predict the contamination level of leftover cooked food, contributing to reducing food waste generated primarily by Malaysians.

Keywords: Classification, electronic nose, food safety, food waste, machine learning.

INTRODUCTION

The socioeconomic costs, waste management, and climate change consequences of food loss and waste (FLW) make it a substantial issue (Chauhan et al., 2021). Food loss, which typically happens in the food value chain, is the term used to describe food that deteriorates or is lost before reaching customers. It frequently results from unintended farming techniques or technical shortcomings in infrastructure, packaging, marketing, storage, or other areas. Contrarily, food that is of good quality but is rejected before or beyond expiration and is not consumed is food waste. It usually happens at the retail and consuming phases of the food value chain, frequently due to carelessness or intentional disposal (Lipinski, 2013). The issue of food waste has grown significantly on a global scale, particularly in developing nations like Malaysia. It has contributed substantially to climate change's effects (Tonini et al., 2018). Some scholars have equated "food loss" with "food waste." However, those who distinguish between the two define "food loss" as food wasted at the start of the value-add chain and "food waste" as food lost at the conclusion (Betz et al., 2015).

The Food and Agriculture Organization of the United Nations estimates that 30 percent of global greenhouse gas emissions are caused by food loss. Malaysia produces 38,000 metric tonnes of waste daily, with 15,000 metric tonnes being food waste (Naim & Rahman, 2020; Sim, 2019). The issue has garnered attention from academics, researchers, politicians, international organisations, and NGOs. A public awareness campaign supported by the government has been initiated to address the increase in food waste, but most attempts have failed due to lack of public response. Past studies have also shown the intention of reducing food waste among Malaysian households (Amirudin & Gim, 2019; Wong et al., 2021). However, Malaysians still lack the information and knowledge on food waste issues. If food waste issues continue, Malaysia will overtake the United States and China as the world's third-greatest emitter of greenhouse gases (Food wastage leads to climate change-Forum Air Malaysia, 2018).

To determine whether food is edible, consumers rely on scent and visual clues (Gunders & Bloom, 2017). However, customers may have varied preferences for the dish's appearance and smell. Even if some food products are still edible, some may discard them due to different tastes. When freezing or reheating cooked leftovers in the future, they should be kept in an airtight container or tightly wrapped in freezer bags. However, people like to gather and refrigerate multiple leftovers at once, accumulating surplus food in the refrigerator (Gunders & Bloom, 2017). Customers can overlook a food item concealed beneath other items in the refrigerator drawer until it starts to mildew or smell bad and has to be thrown out.

Food waste monitoring is crucial to assess the extent of food waste issues and identify areas for action. Machine learning and electronic nose technology are increasingly used in food safety to identify and classify food contaminants accurately. Machine learning is a computer-assisted method for identifying patterns in data, while an electronic nose can identify odours by examining chemical components and identifying distinctive components. Electronic noses mimic the scent organ in humans, making them easy to use. The primary objective of this work is to detect and forecast contamination levels using a machine-learning system, accurately classifying various contamination levels while identifying rotten foods, taste, texture, and colour. Electronic noses are reliable, simple to use, and inexpensive tools used to measure volatile gases, which are used to determine food safety and quality. This new development for categorising pollutants is essential for consumers to identify spoiled food before discarding it, especially for leftover cooked food. This research work's proposed machine learning and contamination level classification method is expected to enhance food waste management in Malaysia and the environment.

RELATED WORKS

Food Waste in Malaysia

In general, food materials that are lost during the production, processing, and preparation stages of the food supply chain are categorised as follows: (1) food waste; (2) unavoidable food waste, which includes things like fruit peel and core; and (3) avoidable food waste, which refers to edible food that is lost during the consumption phase (Ong et al., 2019). In a household, food waste is the outcome of all three activities: purchasing, preparing, and eating food. When preparing food, particularly in the kitchen, individuals may remove inedible, faulty, or broken elements of the meal and edible components such as skins to acquire the appropriate nutritional characteristics. Although pets can consume a smaller quantity of food rejects and leftovers than people, the amount of food accessible for human consumption begins to diminish. A few instances of domestic food waste include food that is grown on-site, food that is taken away, and food that is sold at retail establishments.

Domestic waste is Malaysia's primary municipal solid waste (MSW) source. 44.5 percent of total solid waste collection figures, or 6.1 million tonnes yearly, are attributed to the residential sector, according to the Khazanah Research Institute (KRI). According to Ismail et al. (2020), food waste is half of MSW's total waste composition, separated into 20 categories. In contrast to South-

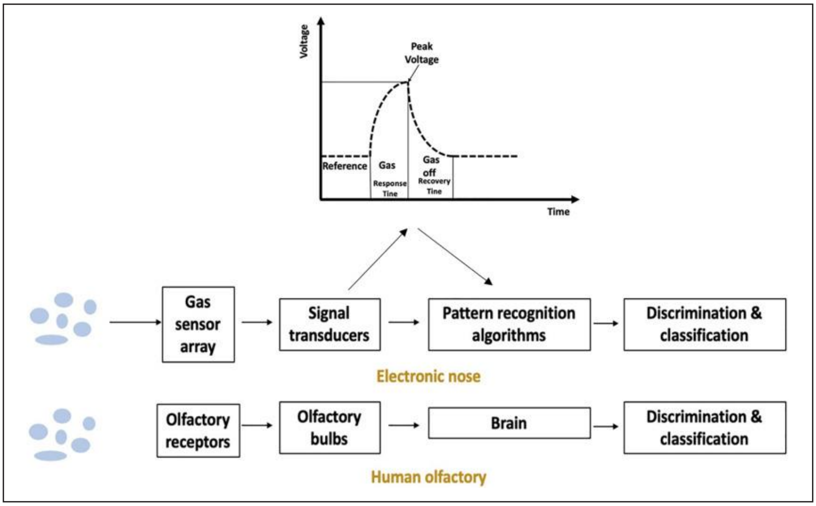
Eastern (40%) and Central European urban areas, where food waste makes up approximately 24 percent of municipal solid waste, food waste makes up the majority of household garbage in Malaysia (Zulkifli et al., 2019). Certain ingredients in food break down and release smells when refrigerated because of the interaction between bacteria and enzymes. The same holds for each household's dustbin, gradually filling up. The stink it released forced the customers to hold their breath at one point. Among the substances that contributed most to the garbage's odour were nitrogen (ammonia) and sulphur (hydrogen sulphide). On the other hand, the waste stench contained other substances. Proteins would degrade into spoilage amines and subsequently degrade into ammonia, hydrogen sulphide, and ethyl mercaptan, according to Huang et al. (2014). Fatty acids would be produced by the breakdown of fat and subsequently transformed into aldehydes. When carbs break down, alcohols, ketones, and aldehydes are produced (Huang et al., 2014).

Electronic Nose (E-Nose)

Many scholarly articles addressed the use of machine learning in electronic nose technology to discern between various food or beverage characteristics. Conventional techniques, including gas and liquid chromatography, microbiological cell counting, mass spectrometry, and high-performance liquid chromatography (HPLC), are labour- and time-intensive since they require lengthy sample preparation procedures (Qiu & Wang, 2017). The food industry uses e-nose technology due to its affordability, ease of usage, and close relationship with sensory panels. Electronic noses are currently regarded as potential instruments for early and quick detection of contamination and defects in various food safety assessment areas. By studying its distinct components and analysing its chemical composition, an e-nose can identify an odour (Gębicki & Szulczyński, 2018; Roy & Yadav, 2022). A group of electronic gas sensors called an electronic nose, or "E-nose," can identify volatile substances in food product samples' headspaces with exceptional selectivity and sensitivity. The sense of smell is similar to that of humans (Abu-Khalaf, 2021; Gu et al., 2019; Liu et al., 2019). The relationship between an electronic nose and a human nose is shown in Figure 1.

Figure 1

Sensing-interpreting-discriminating Process of an Electronic Nose (Tan & Xu, 2020).



To mimic a real olfactory system, gas sensors and sensing materials were used in place of smell receptor cells, as Figure 1 illustrates. The brain and neural network are replaced by computational algorithms, artificial neural networks, and data analysis software (Hsieh & Yao, 2018). During the detection process, several volatile compounds interact with the sensor array, and each sensor stores data as a distinct pattern, or "fingerprint" (Jiang & Liu, 2020; Oates et al., 2018, 2020). The signals are sent to a computer system that uses multivariate data analysis (MVDA) techniques to distinguish and identify the fingerprinting of measured samples (Hsieh & Yao, 2018). E-nose is an objective, automated, non-destructive technology that has grown in popularity and significance as a detection system in a number of industries because of its low cost, high sensitivity, and straightforward design (Liu et al., 2019).

Many studies demonstrated the widespread use of e-nose in assessing food quality-related properties, including the detection of Salmonella (Gonçalves et al., 2023), the quality status of fruits (Buratti et al., 2018; Qiu et al., 2014, 2015), the quality of oils (Hosseini et al., 2023), and the detection of contaminated foods (Feyzioglu & Taspinar, 2023; Putri et al., 2023; Tian et al., 2023).

However, this work distinguished itself from similar past works by incorporating machine learning and an electronic nose to assess the quality of leftover food after days of storage. A related work was recently conducted by Wan Azman et al. (2022), but they used a different classification technique—binary classification—and produced different results. In order to assist consumers in determining which food to discard, it is imperative that the electronic nose reliably detect tainted food. The proposed machine learning approach and method for categorising contamination levels in this work were thought to contribute to developing a unique framework for enhancing environmental preservation and food waste management in Malaysia.

Machine Learning Algorithms

Multivariate statistics is a branch of statistics, while machine learning is a branch of computer science and artificial intelligence. Despite coming from different fields, the two ideas often overlap because they can analyse large, high-dimensional datasets. For example, multivariate statistics and machine learning concentrate on the underlying interactions between components, while multivariate statistics focuses on the algorithms and their predictions. Consequently, there is considerable interest in both methods of assessing the data generated by the electronic nose (Pampoukis et al., 2022). Among the most popular machine learning algorithms are the following ones: A number of them will be used in this work, including Support Vector Machines (SVM), k-nearest Neighbours (k-NN), decision tree-based algorithms (like Random Forest), and artificial neural networks (ANN).

This work will utilise four popular machine-learning techniques: Random Forest (RF), k-NN, SVM, and LDA. Various applications of SVM have been explored in previous studies, including face recognition, chemical classification, text classification, bioinformatics, and data mining (Chauhan et al., 2019). The SVM technique uses a kernel function to transform the original data points from their input space into a higher dimensional feature space (Tan & Xu, 2020). Several commonly used kernel functions include sigmoid, linear, nonlinear, polynomial, Gaussian kernel, and Radial Basis Function (RBF) (Papadopoulou et al., 2013). As explained in a study by Mohamed (2017), the SVM algorithm utilises the concept of margin to effectively determine the optimal distance between the hyperplanes in classification problems. Each feature is assigned a coordinate value, and every data item is plotted as a point in an n-dimensional space, where n corresponds to the number of features.

Identifying the hyper-plane that effectively separates the two classes is crucial in performing performance classification (Kefi-Fatteh et al., 2019). Discovering a linear feature combination that can effectively distinguish between multiple objects is the objective of linear discriminant analysis (LDA), a statistical method utilised in machine learning and pattern recognition. A study by Mahmodi et al. (2019) found that LDA shares similarities with variance and regression analysis. These methods represent the dependent variable as a linear combination of the other independent variables. The LDA technique assumes that each class follows a normal distribution with similar dispersion. This work aims to map samples from N-dimensional space onto a line. Regarding problems related to K classes, m = min (K-1, N) lines were needed. Thus, the samples can be mapped using a set of linear functions. As per the study by Jiarpinijnun et al. (2020), the LDA technique has improved inter-group discrimination by reducing intra-group variance and optimising inter-group variation.

The k-NN method calculates object distances during training using digitalised feature set data. Nearest neighbours offer the shortest distances. The distance between elements is measured using Minkowski, Mahalanobis, or Euclidean methods. According to Xu et al. (2019), the k-NN classifier relies significantly on parameter k for model identification accuracy. To improve the k-NN classifier's performance, modify k for further analysis (Chen et al., 2011). An RF uses numerous trained decision trees to "vote" for a sample's class. The majority vote selects the sample (Schroeder et al., 2019). Due to the vast number of decision trees, selecting qualities for training can sometimes include selecting the one with the largest information gain; instead, it can be done randomly from all of them. There are many radiofrequency training methods. For decision tree training, m random samples with replacement (bootstrap RF) are selected from the training sample pool. Alternative training involves retaining weights above the original training set S and changing them based on classification success. Elith et al. (2008) found that improperly classified cases have higher weights.

Imbalance Classification

Datasets with disparities in the dependent variable are said to be unbalanced. This binary classification issue frequently occurs, leading to the classifier's output being biased in favour of the majority class (Bejjanki et al., 2020). The class imbalance issue has been demonstrated to affect a wide range of real-world applications, including face recognition, text mining, software defect prediction, and remote sensing (Feng et al., 2018). An unbalanced dataset's uneven value distribution across class label attributes can lead to an erroneous classification approach (Bejjanki et al., 2020). There were several methods for balancing the imbalanced data. The unbalanced datasets in this work are balanced via oversampling and undersampling. The data is analysed using the R Studio programme. According to Mohammed et al. (2020), oversampling can be achieved by producing more minority-class instances or samples and repeating some examples. Information loss is prevented via oversampling. Nevertheless, data replication causes problems for it. On the other hand, undersampling lowers the number of majority target instances or samples. Some majority class samples are reduced in the undersampling method to balance the data, which results in a loss of information (Bindra & Sood, 2019).

METHODOLOGIES

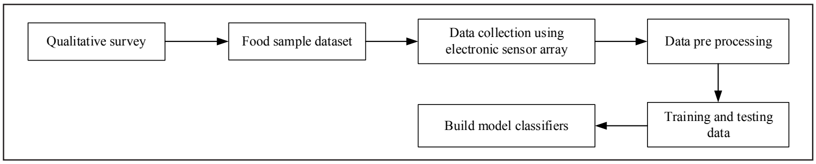
The Proposed Framework

A questionnaire was initially created for this work to comprehend the variations in food waste levels throughout houses and to evaluate household food waste in large-scale studies. However, the survey's metrics depend on respondents' awareness of their food waste levels and how the survey's findings can be used to identify the kinds of discarded food and forecast household behaviour related to food loss and waste management. The questionnaire aims to understand better the factors that influence consumer food waste and the underlying causes of behaviours and practices related to food waste. A questionnaire was also created to compile the list of food samples that would be utilised during the data collection phase. This work's sensing and classification units are two more significant and connected topics. The work's sensing unit comprises the odour sample, electronic sensor array, and pre-processing odour signal circuit. The classification unit comprises the classifier, training data, and data pre-processing.

The basis for the research work described in this paper is shown in Figure 2. The food sample used in this study is cooked leftover food with a bad odour, usually caused by food expiry or decomposition. The object is stored in a cover-mounted, hermetically sealed container with sensors. The presence of gases is usually indicated by the food sample emitting an unpleasant odour. The electronic sensor array will detect the gas emitted by the food sample. Sensor values, or data, will be produced by the sensors. The data pre-processing stage recorded the obtained values from the sensor. The classifier model is then built using the dataset. Lastly, the classifier model was built after training and testing the datasets.

Figure 2

The Framework of the Proposed Work



Preparing Food Sample

The experiment aims to determine the indicator of contamination by assessing the odour produced from six (6) food samples consisting of leftover cooked meals often encountered in Malaysian households. Among the dishes that can be found on the menu are the following: *Ayam Goreng Berempah* (Spicy Fried Chicken), *Ayam Masak Lemak* (Chicken Cooked with Coconut Milk), *Ayam Masak Kunyit* (Turmeric Chicken), *Nasi Goreng Daging* (Fried Rice with Beef), *Sup Daging* (Beef Soup) and *Sawi Masak Air* (Vegetable Soup). These food samples were selected to determine how different cooking methods influenced the degradation rate. In addition, they were chosen to investigate the impact of various cooking methodologies on the classification outcomes. Table 1 shows how each of the food samples were cooked.

Table 1

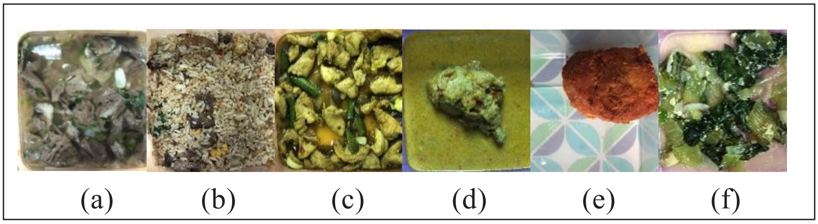
Methods of Cooking for Each Food Sample

Food Sample	Methods of Cooking
Ayam Goreng Berempah (Spicy Fried Chicken)	Deep fried
Ayam Masak Lemak (Chicken Cooked with Coconut Milk)	Cooked with coconut milk as gravy
Ayam Masak Kunyit (Turmeric Chicken)	Stir fried
Nasi Goreng Daging (Fried Rice with Beef)	Fried with beef strips
Sup Daging (Beef Soup)	Soup
Sawi Masak Air (Vegetable Soup)	Soup

For five days, the food samples were kept in a plastic container that was both sterile and airtight, and it had sensors attached to its lid. On a daily basis, work was completed from 8:30 in the morning until 5:30 in the evening. There will be a fifteen-minute interval in which the lid of the container will be removed for sensor cleaning. The experiment was conducted when the temperature was between 28 and 30 degrees Celsius. Food items in their fresh form are depicted in Figures 3(a) through 3(f). The dietary samples were observed every day for the next five days to identify any apparent shifts.

Figure 3

Food Samples in Fresh Condition: (a) Sup Daging, (b) Nasi Goreng Daging, (c) Ayam Masak Kunyit, (d) Ayam Masak Lemak, (e) Ayam Goreng Berempah and (f) Sawi Masak Air



Meanwhile, Figure 4 and Figure 5 below show the sensor values recorded automatically into Microsoft Excel and the sample of sensor readings after they were converted from analogue values to voltage values.

Figure 4

Analogue Sensor Reading Sample

AML_CombinedV3 | MQ137_Ayam Masak Lemak

	A	B	C	D
1	Date	Time	Millis	RandomValue
2	11/8/2020	3:52:09 PM	15308	947
3	11/8/2020	3:52:09 PM	30309	494
4	11/8/2020	3:52:12 PM	45309	652
5	11/8/2020	3:52:27 PM	60310	324
6	11/8/2020	3:52:42 PM	75311	593
7	11/8/2020	3:52:57 PM	90311	758
8	11/8/2020	3:53:12 PM	105312	405
9	11/8/2020	3:53:27 PM	120312	465
10	11/8/2020	3:53:42 PM	135313	552
11	11/8/2020	3:53:58 PM	150313	324
12	11/8/2020	3:54:13 PM	165314	698
13	11/8/2020	3:54:28 PM	180315	392
14	11/8/2020	3:54:43 PM	195315	706
15	11/8/2020	3:54:58 PM	210316	578
16	11/8/2020	3:55:13 PM	225316	57
17	11/8/2020	3:55:28 PM	240317	506
18	11/8/2020	3:55:43 PM	255318	101
19	11/8/2020	3:55:58 PM	270318	590
20	11/8/2020	3:56:13 PM	285319	509
21	11/8/2020	3:56:29 PM	300319	847
22	11/8/2020	3:56:44 PM	315320	888

Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | Day 6 | Day 7 | Sheet1

Figure 5

Voltage Reading and Sensory Observation

AML_HourlyMQ137 (in V)

	A	B	C	D	E	F	G	H	I
1	Hour 0	Hour 1	Hour 2	Hour 3	Hour 4	Hour 5	Hour 6	Hour 7	Level
2	4.6	0.6	3.5	2.3	4.1	0.6	3.7	1.1	Not Spoiled
3	2.4	2.9	1.1	3.7	2.8	2.8	2.6	3.0	Not Spoiled
4	3.2	2.5	4.1	0.0	0.6	4.9	0.6	2.9	Not Spoiled
5	1.6	4.2	4.3	1.9	2.4	3.4	1.0	4.1	Not Spoiled
6	2.9	3.9	1.3	1.4	5.0	1.8	1.7	2.0	Not Spoiled
7	3.7	4.8	4.9	0.4	4.0	0.2	4.1	2.3	Not Spoiled
8	2.0	1.4	1.4	4.2	2.4	4.2	0.0	3.9	Not Spoiled
9	2.3	3.4	2.2	2.7	4.6	0.8	1.4	0.8	Not Spoiled
10	2.7	1.8	0.7	0.4	1.7	4.5	0.9	0.2	Not Spoiled
11	1.6	1.4	1.4	1.9	1.2	4.6	2.6	0.0	Not Spoiled
12	3.4	3.0	3.9	4.0	4.7	3.6	2.3	3.7	Not Spoiled
13	1.9	1.9	3.0	1.1	0.4	3.8	1.9	2.0	Not Spoiled
14	3.5	2.2	4.7	3.4	2.9	1.3	3.9	3.3	Not Spoiled
15	2.8	1.4	0.6	2.3	2.3	2.9	1.7	3.6	Not Spoiled
16	0.3	3.0	4.9	5.0	3.6	3.6	1.3	3.8	Not Spoiled
17	2.5	2.2	0.1	3.3	3.1	3.4	0.7	1.8	Not Spoiled
18	0.5	4.9	4.5	1.5	4.9	2.6	0.3	3.5	Not Spoiled
19	2.9	2.0	1.8	3.0	0.1	0.0	5.0	0.9	Not Spoiled
20	2.5	1.9	0.1	2.2	2.4	0.2	1.6	3.8	Not Spoiled

As seen in Figure 5 above, there were no uniform changes in the sensor reading values; hence, the contamination categories were chosen using the observation done during the five days of the experiment.

Setting Up the Experiment

In order to identify the gas emissions that are coming from the sample, four (4) gas sensors were employed. In this work, Data transmitted from the sensor to the computer is processed and transformed with the assistance of a microcontroller called Arduino and a data-collecting tool called PLX-DAQ. The following table, Table 2, provides a list of the sensors utilised in this experiment, as well as the types of gases that the sensors could recognise.

Table 2

Types of Sensors Used in this Work

Sensor Name	Type of Gas Detected	Type of Gas Detected in This Work
MQ-2	LPG, Propane, Hydrogen, Methane, Smoke	Methane
MQ-136	Hydrogen Sulphide	Hydrogen Sulphide
MQ-137	Ammonia	Ammonia
MQ-138	n-Hexane, Benzene, NH3, Alcohol, Smoke, CO	NH3, Alcohol

Methods of Data Analysis

In the current work, SVM, k-NN, RF, and LDA were utilised to analyse and classify the data acquired from gas sensors, as well as to assess the level of food contamination. The inaccurate performance of the classification algorithm and the unequal distribution of class label features may have been caused by the fact that the datasets that were used in this investigation had dependent variable counts that were not balanced. By employing oversampling and undersampling strategies, the unbalanced datasets were brought into equilibrium to resolve this problem. The practice of producing more samples or examples of minority groups and then using some of those samples or examples again is referred to as oversampling. On the other hand, undersampling is the phrase that is used to represent a decrease in the number of samples or occurrences that are the majority of the target population. R Studio was the data analysis application used most frequently for this investigation at the time.

The purpose of this work was to do multi-classification, and it did so by analysing 17 features and 300 rows of sensor findings. As the reference group for the multi-classification, the categories “Fresh,” “Semi-Fresh,” and “Spoiled” were utilised. These designations were derived from observations of the food samples made during the experiment, which lasted for five days. The food sample was classified as “Fresh” on Days 1 and 2, “Semi-Fresh” on Days 3 and 4, and “Spoiled” on Days 4 and 5. Following loading the datasets into the programme, an 80:20 data partitioning technique was carried out. This procedure specified that 80 percent of the datasets would be defined as training sets, while the remaining 20 percent would be marked as test sets. In the course of this procedure, every data point from the relevant classes was chosen at random. A control function for multi-classification was also built using k-cross validation, with a value of k equal to 10. This was done in accordance with previous research.

RESULTS AND ANALYSIS

The results obtained in this work will be discussed in this section. The observation made during the five days of the experiment will include the confusion matrices used in this section to explain the classification test results. The different rows for each class represent the prediction, and the columns display the actual number of samples for each class. Three (3) confusion matrices, one for each of the imbalanced, oversampled, and undersampled datasets, will be present in each food sample, reflecting the classification tasks. The k-NN, SVM, RF, and LDA classification findings will also be separated into four parts of the classification result.

Freshness Sensory Observation Results

The spoilage level was observed for five days in a row for each food sample. Based on the observation, most of the food was fresh on the first day of the experiment, semi-fresh on the second day and spoiled starting from the third day and for the other two remaining days. Table 3 shows the freshness observation results for five days of the experiment.

Table 3

Freshness Sensory Observation Results

Food Sample	Sup Daging (SD)	Nasi Goreng Daging (NGD)	Ayam Masak Kunyit (AMK)	Ayam Masak Lemak (AML)	Ayam Goreng Berempah (AGB)	Sawi Masak Air (SMA)	Observation
Day 1							Fresh
Day 2							Semi-fresh

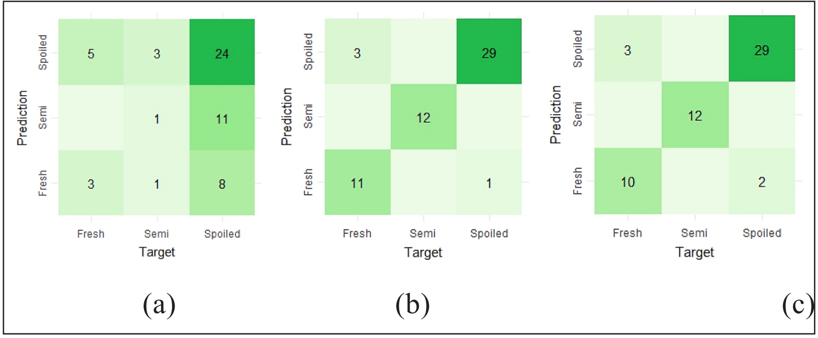
Food Sample	Sup Daging (SD)	Nasi Goreng Daging (NGD)	Ayam Masak Kunyit (AMK)	Ayam Masak Lemak (AML)	Ayam Goreng Berempah (AGB)	Sawi Masak Air (SMA)	Observation
Day 3							Spoiled
Day 4							Spoiled
Day 5							Spoiled

Classification Results of k-NN

Figure 6 illustrates the confusion matrices for the *Ayam Goreng Berempah* sample. When the dataset for *Ayam Goreng Berempah* was not balanced, the first confusion matrix indicates that k-Nearest Neighbours could not accurately classify food contamination levels.

Figure 6

k-NN Classification of Ayam Goreng Berempah: (a) Unbalanced, (b) Oversampled, and (c) Undersampled



It resulted in an average accuracy rate of 50 percent for this dataset. On the other hand, when the dataset was balanced by employing the oversampling and undersampling procedures, the average accuracy increased to 92.86 percent and 91.07 percent, respectively. Figure 7 presents the confusion matrices for *Ayam Masak Kunyit* for a visual representation. As indicated by the confusion matrix of the unbalanced dataset, the k-Nearest Neighbour algorithm could not accurately classify the various degrees of food contamination. As a consequence of this, the average accuracy of the dataset was 42.86 percent. However, after the dataset was balanced using the oversampling and undersampling procedures, the average accuracy increased to 92.86 percent and 91.07 percent, respectively. It means that the accuracy was significantly higher.

Figure 7

k-NN Classification of Ayam Masak Kunyit: (a) Unbalanced, (b) Oversampled, and (c) Undersampled

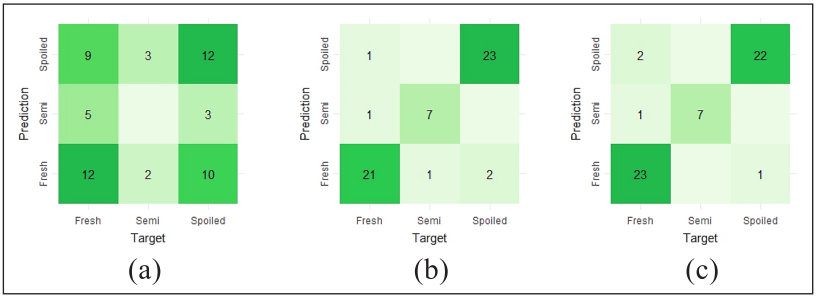
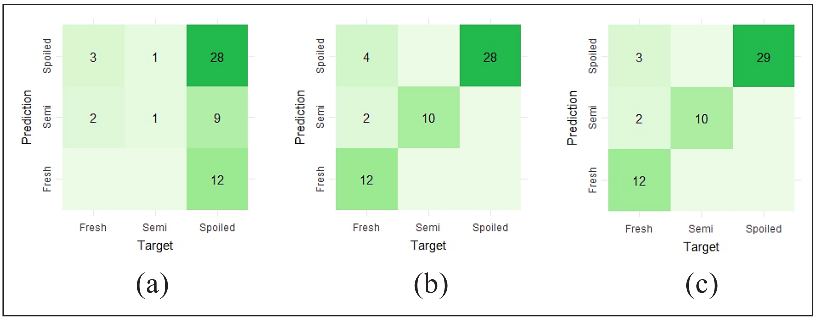


Figure 8 presents the confusion matrix for the classification tasks carried out on *Ayam Masak Lemak*. When the datasets were not in equilibrium, the average accuracy of the dataset was 51.8 percent. Using strategies such as oversampling and undersampling to achieve a balanced dataset increased the average accuracy to 89.29 percent and 91.07 percent, respectively. In the end, only two samples from the "Semi-Fresh" class and four samples from the "Spoiled" class were incorrectly labelled as "Fresh" due to the practice of oversampling. In the datasets that were undersampled, there were two samples from the "Semi-Fresh" class and three samples from the "Spoiled" class that were incorrectly classed as "Fresh".

Figure 8

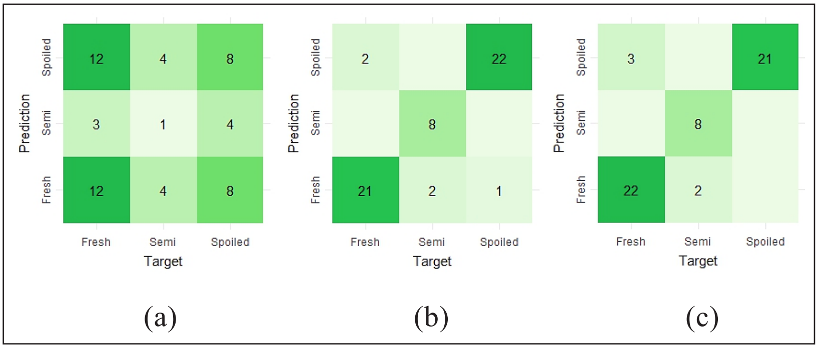
k-NN Classification of Ayam Masak Lemak: (a) Unbalanced, (b) Oversampled, and (c) Undersampled.



According to Figure 9, which can be seen below, the confusion matrix for the *Nasi Goreng Daging* classification tasks is displayed. The datasets had an average accuracy of 37.5 percent when they were not in balance with one another. Using oversampling and undersampling to achieve a balanced dataset improved the average classification accuracy to 91.07 percent.

Figure 9

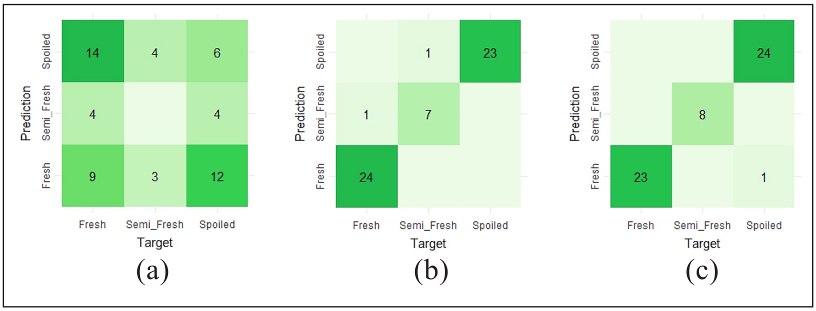
k-NN Classification of Nasi Goreng Daging: (a) Unbalanced, (b) Oversampled, and (c) Undersampled



While two samples from the "Fresh" class were incorrectly recorded as "Semi-Fresh" and one sample as "Spoiled," two samples from the "Spoiled" class were incorrectly reported as "Fresh." This was due to the fact that oversampling occurred. In comparison, the fact that only three samples from the "Spoiled" class and two samples from the "Fresh" class were incorrectly classified as "Semi-Fresh" was the consequence of undersampling. The confusion matrices for the dataset *Sup Daging* are displayed in Figure 10, which may be seen below. The first confusion matrix demonstrates that k-Nearest Neighbours was unable to appropriately classify the degrees of food contamination when the dataset for *Sup Daging* was uneven. Because of this, the average accuracy of the dataset was 26.78 percent.

Figure 10

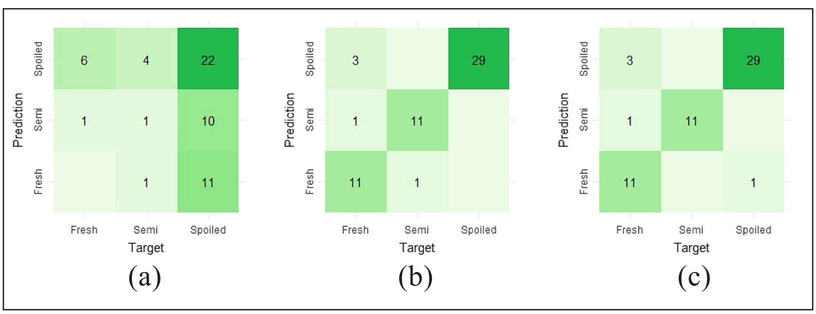
k-NN Classification of Sup Daging: (a) Unbalanced, (b) Oversampled, and (c) Undersampled



The oversampling and undersampling processes were used to achieve a balanced dataset, which increased the average accuracy to 96.43 percent and 98.21 percent, respectively. On the other hand, the confusion matrix for the *Sawi Masak Air* classification issue is depicted in Figure 11, which may be seen below. The first image makes it very evident that the k-NN approach could not be used to classify the data into the appropriate categories appropriately. It can be concluded that the average accuracy for this dataset was 41.07 percent.

Figure 11

k-NN Classification of Sawi Masak Air: (a) Unbalanced, (b) Oversampled, and (c) Undersampled.



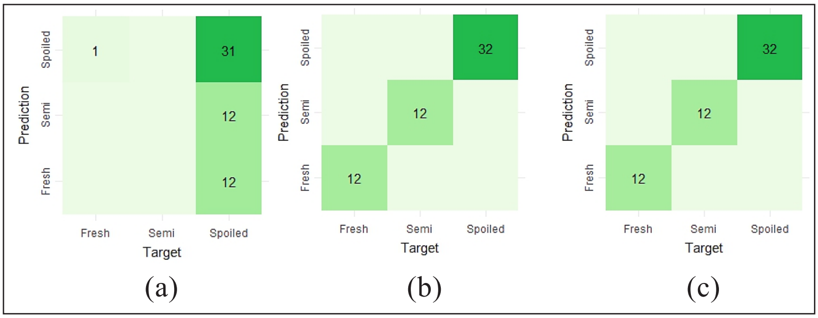
The classification accuracy increased to 91.07 percent when the dataset used the oversampling method to provide a more balanced representation. Following the completion of the balanced dataset, the undersampling procedure resulted in an improvement in classification accuracy to 91.07 percent. One sample from the "Semi-Fresh" class and three samples from the "Spoiled" class were incorrectly categorised as "Fresh" out of the samples that were oversampled. Additionally, a sample that was supposed to be in the "Fresh" category was labelled as "Semi-Fresh" by mistake. When it comes to the datasets that had undersampling, there were three samples from the "Spoiled" class and one sample from the "Semi-Fresh" class that were incorrectly classed as "Fresh." The "Spoiled" label was incorrectly applied to one of the samples that belonged to the "Fresh" category.

Classification Results of SVM

The results of the classification of the SVM method will be discussed in this section. Ayam Goreng Berempah's confusion matrices are depicted in Figure 12, which may be seen below. For example, when the dataset for *Ayam Goreng Berempah* was biassed, just one sample was incorrectly labelled as "Fresh" from the "Spoiled" class. The first confusion matrix demonstrates this. On the other hand, the data points that belonged to the "Fresh" and "Semi-Fresh" classifications were categorised in an entirely erroneous manner.

Figure 12

SVM Classification of Ayam Goreng Berempah: (a) Unbalanced, (b) Oversampled, and (c) Undersampled



Therefore, the accuracy of this dataset was, on average, 55.36 percent. Despite this, the average accuracy increased to one hundred per cent when the dataset was balanced by employing the procedures of oversampling and undersampling. The confusion matrices for the *Ayam Masak Kunyit* are shown in Figure 13, which may be found below. The confusion matrix for the unbalanced dataset suggested that the SVM could not consistently identify food contamination levels. It was similar to the situation with the other datasets stated earlier. It helps explain why this dataset's average accuracy was 33.93 percent. However, when the dataset was balanced using the oversampling and undersampling methodologies, the algorithm attained an average accuracy of 100 percent.

Figure 13

SVM Classification of Ayam Masak Kunyit: (a) Unbalanced, (b) Oversampled, and (c) Undersampled

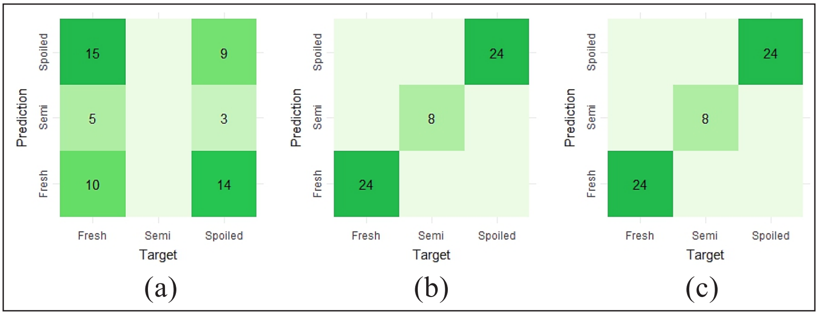
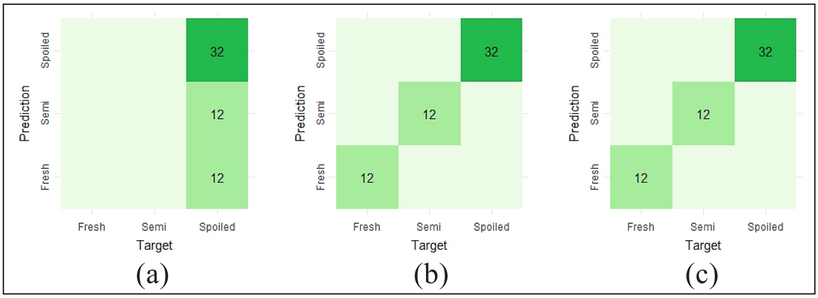


Figure 14 displays the confusion matrix for the classification tasks finished on Ayam Masak Lemak. In situations when the datasets were not in equilibrium, the average accuracy of this dataset was 57.14 percent. On the other hand, when the dataset was balanced by employing the oversampling and undersampling processes, there were no misclassifications, and the classification accuracy was 100 percent.

Figure 14

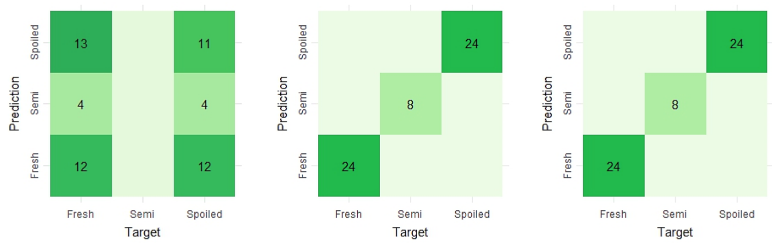
SVM Classification of Ayam Masak Lemak: (a) Unbalanced, (b) Oversampled, and (c) Undersampled



The classification tasks for the *Nasi Goreng Daging* datasets are depicted in the confusion matrix given below in Figure 15. When the datasets were deemed to be unbalanced, the average accuracy for this dataset was found to be 41.07 percent.

Figure 15

SVM Classification of Nasi Goreng Daging: (a) Unbalanced, (b) Oversampled, and (c) Undersampled



(a) (b) (c)

An average classification accuracy of 100 percent was achieved for this dataset collection by balancing the dataset using oversampling and undersampling procedures. The confusion matrices for the dataset *Sup Daging* are displayed in Figure 16, which may be seen below. The first confusion matrix demonstrates that the Support Vector Machine (SVM) could not appropriately categorise food contamination levels when the *Sup Daging* dataset had uneven data. As a consequence of this, the average accuracy for this dataset was 39.29 percent. Nevertheless, no misclassification errors occurred after the dataset was balanced by utilising the oversampling and undersampling processes. It is because the algorithm accurately categorised the dataset into the appropriate categories, achieving a classification accuracy of 100 percent.

Figure 16

SVM Classification of Sup Daging: (a) Unbalanced, (b) Oversampled, and (c) Undersampled

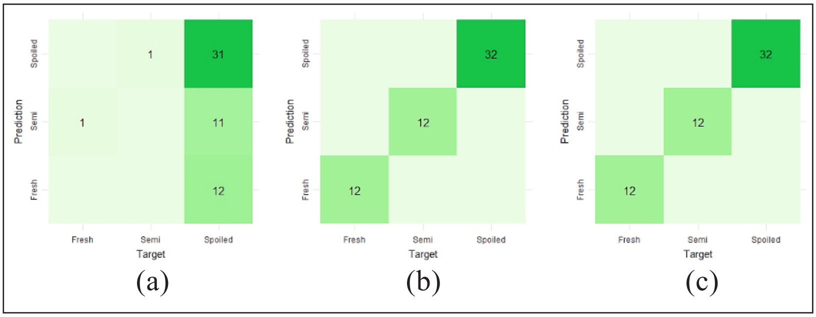
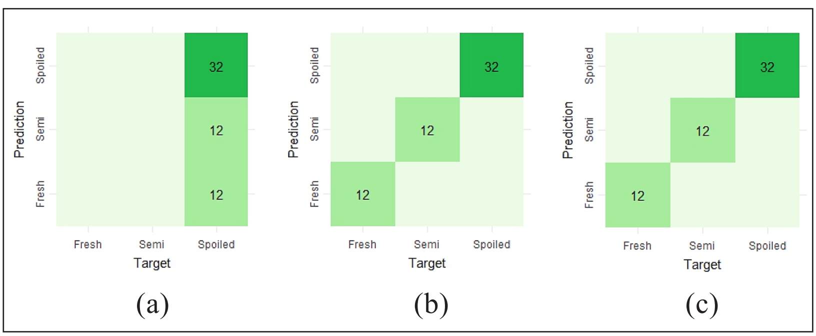


Figure 17 illustrates the confusion matrix used for the *Sawi Masak Air* classification tasks. The performance of the SVM method in categorising only the "Spoiled" class resulted in an average accuracy of 57.14 percent throughout this dataset, as seen in the first image. However, no misclassification errors occurred when the dataset was balanced by utilising the oversampling and undersampling procedures. This is because the algorithm correctly classified the dataset into the correct categories.

Figure 17

SVM Classification of Sawi Masak Air: (a) Unbalanced, (b) Oversampled, and (c) Undersampled

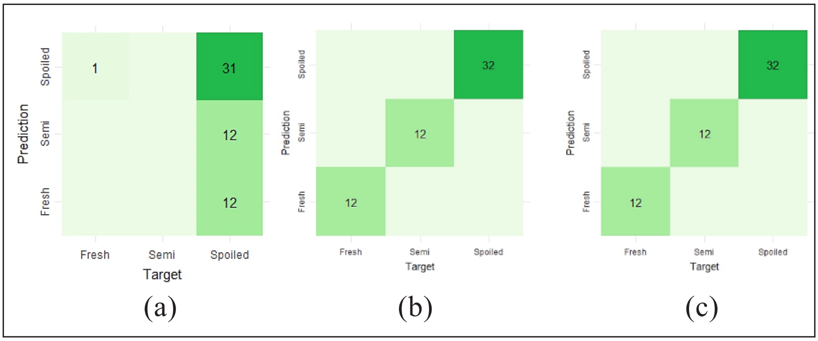


Classification Result of RF

The classification results of the RF algorithm will be examined in this section. The confusion matrices for the dataset *Ayam Goreng Berempah* are shown in Figure 18. According to the first confusion matrix, just one sample—in an unbalanced dataset for *Ayam Goreng Berempah*—was mistakenly labelled as “Fresh” rather than “Spoiled.” The “Fresh” and “Semi-Fresh” courses’ data points were completely mislabeled.

Figure 18

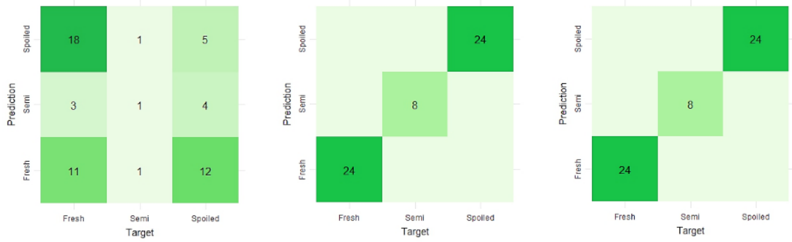
RF Classification of Ayam Goreng Berempah: (a) Unbalanced, (b) Oversampled, and (c) Undersampled



The average accuracy for this dataset was, therefore, 55.36 percent. The average accuracy climbed to 100 percent; nevertheless, once the dataset was balanced using the oversampling and undersampling techniques. Conversely, Figure 19 displays the confusion matrices for the *Ayam Masak Kunyit* dataset. The dataset previously mentioned that the confusion matrix of the imbalanced dataset indicated that the RF could not reliably distinguish the levels of food contamination. The dataset’s average accuracy was, therefore, 30.36 percent. However, the algorithm achieved an average accuracy of 100 percent when the dataset was balanced using the oversampling and undersampling strategies.

Figure 19

RF Classification of Ayam Masak Kunyit: (a) Unbalanced, (b) Oversampled, and (c) Undersampled

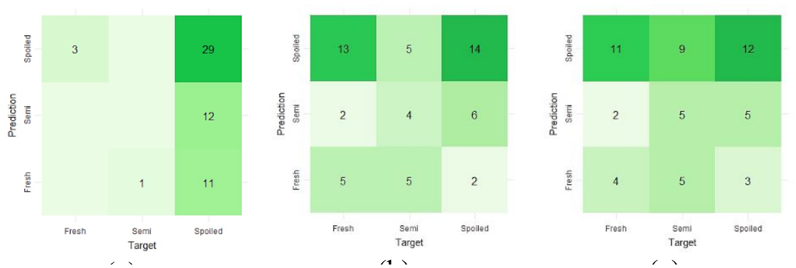


(a) (b) (c)

The confusion matrix for the classification tasks completed on *Ayam Masak Lemak* is then displayed in Figure 20. When the datasets were out of balance, this dataset's average accuracy was 55.35 percent. As can be seen from Figure 20(a), just one sample was incorrectly classified as "Semi-Fresh" from the "Spoiled" class when the *Ayam Masak Lemak* dataset was unbalanced. In the meantime, the "Fresh" and "Semi-Fresh" class data points were entirely mislabeled.

Figure 20

RF Classification of Ayam Masak Lemak: (a) Unbalanced, (b) Oversampled, and (c) Undersampled

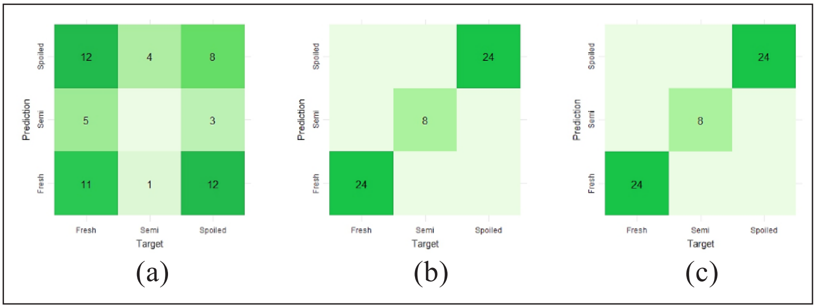


(a) (b) (c)

On the other hand, the classification accuracy was 100 percent, and there were no incorrect classifications when the dataset was balanced by employing the oversampling and undersampling approaches. A representation of the confusion matrix for the *Nasi Goreng Daging* classification tasks can be found in Figure 21, which can be found below. When the datasets were not in equilibrium, the average accuracy for this dataset was 33.93 percent, according to the data. As a result of the use of oversampling and undersampling strategies to achieve a balanced dataset, the classification average accuracy for this particular dataset also approached a value of 100 percent.

Figure 21

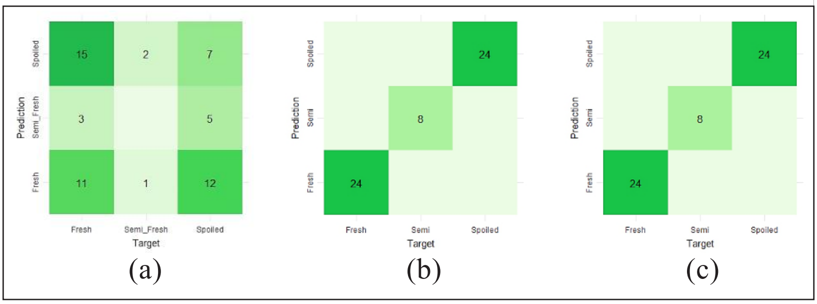
RF Classification of Nasi Goreng Daging: (a) Unbalanced, (b) Oversampled, and (c) Undersampled



A visualisation of the confusion matrices for the dataset *Sup Daging* can be found in Figure 22, which can be found below. RF could not correctly categorise food contamination levels when the dataset for *Sup Daging* was uneven, as illustrated by the first confusion matrix. This was the case because the dataset displayed inconsistent values. The average accuracy for this dataset was in the range of 32.14 percent. When the dataset was balanced by utilising the oversampling and undersampling procedures, the algorithm correctly classified the dataset into the relevant groups, achieving a classification accuracy of 100 percent. As a consequence of this, there was no error in classification.

Figure 22

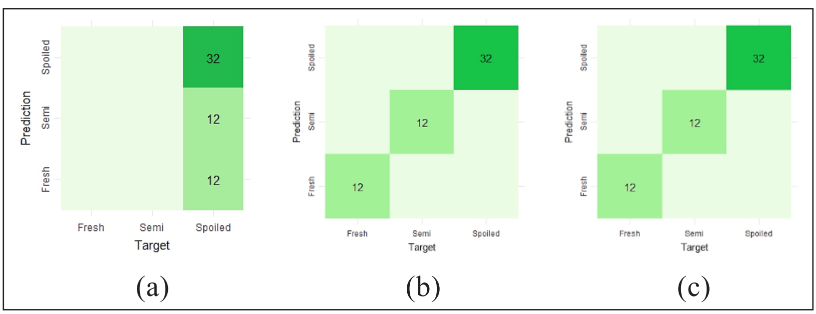
RF Classification of Sup Daging: (a) Unbalanced, (b) Oversampled, and (c) Undersampled



The confusion matrix created for the *Sawi Masak Air* classification tasks is presented in Figure 23. As seen in the first picture, the RF technique can only accurately classify the "Spoiled" category. This dataset has an average accuracy of 32.14 percent due to this measurement. However, no misclassification errors occurred when the dataset was balanced by utilising the oversampling and undersampling procedures. This is because the algorithm could appropriately classify the dataset into suitable classes.

Figure 23

RF Classification of Sawi Masak Air: (a) Unbalanced, (b) Oversampled, and (c) Undersampled

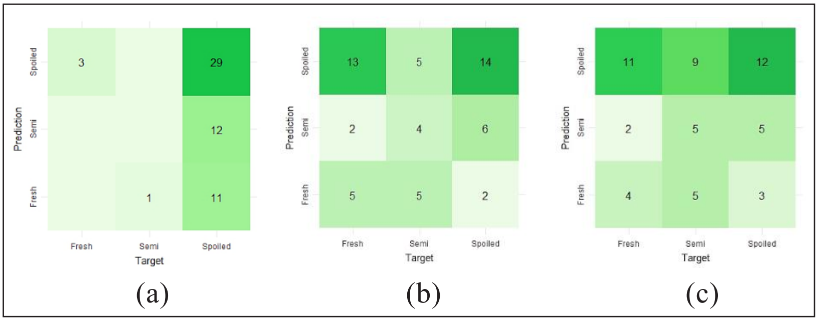


Classification Result of LDA

The classification outcomes for the LDA method will be covered in this section. The *Ayam Goreng Berempah* dataset's confusion matrices are displayed in Figure 24 below. The first confusion matrix demonstrates that three samples were incorrectly identified as "Fresh" from the "Spoiled" class when the *Ayam Goreng Berempah* dataset was unbalanced. On the other hand, the "Fresh" and "Semi-Fresh" class data points were entirely incorrectly classified. Consequently, 51.79 percent was the dataset's average accuracy.

Figure 24

LDA Classification of Ayam Goreng Berempah: (a) Unbalanced, (b) Oversampled, and (c) Undersampled



However, when the dataset was balanced by utilising oversampling and undersampling, the classification accuracy went up to 41.07 percent and 37.50 percent, respectively. The overfitting that occurred during the oversampling process and the data loss that occurred during the undersampling operations were the causes of these outcomes. Oversampling was used to accomplish the task of balancing the dataset, and in order to prevent any loss of information, the data points were duplicated. Nevertheless, it needs help with data duplication, which ultimately results in overfitting. Undersampling is a strategy that involves reducing specific samples of the majority class to achieve a more balanced data set. However, this method results in a loss of information because of the data reduction.

Meanwhile, Figure 25 presents the confusion matrices corresponding to the *Ayam Masak Kunyit* dataset. As was the case with the dataset discussed before, the confusion matrix for the imbalanced dataset indicated that the LDA could not accurately categorise the levels of food contamination. As a consequence of this, the average accuracy of this dataset was 32.14 percent. The algorithm, however, was not successful in achieving an average accuracy of 100 percent when the dataset was balanced by employing the oversampling and undersampling approach. An overfitting of the dataset caused by oversampling and a loss of data caused by undersampling both contributed to worsening the results.

Figure 25

LDA Classification of Ayam Masak Kunyit: (a) Unbalanced, (b) Oversampled, and (c) Undersampled

(a) (b) (c)

After that, the confusion matrix for the classification tasks that were completed on *Ayam Masak Lemak* is displayed in Figure 26. In situations when the datasets were not in equilibrium, the average accuracy of this dataset was 55.35 percent. The first confusion matrix demonstrates that when the dataset for *Ayam Masak Lemak* was imbalanced, just one sample was incorrectly identified as "Semi-Fresh" from the "Spoiled" class. While this was happening, just two samples were accurately identified as "Semi-Fresh," while the remaining samples were utterly misclassified.

Figure 26

LDA Classification of Ayam Masak Lemak: (a) Unbalanced, (b) Oversampled, and (c) Undersampled

(a) (b) (c)

When the dataset was balanced using the oversampling and undersampling approach, the algorithm could not attain an average accuracy of one hundred per cent. It is the same situation found in the *Ayam Masak Lemak* dataset. The classification results for the dataset with oversampling only reached up to 35.71 percent, whereas the results for the dataset with undersampling reached 37.50 percent. The overfitting that occurred during the oversampling and the loss of data during the undersampling were the causes of these outcomes. The confusion matrix for the classification jobs that *Nasi Goreng Daging* completed is displayed below in Figure 27. The average accuracy for this dataset was found to be 44.64 percent when the datasets were not balanced. The confusion matrix below demonstrates that LDA was not successful in classifying most of the data points into their respective classes, even if the data were not balanced.

Figure 27

LDA Classification of Nasi Goreng Daging: (a) Unbalanced, (b) Oversampled, and (c) Undersampled

(a) (b) (c)

Although the dataset was adequately balanced by utilising the oversampling and undersampling approach, the classification accuracy results were not even close to being sufficient. Regarding oversampling and undersampling, the classification results only reached 39.29 percent and 33.93 percent, respectively. In the same way as previous datasets did, this dataset experienced the problem of overfitting due to oversampling and data losses due to undersampling. In Figure 28, which can be seen below, the LDA classification results for the dataset *Sup Daging* are displayed respectively.

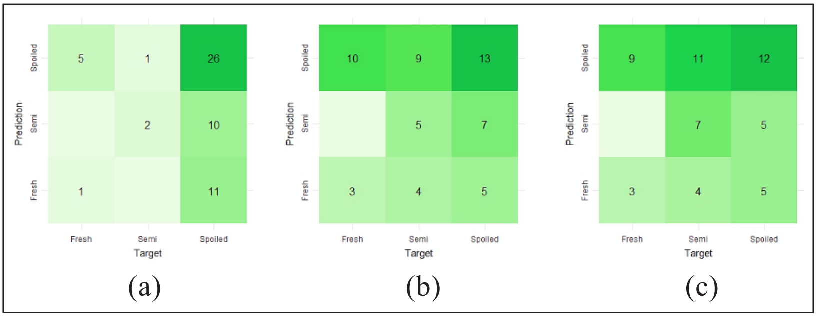
Figure 28

LDA Classification of Sup Daging: (a) Unbalanced, (b) Oversampled, and (c) Undersampled

As with other datasets mentioned above, the average accuracy for this dataset was 42.86 percent when the datasets under consideration were not balanced. Even though the data were not balanced, the LDA could not correctly categorise most of the data points into their respective classes, as seen by the confusion matrix. The method could not attain an average accuracy of 100 percent when the dataset was balanced by employing oversampling and undersampling. The dataset was overfit, and the data was lost in undersampling. As a result, the results were more unfavourable. The confusion matrix for the classification tasks completed for *Sawi Masak Air* is displayed in Figure 29, which may be seen below.

Figure 29

LDA Classification of Sawi Masak Air: (a) Unbalanced, (b) Oversampled, and (c) Undersampled



When the dataset for *Sawi Masak Air* was imbalanced, only 26 samples were correctly identified as "Spoiled," as seen by the first confusion matrix showing the analysis results. Only two samples were accurately identified as "Semi-Fresh," and only one was identified as "Fresh," resulting in an average accuracy rate of 51.79 percent. The classification accuracy, on the other hand, declined for both of the balancing strategies when the dataset was balanced using oversampling and undersampling. In the case of oversampling, the classification accuracy only reached 37.50 percent and 39.29 percent. These results were brought about by overfitting that occurred during the oversampling procedure and data loss that occurred during the undersampling procedure.

It was evident from the results presented in this work that variations in the degree of precision were observed in the classification of distinct food samples. This variation can be attributed to the diverse methods of food preparation employed in the samples, including frying, simmering proteins or vegetables in soups, and employing coconut milk to prepare proteins or vegetables. Various cooking techniques may also impact the food samples' rate of deterioration. Therefore, it can be deduced that the methodologies utilised in preparing the food samples influence the precision of the classification results.

CONCLUSION

An e-nose application utilising machine learning and food odours was developed to detect and categorise the level of contamination in uneaten prepared food. The dataset for this work consisted of traditional Malaysian lunch and dinner items, which were then stored as leftovers. The samples were analysed using machine learning techniques such as k-NN, SVMs, RF and LDA. The samples were categorised into three groups: "Fresh," "Semi-Fresh," and "Spoiled." The work demonstrated that k-NN, SVM and RF achieved an accuracy between 90 percent and 100 percent when utilising oversampling and undersampling methodologies. This suggests that these methods effectively classified the contamination level of the dataset. However, based on the results obtained, the LDA algorithm could not perform well in the classification task due to the datasets experiencing overfitting and data loss after the balancing stage. Nevertheless, the work's findings suggest that the electronic nose showed promise as a tool for classifying the level of contamination in uneaten prepared food.

ACKNOWLEDGMENT

This work was supported by the Ministry of Higher Education through the Fundamental Research Grant Scheme (FRGS/1/2022/STG07/UNIMAP/02/5).

REFERENCES

Abu-Khalaf, N. (2021). Identification and quantification of olive oil quality parameters using an electronic nose. *Agriculture (Switzerland), 11*(7). https://doi.org/10.3390/agriculture11070674

Amirudin, N., & Gim, T. H. T. (2019). Impact of perceived food accessibility on household food waste behaviors: A case of the Klang Valley, Malaysia. *Resources, Conservation and Recycling, 151*. https://doi.org/10.1016/j.resconrec.2019.05.011

Bejjanki, K. K., Gyani, J., & Gugulothu, N. (2020). Class imbalance reduction (CIR): A novel approach to software defect prediction in the presence of class imbalance. *Symmetry, 12*(3). https://doi.org/10.3390/sym12030407

Betz, A., Buchli, J., Göbel, C., & Müller, C. (2015). Food waste in the Swiss food service industry - Magnitude and potential for reduction. *Waste Management, 35,* 218–226. https://doi.org/10.1016/j.wasman.2014.09.015

Bindra, N., & Sood, M. (2019). Detecting DDoS attacks using machine learning techniques and contemporary intrusion detection dataset. *Automatic Control and Computer Sciences, 53*(5), 419–428. https://doi.org/10.3103/S0146411619050043

Buratti, S., Malegori, C., Benedetti, S., Oliveri, P., & Giovanelli, G. (2018). E-nose, e-tongue and e-eye for edible olive oil characterisation and shelf life assessment: A powerful data fusion approach. *Talanta, 182*(January), 131–141. https://doi.org/10.1016/j.talanta.2018.01.096

Camardo Leggieri, M., Mazzoni, M., Fodil, S., Moschini, M., Bertuzzi, T., Prandini, A., & Battilani, P. (2021). An electronic nose supported by an artificial neural network for the rapid detection of aflatoxin B1 and fumonisins in maise. *Food Control, 123*. https://doi.org/10.1016/j.foodcont.2020.107722

Chauhan, C., Dhir, A., Akram, M. U., & Salo, J. (2021). Food loss and waste in food supply chains. A systematic literature review and framework development approach. *Journal of Cleaner Production, 295*. https://doi.org/10.1016/j.jclepro.2021.126438

Chauhan, V. K., Dahiya, K., & Sharma, A. (2019). Problem formulations and solvers in linear SVM: A review. *Artificial Intelligence Review, 52*(2), 803–855. https://doi.org/10.1007/s10462-018-9614-6

Chen, Q., Zhao, J., Chen, Z., Lin, H., & Zhao, D. A. (2011). Discrimination of green tea quality using the electronic nose technique and the human panel test, comparison of linear and nonlinear classification tools. *Sensors and Actuators, B: Chemical, 159*(1), 294–300. https://doi.org/10.1016/j.snb.2011.07.009

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology, 77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Feng, W., Huang, W., & Ren, J. (2018). Class imbalance ensemble learning based on the margin theory. *Applied Sciences (Switzerland), 8*(5). https://doi.org/10.3390/app8050815

Feyzioglu, A., & Taspinar, Y. S. (2023). Beef quality classification with reduced e-nose data features according to beef cut types. *Sensors, 23*(4), 2222. https://doi.org/10.3390/s23042222

Food wastage leads to climate change — Forum Air Malaysia | What You Think | Malay Mail. (n.d.). Retrieved September 21, 2021, from https://www.malaymail.com/news/what-you-think/2018/07/20/food-wastage-leads-to-climate-change-forum-air-malaysia/1654305

Gębicki, J., & Szulczyński, B. (2018). Discrimination of selected fungi species based on their odour profile using prototypes of electronic nose instruments. Measurement: *Journal of the International Measurement Confederation, 116*(August 2015), 307–313. https://doi.org/10.1016/j.measurement.2017.11.029

Gonçalves, W. B., Teixeira, W. S. R., Sampaio, A. N. da C. E., Martins, O. A., Cervantes, E. P., Mioni, M. de S. R., Gruber, J., & Pereira, J. G. (2023). Combination of the electronic nose with microbiology as a tool for rapid detection of Salmonella. *Journal of Microbiological Methods, 212,* 106805. https://doi.org/10.1016/j.mimet.2023.106805

Gu, S., Wang, J., & Wang, Y. (2019). Early discrimination and growth tracking of Aspergillus spp. contamination in rice kernels using electronic nose. *Food Chemistry, 292*(April), 325–335. https://doi.org/10.1016/j.foodchem.2019.04.054

Gunders, D., & Bloom, J. (2017). Wasted: How America is losing up to 40% of its food from farm to fork to landfill. *Natural Resources Defense Council, 1*(August), 1–58. www.suerossi.com

Hosseini, H., Minaei, S., & Beheshti, B. (2023). A dedicated electronic nose combined with chemometric methods for detection of adulteration in sesame oil. *Journal of Food Science and Technology, 60*(10), 2681–2694. https://doi.org/10.1007/s13197-023-05792-2

Hsieh, Y. C., & Yao, D. J. (2018). Intelligent gas-sensing systems and their applications. *Journal of Micromechanics and Microengineering, 28*(9), 093001. https://doi.org/10.1088/1361-6439/aac849

Huang, L., Zhao, J., Chen, Q., & Zhang, Y. (2014). Nondestructive measurement of total volatile basic nitrogen (TVB-N) in pork meat by integrating near infrared spectroscopy, computer vision and electronic nose techniques. *Food Chemistry, 145,* 228–236. https://doi.org/10.1016/j.foodchem.2013.06.073

Interest, C. (2017). Talking Trash – The Chemistry Behind the Smell of Garbage | SciTech Connect. *Elsevier SciTech Connect.* https://scitechconnect.elsevier.com/talking-trash-chemistry-behind-garbage-smell/

Ismail, M. H., Ghazi, T. I. M., Hamzah, M. H., Manaf, L. A., Tahir, R. M., Nasir, A. M., & Omar, A. E. (2020). Impact of movement control order (MCO) due to coronavirus disease (Covid-19) on food waste generation: A case study in Klang Valley, Malaysia. *Sustainability (Switzerland), 12*(21), 1–17. https://doi.org/10.3390/su12218848

Jiang, S., & Liu, Y. (2020). Gas sensors for volatile compounds analysis in muscle foods: A review. *TrAC - Trends in Analytical Chemistry, 126,* 115877. https://doi.org/10.1016/j.trac.2020.115877

Jiarpinijnun, A., Osako, K., & Siripatrawan, U. (2020). Visualisation of volatomic profiles for early detection of fungal infection on storage Jasmine brown rice using electronic nose coupled with chemometrics. Measurement: *Journal of the International Measurement Confederation, 157,* 107561. https://doi.org/10.1016/j.measurement.2020.107561

Kefi-Fatteh, T., Ksantini, R., Kaâniche, M. B., & Bouhoula, A. (2019). A novel incremental one-class support vector machine based on low variance direction. *Pattern Recognition, 91,* 308–321. https://doi.org/10.1016/j.patcog.2019.02.027

Li, Q., Gu, Y., & Jia, J. (2017). Classification of multiple Chinese liquors by means of a QCM-based e-nose and MDS-SVM classifier. *Sensors (Switzerland), 17*(2). https://doi.org/10.3390/s17020272

Li, Q., Gu, Y., & Wang, N. F. (2017). Application of random forest classifier by means of a QCM-based e-nose in the identification of Chinese liquor flavors. *IEEE Sensors Journal, 17*(6), 1788–1794. https://doi.org/10.1109/JSEN.2017.2657653

Lipinski, B. (2013). By the numbers: Reducing food loss and waste. *World Resources Institute.* 1–4. http://www.wri.org/blog/2013/06/numbers-reducing-food-loss-and-waste

Liu, H., Li, Q., Yan, B., Zhang, L., & Gu, Y. (2019). Bionic electronic nose based on mos sensors array and machine learning algorithms used for wine properties detection. *Sensors (Switzerland), 19*(1). https://doi.org/10.3390/s19010045

Mahmodi, K., Mostafaei, M., & Mirzaee-Ghaleh, E. (2019). Detection and classification of diesel-biodiesel blends by LDA, QDA and SVM approaches using an electronic nose. *Fuel, 258*(July), 116114. https://doi.org/10.1016/j.fuel.2019.116114

Mohamed, A. E. (2017). Comparative study of four supervised machine learning techniques for classification. *International Journal of Applied Science and Technology, 7*(2), 5–18. https://doi.org/10.15546/aeei-2014-0021

Mohammed, R., Rawashdeh, J., & Abdullah, M. (2020). Machine learning with oversampling and undersampling techniques: Overview study and experimental results. *2020 11th International Conference on Information and Communication Systems, ICICS 2020, April,* 243–248. https://doi.org/10.1109/ICICS49469.2020.239556

Naim, M. N., & Rahman, H. A. (2020). Knowledge, attitude and practice on food waste management among food vendors in Universiti Putra Malaysia, Serdang, Selangor. *Malaysian Journal of Medicine and Health Sciences, 16*(8), 118–127.

Oates, M. J., Abu-Khalaf, N., Molina-Cabrera, C., Ruiz-Canales, A., Ramos, J., & Bahder, B. W. (2020). Detection of lethal bronzing disease in cabbage palms (Sabal palmetto) using a low-cost electronic nose. *Biosensors, 10*(11). https://doi.org/10.3390/bios10110188

Oates, M. J., Fox, P., Sanchez-Rodriguez, L., Carbonell-Barrachina, Á. A., & Ruiz-Canales, A. (2018). DFT based classification of olive oil type using a sinusoidally heated, low cost electronic nose. *Computers and Electronics in Agriculture, 155,* 348–358. https://doi.org/10.1016/j.compag.2018.10.026

Ong, S. Q., Lee, B. B., Tan, G. P., & Maniam, S. (2019). Food waste pattern in tertiary education institution in Penang: Quantitative comparison of food waste composition between semester break and start. *Pertanika Journal of Social Sciences and Humanities, 27*(1), 379–389.

Ordukaya, E., & Karlik, B. (2017). Quality control of olive oils using machine learning and electronic nose. *Journal of Food Quality, 2017,* 1-7. https://doi.org/10.1155/2017/9272404

Pampoukis, G., Lytou, A. E., Argyri, A. A., Panagou, E. Z., & Nychas, G.-J. E. (2022). Recent advances and applications of rapid microbial assessment from a food safety perspective. *Sensors, 22*(7), 2800. https://doi.org/10.3390/s22072800

Papadopoulou, O. S., Panagou, E. Z., Mohareb, F. R., & Nychas, G. J. E. (2013). Sensory and microbiological quality assessment of beef fillets using a portable electronic nose in tandem with support vector machine analysis. *Food Research International, 50*(1), 241–249. https://doi.org/10.1016/j.foodres.2012.10.020

Putri, L. A., Rahman, I., Puspita, M., Hidayat, S. N., Dharmawan, A. B., Rianjanu, A., Wibirama, S., Roto, R., Triyana, K., & Wasisto, H. S. (2023). Rapid analysis of meat floss origin using a supervised machine learning-based electronic nose towards food authentication. *NPJ Science of Food, 7*(1), 31. https://doi.org/10.1038/s41538-023-00205-2

Qiu, S., Gao, L., & Wang, J. (2014). Classification and regression of ELM, LVQ and SVM for E-nose data of strawberry juice. *Journal of Food Engineering, 144,* 77–85. https://doi.org/10.1016/j.jfoodeng.2014.07.015

Qiu, S., Wang, J., Tang, C., & Du, D. (2015). Comparison of ELM, RF, and SVM on E-nose and E-tongue to trace the quality status of mandarin (Citrus unshiu Marc.). *Journal of Food Engineering, 166,* 193–203. https://doi.org/10.1016/j.jfoodeng.2015.06.007

Qiu, S., & Wang, J. (2017). The prediction of food additives in the fruit juice based on electronic nose with chemometrics. *Food Chemistry, 230,* 208–214. https://doi.org/10.1016/j.foodchem.2017.03.011

Roy, M., & Yadav, B. K. (2021). Electronic nose for detection of food adulteration: A review. *Journal of Food Science and Technology,* 1-13. https://doi.org/10.1007/s13197-021-05057-w

Schroeder, V., Evans, E. D., Wu, Y. C. M., Voll, C. C. A., McDonald, B. R., Savagatrup, S., & Swager, T. M. (2019). Chemiresistive sensor array and machine learning classification of food. *ACS Sensors, 4*(8), 2101–2108. https://doi.org/10.1021/acssensors.9b00825

Tan, J., & Xu, J. (2020). Applications of electronic nose (e-nose) and electronic tongue (e-tongue) in food quality-related properties determination: A review. *Artificial Intelligence in Agriculture, 4,* 104–115. https://doi.org/10.1016/j.aiia.2020.06.003

Tian, H., Wu, D., Chen, B., Yuan, H., Yu, H., Lou, X., & Chen, C. (2023). Rapid identification and quantification of vegetable oil adulteration in raw milk using a flash gas chromatography electronic nose combined with machine learning. *Food Control, 150,* 109758. https://doi.org/10.1016/J.FOODCONT.2023.109758

Tonini, D., Albizzati, P. F., & Astrup, T. F. (2018). Environmental impacts of food waste: Learnings and challenges from a case work on UK. *Waste Management, 76,* 744–766. https://doi.org/10.1016/j.wasman.2018.03.032

Wan Azman, W. N. F. S., Ku Azir, K. N. F., & Amir, A. (2022). Classification of odour in the leftover cooked food to determine contamination using machine learning. *Lecture Notes in Electrical Engineering,* 835, 831–841. https://doi.org/10.1007/978-981-16-8515-6_63

Wong, K. K. S., Sharifuddin, J., Nik Masdek, N. R., Wong, W. L., & Lai, K. S. (2021). Determinants of household's intention of practicing sustainable food waste management in Malaysia.

IOP Conference Series: Earth and Environmental Science, 756(1). https://doi.org/10.1088/1755-1315/756/1/012036

Xu, M., Wang, J., & Gu, S. (2019). Rapid identification of tea quality by E-nose and computer vision combining with a synergetic data fusion strategy. *Journal of Food Engineering, 241*(June 2018), 10–17. https://doi.org/10.1016/j.jfoodeng.2018.07.020

Zulkifli, A. A., Yusoff, M. Z. M., Manaf, L. A., Zakaria, M. R., Roslan, A. M., Ariffin, H., Shirai, Y., & Hassan, M. A. (2019). Assessment of municipal solid waste generation in Universiti Putra Malaysia and its potential for green energy production. *Sustainability (Switzerland), 11*(14). https://doi.org/10.3390/su11143909