Smart Agricultural Technology 9 (2024) 100624

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Revolutionizing agri-food technology: Development and validation of the Portable Intelligent Oil Recognition System (PIORS)

Montaser N.A. Ramadan [a], Mohammed A.H. Ali [a,*], Shin Yee Khoo [a,b], Layth Hamad [c], Mohammad Alkhedher [d]

[a] Department of Mechanical Engineering, Faculty of Engineering, Universiti Malaya, 50603, Kuala Lumpur, Malaysia
[b] Centre of Research Industry 4.0 (CRI 4.0), Faculty of Engineering, Universiti Malaya, 50603, Kuala Lumpur, Malaysia
[c] Qatar Mobility Innovations Center, Qatar University, Doha, Qatar
[d] Mechanical and Industrial Engineering, Abu Dhabi University, Abu Dhabi, United Arab Emirates

ARTICLE INFO

Keywords:
ESP32
Food safety
IoT
MOX
Oil recognition
PIORS

ABSTRACT

The increasing demand for high-quality vegetable oils has amplified the need for efficient, and accurate, oil recognition systems. This paper introduces the Portable Intelligent Oil Recognition System (PIORS), a pioneering advancement in the field. PIORS stands as the first-ever portable, AI-powered device tailored for rapid and accurate oil type identification, capable of delivering results within a mere five seconds. Using advanced machine learning, PIORS can distinguish between various oil types such as sesame, black seed, flaxseed, and almond oils, ensuring the quality and safety of the final product. The device's innovative design integrates an advanced sensor array with seamless Bluetooth connectivity, offering real-time data synchronization with mobile applications. Several Machine learning models, including Support Vector Machines (SVM), AdaBoost, Random Forest, K-Nearest Neighbors (K-NN), Gradient-Boosting Decision Tree (GBDT), and Extreme Gradient Boosting (XGBoost) were implemented and thoroughly validated to obtain correct categorization. The results show that the XGBoost instantaneous classification algorithm continuously beat the competition, obtaining an astounding success rate of 97 percent in predicting and differentiating between the four oil classes. PIORS not only sets a new standard for speed and efficiency in oil quality assessment but also paves the way for groundbreaking applications in food safety, environmental monitoring, and quality control processes. This paper details the development, implementation, and validation of PIORS, showcasing its potential to revolutionize the agri-food industry with its cutting-edge, AI-driven approach to oil recognition.

1. Introduction

Vegetable oils are essential in a wide range of industrial and food applications [1]. With increased concerns about food quality and adulteration, there is a greater need for effective detection and recognition systems for various types of oils [2]. Traditional approaches like Gas Chromatography (GC) and Mass Spectrometry (MS) have been central to the analysis and recognition of vegetable oils. These methods focus on identifying and quantifying the fatty acid composition, which can serve as a fingerprint for different types of oils. Despite their widespread use, these traditional techniques have certain limitations [3]. The main challenge with GC in analyzing vegetable oils lies in its extensive and time-consuming sample preparation process, which often includes the derivatization of fatty acids. This can introduce errors and biases, affecting the accuracy and reliability of the results. Additionally, GC is primarily focused on analyzing volatile components, limiting its ability to provide comprehensive insights into the complex matrix of vegetable oils, especially for non-volatile minor components that can be crucial for oil authentication and quality assessment [4]. MS, while powerful for identifying and quantifying the molecular composition of vegetable oils, comes with its own set of challenges. The complexity and high cost of MS equipment can be prohibitive for some laboratories. The technique is also susceptible to matrix effects, where the presence of other compounds in the sample can interfere with the accurate quantification of target molecules. Regular maintenance and skilled operation are essential for reliable MS analysis, adding to the operational overhead and

* Corresponding author.
E-mail address: hashem@um.edu.my (M.A.H. Ali).

https://doi.org/10.1016/j.atech.2024.100624
Received 4 September 2024; Received in revised form 23 October 2024; Accepted 23 October 2024
Available online 7 November 2024

requiring specialized knowledge [5]. The existing systems have several drawbacks, including high operational complexity, long analysis times, and vulnerability to environmental influences [6]. A comprehensive solution that addresses these issues could considerably advance present technologies. This paper describes an AI-powered vegetable oils recognition system that uses metal oxide (MOX) sensors [7] and machine learning techniques to classify various vegetable oils such as sesame, black seed, flaxseed, and almond oils in real-time. For the interpretation of sensor data, common machine learning algorithms are used such as Support Vector Machine (SVM), AdaBoost, Random Forest, K-Nearest Neighbors (K-NN), and Gradient-Boosting Decision Tree (GBDT) [8]. Because of the growing demand for fast and accurate recognition of vegetable oils, particularly in quality control and fraud prevention [9], this research is particularly pertinent. The contributions of the paper can be summarized as follows:

- Advanced AI Integration for Unmatched Accuracy: The system's integration of advanced machine learning algorithms, including Support Vector Machine (SVM), AdaBoost, Random Forest, K-Nearest Neighbors (K-NN), Gradient-Boosting Decision Tree (GBDT), and particularly XGBoost, represents a significant technological advancement. The XGBoost algorithm, in particular, demonstrates extraordinary performance, setting a new standard in the precision of oil classification.
- Innovation in Portability and Speed: PIORS revolutionizes the field of agri-food technology by being the first to introduce a portable, AI-powered intelligent oil recognition device. PIORS has the distinct capacity to quickly and reliably detect the type of oil, unlike any previous systems. This rapid analysis is not just a leap in speed but also in convenience, enabling on-the-spot, real-time testing in various settings without the need for bulky or stationary equipment.
- Seamless Connectivity and User-Friendly Interface: Beyond its analytical capabilities, PIORS is designed with the end-user in mind. It features seamless Bluetooth connectivity that allows for real-time data transmission to a mobile application, offering a user-friendly interface for immediate display and interpretation of results. This aspect of design not only enhances the usability of the device but also ensures that it can be effectively utilized by individuals with varying levels of technical expertise, thereby broadening its applicability across different sectors of the agri-food industry.

2. Related work

Several different sensor technologies, including optical, electrochemical, and electronic nasal sensors, have been utilized in food quality monitoring [10]. Recently, there has been a surge in enthusiasm for using metal oxide gas detectors for monitoring food quality due to their great sensitivity and relatively low cost [11]. The electronic nose's sensors provide a wealth of information, and the idea of electronic olfactometry relies heavily on the processing of the data produced by the device. Applications for electronic olfactometry can be found in a variety of scientific domains, including environmental analysis [12], food and drinks quality control [13], monitoring biological processes [14], and even disease detection [15]. In recent years, research has begun to combine sensor technology with machine learning algorithms in order to provide real-time monitoring in the food business [16]. However, applications that specifically focus on vegetable oils are still in their infancy, making this a potentially fruitful area for additional investigation. While the focus on vegetable oil applications is emerging, with studies like [17] showcasing the use of gas chromatography-mass spectrometry (GC–MS SPME), organoleptic analysis, and Small Sensor Systems (S3) for evaluating EVOO quality in Umbria, Italy, these approaches reveal significant variations in volatile organic compounds. Although S3 successfully differentiates among EVOOs, highlighting its speed and cost-effectiveness, the method's reliance on sensor systems introduces potential challenges in sensor drift and environmental influence, not fully explored in this context. The study in [18] emphasizes Principal Component Analysis (PCA) for identifying unknown vegetable oils through infrared spectra, underscoring PCA's ability to simplify complex datasets. However, the practical application of such chemometric techniques necessitates a deep understanding of multivariate data analysis, possibly limiting its accessibility to specialists. In [19], Laser-Induced Breakdown Spectroscopy (LIBS) offers a novel parameter for estimating saturated fatty acids in oils based on the "Swan band" of C2 emission. While this method shows a promising correlation between C2 emission and fatty acid content, its specificity and sensitivity in broader applications remain unaddressed. Fourier Transform Infrared (FT-IR) spectroscopy's utility in detecting oil adulteration, as demonstrated in [20], benefits from enhancements like wavelength selection and multiplicative signal correction (MSC). Despite its accuracy in identifying oils and predicting physical properties, the technique's effectiveness can be hampered by the low selectivity for comparable substances. The research in [21] presents an efficient ANN-based method for classifying oils such as Canola and Soybean using induced fluorescence spectra. While this approach allows for quick classification with minimal data manipulation, the adaptability of the method to accurately detect adulteration or subtle quality differences without extensive data preprocessing is not clear. Furthermore, [22] delves into adulteration detection using fatty acid profiles via GC/MS, combining unsupervised and supervised statistical methods. Although it achieves sensitive adulteration detection to a level of 10%, the methodology's performance in the face of adulterants closely mimicking target oil profiles is not thoroughly evaluated. The potential for machine learning systems to analyze complicated sensor data has been demonstrated in [23,24]. SVM, GBDT, and neural networks have all been used successfully to categorize various gases and food items [25]. Overall, existing systems encounter challenges such as sensor drift, the impact of environmental variables, and limited selectivity among similar substances [26]. Despite recent advancements showing promising avenues for authenticating and assessing the quality of vegetable oils, these challenges highlight the essential need for ongoing refinement and comprehensive validation of analytical techniques within the agri-food sector.

3. Materials and methods

3.1. System description

The proposed project is based on an array of eight metal oxide (MOX) gas sensors, each of which is linked to a microcontroller. The microcontroller unit (MCU) is in charge of controlling the sensors' operations and reading data from each one sequentially. As shown in Fig. 1, various vegetable oils, including sesame, black seed, flaxseed, and almond oils, are used for testing and classification. Next, the main controller is a 32-bit ESP32 microcontroller. The ESP32 is a low-cost, low-power system on a chip (SoC) series created by Espressif Systems. It is notable for its integrated Wi-Fi and dual-mode Bluetooth capabilities, making it a popular choice for building Internet of Things (IoT) applications and projects [27]. The MCU will preprocess all of the gathered data before applying it to a ready-made machine-learning model that has already been created and uploaded. The identification of the oil should be the final outcome.

3.2. Hardware design of measurement system

The proposed AI-powered oil detection system is built around a pattern recognition module and a series of gas sensors. This collection of gas sensors contains several varieties, each with a different level of specificity. Each sensor can operate independently, transforming chemical information from various gas mixtures into a detectable signal. Each sensor's unique sensitivity and selectivity results in a wide range of responses from the sensor array as a whole. Then, the signal analysis module examines these signals to accomplish the oil classification task.

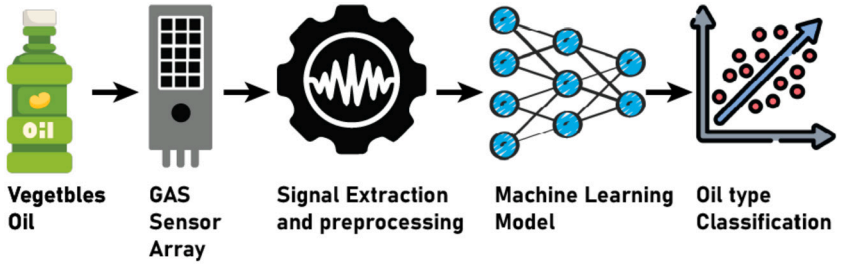


Fig. 1. Proposed AI-powered oil recognition overview.

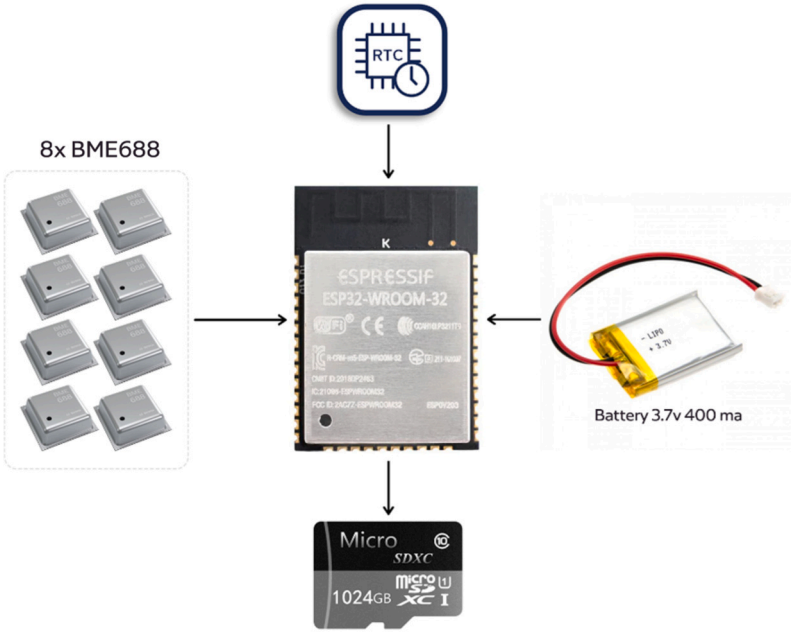


Fig. 2. Sensor Array and acquisition system.

Fig. 2 depicts a detailed breakdown of the planned project's components. The sensor array and the control system are the two primary components of the hardware. The sensor array consists of an 8-sensor MOX gas custom array using the BME688 sensor (Bosch Sensortec, Reutlingen, Germany). A 32-bit ESP32 microcontroller with integrated 16 MB Flash memory, Bluetooth and Wi-Fi, an SD card slot, a Real-Time Clock (RTC), and a 1.5-watt lithium battery power the control system.

Fig. 3 shows that the BME 688 sensor has a size of 3.00 mm x 3.00 mm x 0.93 mm, which is relatively small compared to the human finger. The core of the gas sensor is a metal-oxide semiconductor layer. When the sensor is heated and exposed to air, the metal oxide's surface interacts with the gases present. Different gases cause changes in the surface chemistry of the metal oxide, altering its electrical resistance. The BME688 periodically heats the MOX sensor to a high temperature to clean the surface, ensuring accurate and stable readings over time. This process also allows the sensor to differentiate between different gases based on their specific interaction patterns with the heated metal oxide. The presence of target gases in the environment changes the conductivity or resistance of the MOX layer. By measuring these changes in resistance, the sensor can determine the concentration of certain gases. The pattern and magnitude of resistance change provide insights into the type and quantity of gases present [28]. The BME688 sensor is equipped to detect a broad spectrum of gases, notably Volatile Organic Compounds (VOCs) which are prevalent in various environments due to their emission from paints, adhesives, cleaning products, and office equipment. These compounds are significant indicators of indoor air quality. Additionally, the sensor is capable of identifying breath VOCs, which emanate from human breath, and has applications in health monitoring, highlighting its utility in non-invasive wellness and fitness tracking devices. It also detects odorous gases, which are crucial for applications aimed at maintaining or improving the ambiance of living spaces, such as smart home systems that monitor and control air quality to ensure a comfortable and odor-free environment [7].

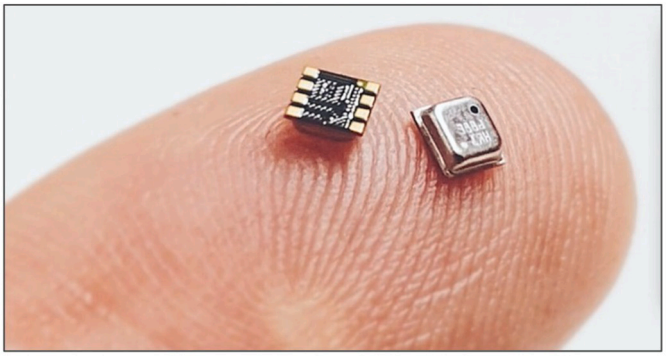


Fig. 3. The BME688 sensor. (https://www.bosch-sensortec.com/products/environmental-sensors/gas-sensors/bme688/).

The BME688 sensor is used as the foundation for electronic nose technology. The electronic nose measures the resistance in the detecting layer of the integrated MOX gas sensor. The sensor's inherent unpredictability and non-specificity are exploited in this study to construct an electronic nose capable of distinguishing between different oil types. Moreover, the BME688 sensor has two modes of operation: Sleep and

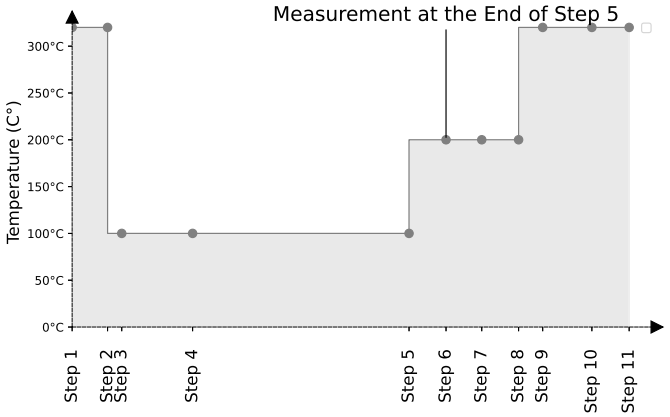


Fig. 4. BME688 heater profile.

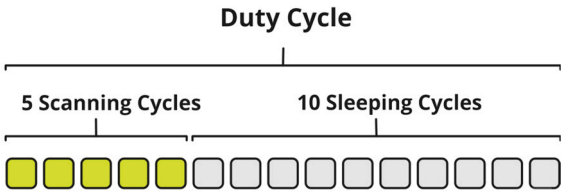


Fig. 5. BME688 duty profile.

Forced [29]. Following the completion of the power-up procedure, the sensor enters Sleep mode, during which it takes no measurements and consumes the least amount of power. Heater Profile (HP) and Duty Cycle (DC) are two of the configurable parameters that each sensor in the array can use. The Heater Profile in the BME688 refers to the controlled application of temperature cycles to the sensor's metal-oxide (MOX) gas-sensitive layer. This heating process is crucial for the sensor's operation, as the detection of gases by MOX sensors is highly dependent on the temperature of the sensing element. The temperature profile that the sensor runs through in ten steps during a scanning cycle is displayed in Fig. 4. The heater profile is designed with a 10-step configuration, each step having a duration set to 140 milliseconds, resulting in a total duration from Step 1 to Step 11 of 1,400 milliseconds. At the end of each step, a measurement is recorded and stored in the data.

By adjusting the duty cycle, the BME688 can significantly reduce its power consumption. During idle periods, the sensor consumes less power, which is critical for battery-operated devices such as smart home sensors, wearable health monitors, and IoT applications. The duty cycle determines how frequently the sensor takes measurements. A higher duty cycle means more frequent measurements, leading to more real-time data but at the cost of higher power consumption. Conversely, a lower duty cycle reduces power usage but at the expense of data freshness and possibly resolution. In the proposed project, the duty cycle with 5 scanning cycles and 10 sleeping cycles is implemented as shown in Fig. 5. Using the built-in micro-hotplate, the sensor runs through a series of temperature steps by a certain Heater Profile during a Scanning Cycle. Furthermore, the sensor is in sleep mode and does not take any readings during a sleeping cycle [28].

The sensor commences a measuring cycle in Forced mode to collect data on temperature (T), pressure (P), humidity (H), and gas resistance (G). When this cycle is completed, the sensor returns to Sleep mode until the next set of measurements is triggered. The BME688 includes built-in registers for precise data collection. These registers allow the sensor to be told to detect humidity or gas concentration only when necessary. For the sake of this investigation, data is collected for all four types of sensors at the same time.

3.3. Design of a 3D oil recognition enclosure

The datasheet states that the BME688 sensor is incredibly sensitive [30], and ought not to be touched directly. To protect the sensor array, a special 3D enclosure is designed as shown in Fig. 6, which is compatible with the main hardware parts shown in Fig. 2. Next, a 3D printer was used to print the enclosure using polylactic acid (PLA). The device has a height of 7.5 cm and a diameter of 4.3 cm. As seen in Fig. 6, there is a primary power switch, an LED indication that shows whether the device is ready for testing or not, and a particular area to input the tested paper once it has been dropped with tested oil. The final findings will be displayed through a mobile phone app.

3.4. Data acquisition

There are many different types of vegetable oils on the market; this study uses four types of oils that have similar colors and are difficult to identify by odor based on human smell. This is the most challenging aspect of this project: developing a system that can distinguish between vegetable oils despite their identical color and odor. The target oils for this project, as shown in Fig. 7, are sesame, black seed, flaxseed, and almond oils, which will be utilized to assess the performance and accuracy of the proposed AI-powered oil recognition system based on an array of eight MOX sensors. Each of these oil types is essential in daily life.

4. Datasets and model training

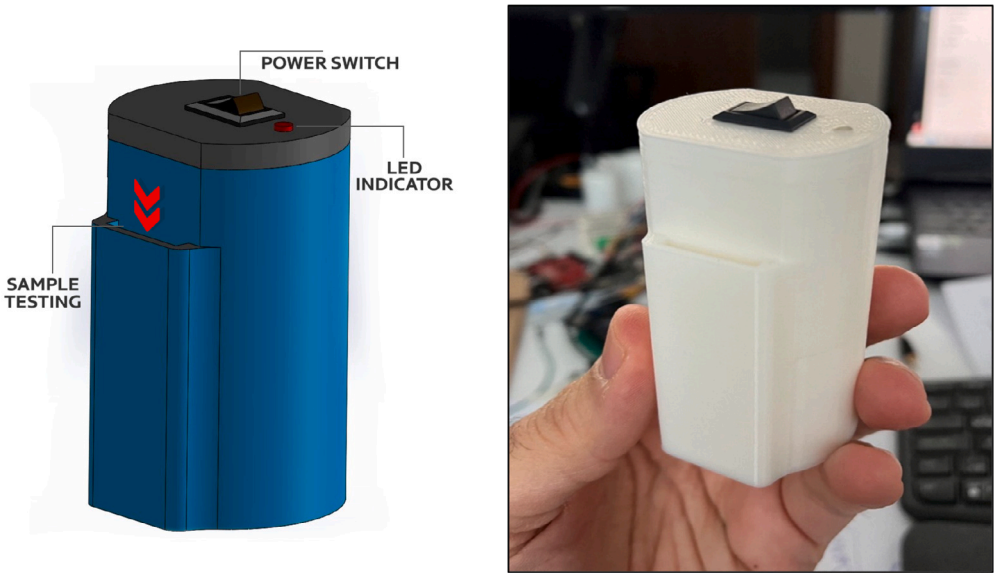
4.1. Training dataset

The dataset was collected using the electronics nose array sensors in four trials. As shown in Fig. 7, the AI recognition main board is installed inside a closed oil bottle with a 100-millimeter oil for data collection. The bottle is completely closed and a small hole is opened for the power cable, the same experiment was repeated with other types of oils. To save power, every 140 milliseconds, readings from each of the eight gas sensors are obtained, and saved to the SD card, then all sensors are then put into sleep mode for an additional 140 milliseconds. Fresh values from the second sensor are taken and written into the SD card in the second phase. The sensors then go back to sleep, Then the remaining sensors undergo the same procedure. Each trial keeps raw data for around 30 minutes, generating almost 20,000 distinct measurements from the eight MOX gas sensors for all types of oils as shown in Table 1. To assure the correctness of the data, a mechanism for normalizing the environment between tests is required. After each trial, fresh air data is collected for 30 minutes before testing the next variety of oil to minimize the environmental effect. The time resolution for the data acquisition used in this study is 500 ms.

Upon conducting eight tests, over 40,000 data points were amassed and saved to an SD card; half of these were disregarded as they were associated with fresh air. Table 1 outlines the size of the dataset for the four vegetable oils examined in the proposed study: Black Seed, Flaxseed, Sesame, and Almond oils. Since these four oils are often similar in color and occasionally in scent, distinguishing between them using human olfactory abilities is quite challenging. Consequently, the dataset is complex and requires preprocessing before subjecting it to evaluation through various machine learning algorithms.

4.2. Data preprocessing and visualization

In practical applications, it's common to come across extensive amounts of unstructured data that can be difficult for machine learning algorithms to manage efficiently. As a result, preprocessing becomes an essential preliminary step before applying various machine learning methods to the raw data. First, the data is zero-mean normalized.



(a) Schematic. (b) Prototype.

Fig. 6. The 3D oil recognition enclosure.

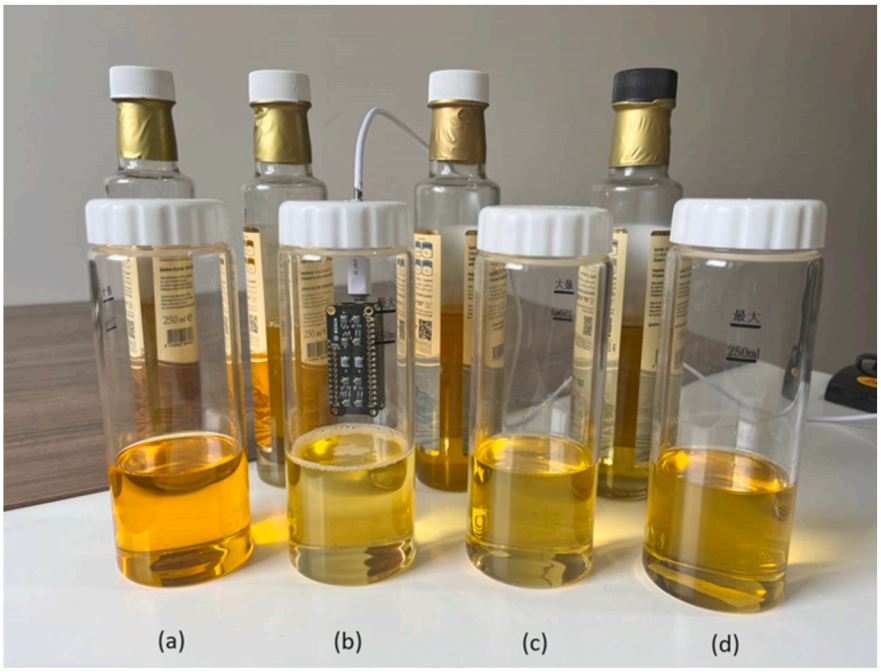


Fig. 7. Data acquisition using an array of gas sensors. (a) sesame, (b) black seed, (c) flaxseed, and (d) almond oils.

Table 1
Dataset description.

Class	No. of Samples (including fresh air data)	No. of samples (excluding fresh air data)
Black Seed Oil	10454	5454
Flaxseed Oil	9172	4172
Sesame Oil	10550	5550
Almond Oil	10000	5000

This step aims to simplify the data representation and potentially improve the effectiveness of subsequent machine-learning techniques. Feature bias is mitigated by removing the mean from each feature. After mean normalization, the dataset comprises multiple features captured at different time intervals. To simplify the data, dimensionality reduction techniques were employed such as Principal Component Analysis (PCA). This study uses the Scikit-learn library for data preprocessing and training the ML models. Fig. 8 highlights the first two principal components as determined by PCA, which were derived from measurements of sesame, black seed, flaxseed, and almond oils.

The figure reveals that after preprocessing, the different classes are effectively grouped. This dimensionality reduction not only removes irrelevant features but also sets a strong foundation for the following classification steps. The distinction between various classes becomes evident at each stage of the process. The array of sensors measures 4 important parameters including temperature (°C), humidity (%), pressure (Pa), and gas resistance (ω). The first three parameters, i.e. temperature, humidity, and pressure should remain constant for the same environmental

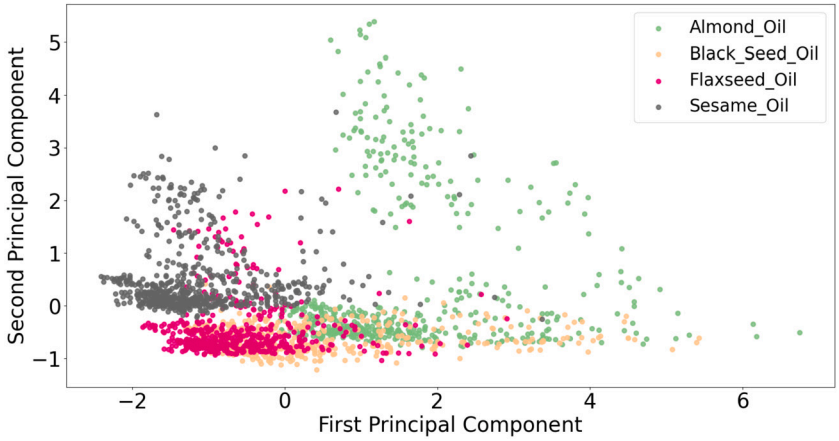


Fig. 8. PCA analysis for different oil types (i.e., sesame, black seed, flaxseed, and almond oils).

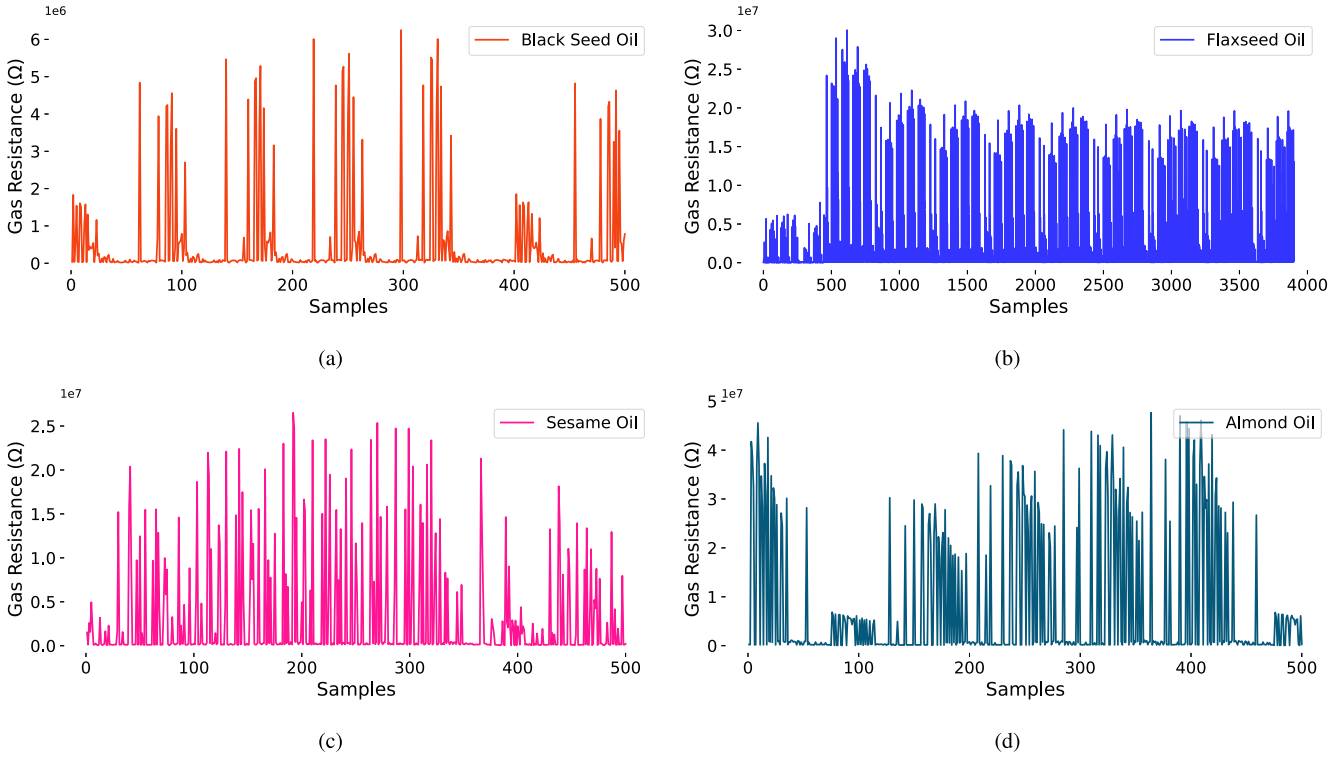


Fig. 9. The gas resistance readings for (a) almond oil, (b) black seed oil, (c) flaxseed oil, and (d) sesame oil.

condition. The value of gas resistance will change according to the examined oil odor, it plays an important role in the oil recognition task in this study. During the data cleaning and preparation stage, the variances among the four different types of oils i.e., Black Seed, Flaxseed, Sesame, and Almond oils are analyzed by plotting the sensor array's outputs. Specifically, Fig. 9 provides a comprehensive analysis of the response of each of the eight sensors to black seed oil over a specified period.

It is evident from these graphs that the changes in gas resistance, expressed in ohms (Ω), are directly related to the volatile substances that leaked from the oils in the hermetically sealed containers. One key observation was the substantial variations in sensor output throughout the measurement period. This dynamic change in output signifies that the sensor's interaction with the volatile compounds is complex and changes over time. What is particularly interesting from these graphical representations is the clear distinction in sensor responses when different oils are subjected to measurement. In other words, the data extracted from the sensor array showed noticeable differences, depending on which one of the four oils was placed in the testing environment. This point underscores the complexity and uniqueness of each oil type and further justifies the need for sophisticated machine-learning models for accurate classification. Fig. 10 displays the raw sensor data collected during a standard sampling cycle in the PIORS. In this figure, the gas resistance readings from each of the eight individual sensors are plotted through separate channels, each represented by a distinct color curve. The curves demonstrate a consistent pattern: they show a rapid initial increase, followed by a more gradual alteration, and then finally a swift decrease. This behavior aligns well with the commonly understood attributes of MOX sensors, known for their quick responsiveness but slower recovery times.

Moreover, as illustrated in Fig. 7, the layout of the eight sensors on the main measuring board is deliberate and strategic. Some sensors are located closer to the oil samples than others, and this proximity variance contributes to the unique sensor responses to the same oil type. These differential sensor outputs come together to form a composite

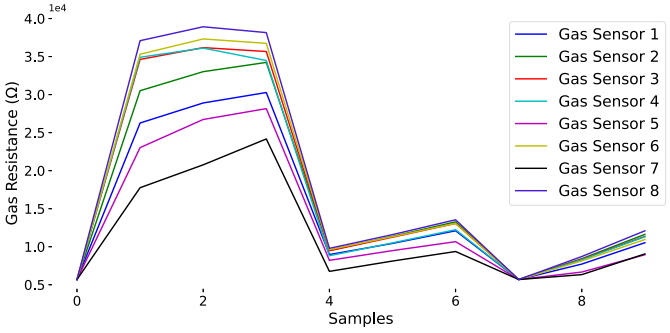


Fig. 10. Gas sensor reading for black seed oil.

"fingerprint" of the sample. This fingerprint serves as the basis for the subsequent Pattern Identification (PI) process, using machine learning techniques to classify the different types of oils based on their distinct sensor responses.

4.3. Model selection

The experiments use seven machine learning methods as listed in this section for oil-type classification. Each model is created and examined independently, and then the results of the analyses are compared to find the best-performing model.

Multi-layer perception (MLP). A Multi-layer Perceptron (MLP) is a sort of feedforward artificial neural network that has at least three layers: the input layer, one or more hidden layers, and the output layer. Except for those in the input layer, all nodes make use of a nonlinear activation function. These MLPs are frequently used in supervised learning tasks as well as a wide range of scientific and engineering problems [31].

AdaBoost. AdaBoost, an acronym for "Adaptive Boosting," is a technique in ensemble learning mainly designed to enhance the efficacy of classification models, though it can be modified for regression problems as well. Created by Yoav Freund and Robert Schapire in 1995, AdaBoost's fundamental idea is to train a series of weak learners—models that perform just slightly better than random chance, like small decision trees—on varying sets of the data. These weak learners are then amalgamated to form a single, more robust model [32]. For tasks like odor classification, where data can be imbalanced or particular classes can be more difficult to distinguish, AdaBoost could be a useful tool. It can focus more on the difficult-to-classify instances, thus improving the overall model's performance. However, care should be taken if the dataset has a lot of noisy data or outliers, as AdaBoost might overemphasize these instances [33].

K-Nearest Neighbors (KNN). K-Nearest Neighbors (KNN) is a straightforward but powerful method utilized for both classification and regression in machine learning. Unlike model-based approaches, KNN is an instance-based learning technique, which means it relies on the complete dataset for decision-making. In the context of odor classification, KNN can transform sensor data into feature vectors and employ an appropriate distance metric to identify the K closest odors in that feature space. This makes KNN especially valuable when the association between sensor measurements and types of odors is intricate and nonlinear, as the algorithm doesn't assume any specific data distribution [34].

Random forest. Random Forests is an ensemble learning method used for classification, regression, and other tasks. It operates by constructing a multitude of decision trees during training and outputs the class that is the mode (for classification) or the mean (for regression) of the classes or values output by individual trees. Random Forests was developed by Leo Breiman and Adele Cutler, building on earlier work on bagging and the random subspace method [27]. For odor classification tasks, Random Forests can be highly effective due to their ability to model complex relationships in the data. Sensor readings from an electronic nose can be used as features for the Random Forest model. Given the robustness of Random Forests to overfitting and their ability to capture feature importance, they can offer significant advantages in discerning subtle differences between various types of odors [35].

Support vector machine. A supervised learning approach called Support Vector Machine (SVM) is employed for classification and regression problems. It aims to find the hyperplane that best divides a dataset into classes. The term "support vector" refers to the data points that are closest to the hyperplane, and these points essentially "support" or define the separating line or hyperplane. SVM was developed by Vladimir Vapnik and his colleagues in the 1990s [36]. SVM can be applied to odor classification tasks where high-dimensional sensor data is involved. The algorithm can effectively separate complex odor classes using hyperplanes in a high-dimensional space. This makes it well-suited for applications where sensor readings from electronic noses or similar devices are used to identify different types of odors [37].

Gradient Boosting Decision Tree (GBDT). Gradient Boosting Decision Trees (GBDT) is an ensemble learning method that aims to build a strong predictive model in the form of an ensemble of weak predictive models, typically decision trees. It's particularly useful for regression and classification problems, as well as ranking tasks. The concept was popularized by statisticians like Leo Breiman and Jerome Friedman [38]. Moreover, GBDT often yields high-accuracy models and is one of the best off-the-shelf algorithms for a wide range of machine learning tasks [39]. In the context of odor classification, GBDT can be a powerful tool for creating highly accurate models based on sensor data from electronic noses or other sensory equipment [40]. Let $\{x_i, y_i\}_i^m = 1$ represent the sample data, where $x_i = (x_{1i}, x_{2i}, \ldots, x_{ri})$ represent the features, and y_i indicates the class label. The detailed steps involved in the Gradient Boosted Decision Trees (GBDT) process are outlined below [41], as:

The starting constant value γ is determined as follows:

$$F_0(x) = argmin\left(-\gamma \sum_{i=1}^{m} L(y_i, \gamma)\right) \tag{1}$$

Where $L(y_i, \gamma)$ represents the loss function. Subsequently, the residual in the direction of the gradient is expressed as:

$$\hat{y}_i = -\left[\frac{\partial_L(y_i, F(x_i))}{\partial F(x_i)}\right]_{(f(x)=f_{(n-1)}(x))} \tag{2}$$

where n represents the number of iterations, and $n = 1, 2, \ldots, N$. The original prototype, denoted by $T(x_i, \alpha_n)$, is derived by fitting it to the sample data. The parameter α_n is computed using the least squares method, as follows:

$$\alpha_n = argmin_{\alpha,\beta} \sum_{i=1}^{m} \left(\hat{y}_i - \beta T\left(x_i; \alpha\right)\right)^2 \tag{3}$$

By optimizing the loss function, the weight of the present version is formulated as:

$$\gamma_n = argmin_{\gamma} \sum_{i=1}^{m} L\left(y_i, F_{n-1}\left(x\right) + \gamma T\left(x_i; \alpha_n\right)\right) \tag{4}$$

The model is then updated iteratively until the conditions for convergence are satisfied:

$$F_n\left(x\right) = F_{n-1}\left(x\right) + \gamma_n T\left(x_i; \alpha_n\right) \tag{5}$$

4.4. XGBoost algorithm

The XGBoost algorithm, developed by Chen and Guestrin [42], builds upon the framework of Gradient Boosted Decision Trees (GBDT). It has gained considerable attention for its exceptional performance in machine learning competitions on platforms like Kaggle [43]. Unlike GBDT, XGBoost incorporates a regularization term into its objective function to mitigate the risk of overfitting. This is how the objective function is described:

$$O = \sum_{i=1}^{n} L\left(y_i, F\left(x_i\right)\right) + \sum_{k=1}^{t} R\left(f_k\right) + C \tag{6}$$

In this equation, $R\left(f_k\right)$ represents the regression term at k^{th} iteration and C is a constant that may be optionally excluded. The $R\left(f_k\right)$ signifies the regularization component at the k^{th} iteration:

$$R\left(f_k\right) = \alpha H + \frac{1}{2}\vartheta \sum_{j=1}^{H} w_j^2 \tag{7}$$

In this explanation, $R\left(f_k\right)$ embodies the regularization term and is influenced by several elements. H is the total number of leaves, ϑ is the penalty variable, and w_j is the output associated with each individual leaf node. Specifically, the leaves serve as the predicted categories based on the established classification rules, while a leaf node refers to a terminal node in the tree that isn't subject to further splitting. In contrast to GBDT, which approximates its objective function using a first-order derivative, XGBoost uses a second-order Taylor series. When using mean square error as the loss function, the goal can be expressed as:

$$O = \sum_{i=1}^{n} \left[p_i w_{q(x_i)} + \frac{1}{2}\left(q_i w_{q(x_i)}^2 \right) \right] + \alpha H + \frac{1}{2}\vartheta \sum_{j=1}^{H} w_j^2 \tag{8}$$

In this setup, $q\left(x_i\right)$ serves as a function that maps data points to their respective leaves. The overall loss is computed by summing up the individual loss values. Because data samples are mapped to decision tree structure (DT), the total loss can be ascertained by adding together the loss scores associated with each leaf node. Consequently, the objective function can also be articulated as a summation of the individual loss values for each leaf node in the decision tree as:

$$O = \sum_{j=1}^{T} \left[p_i w_j + \frac{1}{2}\left(Q_i + \vartheta\right) w_j^2 \right] + \alpha H \tag{9}$$

Where $P_j = \sum_{i \in I_j} p_i$, $Q_j = \sum_{i \in I_j} q_i$, and I_j represents all samples located in leaf node j. In summary, optimizing the objective function is essentially transformed into the task of finding the minimum point of a quadratic function. Furthermore, the inclusion of a regularization term enhances XGBoost's capability to resist overfitting.

5. Performance evaluation

Every machine learning model that was chosen was trained in order to assess how well it performs based on five essential evaluation measures: recall, precision, F1 score, Kappa, and Hamming Loss. The calculation of precision is determined by taking the number of True Positives (TP) and dividing it by the combined count of True Positives and False Positives (FP).

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

Additionally, the Recall is calculated to gauge the model's effectiveness in correctly identifying positive samples. This was determined by dividing the number of True Positives (TP) by the combined total of True Positives and False Negatives (FN). This proportion represents the correctly identified positive cases in relation to all actual positive cases.

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

For a comprehensive assessment of a model's effectiveness, it's crucial to take into account both precision and recall. To achieve this, the F1 score is calculated which is the harmonic mean of both recall and precision, providing a well-rounded insight into the algorithm's behavior:

$$F_1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{12}$$

Additionally, the Kappa statistic is a vital measure, notably beneficial for gauging model efficacy on datasets with class imbalances. This metric offers a numerical representation of a classifier's accuracy. While the theoretical value of Kappa lies between -1 and 1, for the purposes of this study, it has been normalized to a range between 0 and 1. A greater Kappa score suggests superior performance in classification tasks. Assuming N denotes the total number of predictions made, the formula for calculating Kappa can be found in Equation (13)

$$Kappa = \frac{P_0 - P_e}{1 - P_e} \tag{13}$$

where,

$$P_0 = \frac{\sum_{i=1}^{4} TP_i}{N} \tag{14}$$

and

$$P_e = \frac{\sum_{i=1}^{4} \left[\left(TP_i + FP_i\right) \times \left(TP_i + FN_i\right) \right]}{N \times N} \tag{15}$$

Furthermore, Hamming loss [44] serves as an alternative metric useful for evaluating the effectiveness of models designed for multi-class classification. Hamming loss reflects the rate of misclassified labels. A zero value means 100% accuracy. In the equation for this metric, N denotes the total number of predicted samples, and M denotes the number of unique class labels. The mathematical expression for Hamming Loss is detailed in Equation (16).

$$Hamming\ loss\left(y, \hat{y}\right) = \frac{1}{N \times M} \sum_{i=1}^{N} \sum_{j=1}^{M} XOR\left(y_{i,j}, \hat{y}_{i,j}\right) \tag{16}$$

In this context, y and $\hat{y}$ represent the actual class label and the predicted class label, respectively.

6. Results and validation

In the development phase of the proposed oil recognition system, this study conducted two distinct experiments to thoroughly evaluate the performance of the models: one experiment using pure oil samples and another using mixed (impure) oil samples. This dual approach allowed for a comprehensive analysis of how the models perform under different real-world scenarios.

The first experiment involved testing seven distinct machine learning algorithms with pure oil samples of sesame, black seed, flaxseed, and almond oils. Key performance indicators, including accuracy, F1 score, Kappa statistic, and Hamming Loss, were assessed and reported for each model to determine their effectiveness in distinguishing between these similar oils.

The second experiment extended the dataset by incorporating mixed oil samples with varying purity levels (90%, 80%, and 70%). These mixed samples aimed to evaluate the models' ability to accurately classify oils when impurities were present, simulating practical conditions where oils may not be entirely pure.

For both experiments, the dataset was divided into three subsets for training, validation, and testing:

- **Training Subset:** The largest subset, constituting 80% of the total data (including both pure and mixed samples), was used for training the models.

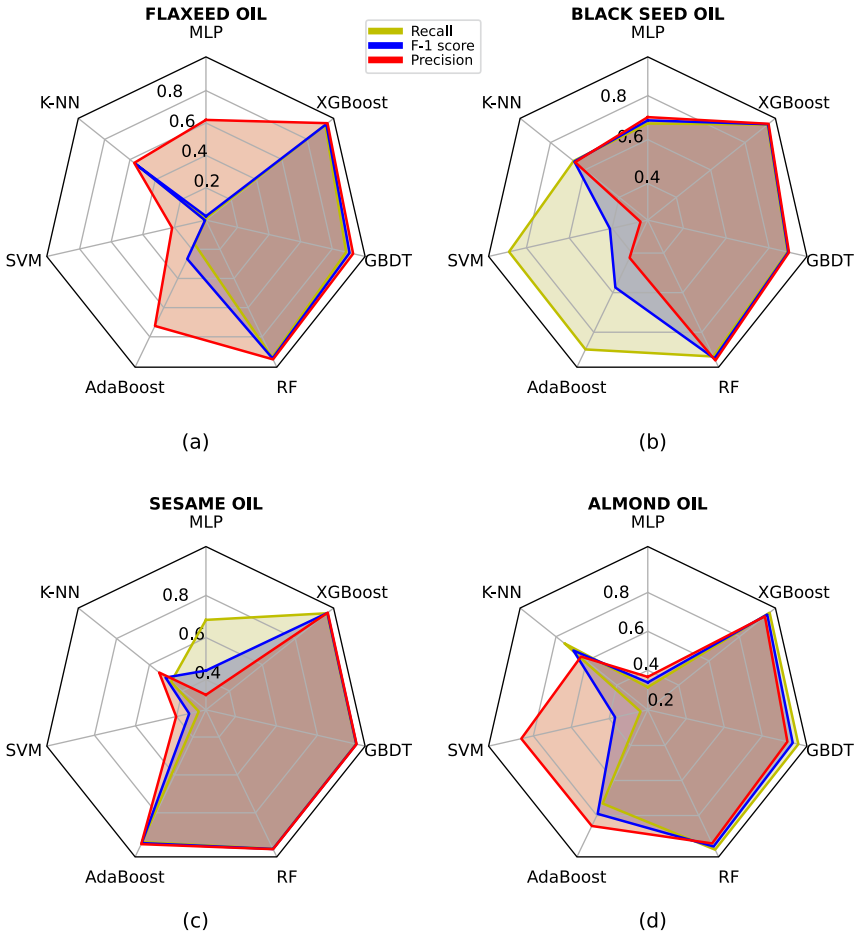


Fig. 11. Performance indicators including recall, F1 score, and precision are used to assess how well machine learning classifiers perform for the four types of oils, (a) flaxseed, (b) black seed, (c) sesame, and (d) almond oils.

- **Validation Subset:** 10% of the data was reserved for model validation, allowing fine-tuning of the models' parameters based on both pure and impure sample scenarios.
- **Testing Subset:** The remaining 10% of the data, including both pure and mixed oil samples, was used to evaluate the models' performance and their generalization ability in distinguishing between pure and impure oils.

The results from both experiments were compared to identify the most effective model. In the pure oil experiment, tree-based ensemble methods, particularly XGBoost, demonstrated high accuracy and consistency. In the mixed oil experiment, the XGBoost model continued to outperform others, though with a slight decrease in accuracy as the level of impurity increased. This comprehensive evaluation confirmed the robustness of the models, providing valuable insights into their applicability and reliability under varying conditions.

6.1. ML model performance

When choosing the ML model, the ultimate criterion is the quality of the model's output i.e., the one that yields superior performance. Evaluating a model's efficacy in this specific scenario necessitates a comprehensive set of metrics, such as accuracy, precision, recall, and the F1 score. All these metrics have been computed for the machine learning models deployed to classify the various vegetable oils, which include almond, black seed, flaxseed, and sesame oils. The results from the seven tested machine learning algorithms are visually represented in a radar chart, as shown in Fig. 11. This chart illustrates significant fluctuations in accuracy levels across the different machine-learning models. It's important to note that the dataset is inherently complex due to the selection of four types of oils that are visually similar in color and nearly indistinguishable in odor. In evaluating the performance of machine learning algorithms across four distinct oil types, it becomes evident that the effectiveness of these models varies significantly. The Support Vector Machines (SVM) algorithm, for instance, showcased notably poor performance, achieving an accuracy rate of only 33%. Its F1 scores, which gauge the balance between precision and recall, exhibited considerable variation—ranging from a mere 1% to 41% across different oil types. This inconsistency suggests potential challenges in SVM's predictive capability, which might stem from its sensitivity to kernel selection or feature space scale. Closely following in terms of underperformance, the Multi-Layer Perceptron (MLP) recorded an accuracy of 41%, with F1 scores fluctuating between 2% and 68%. The variability in F1 scores across oil categories highlights the complexities inherent in classifying nuanced datasets, indicating a pressing need for further refinement of these models or the exploration of alternative algorithms that might better navigate such variability. Conversely, the K-NN and AdaBoost models yielded more promising results, with average accuracy rates of 60% and 67%, respectively. Moreover, models like Random Forests and Gradient Boosting Decision Trees (GBDT) stood out for their exceptional performance, boasting average accuracies of approximately 96% and 94%, respectively. The standout performer, however, was Extreme Gradient Boosting (XGBoost), which not only achieved the highest accuracy at 97% but also maintained remarkable consistency in F1 scores, ranging between 93% and 98%. These outcomes underscore the effectiveness of ensemble methods and tree-based models in handling the complexities of oil-type classification. They demonstrate the potential of such models to achieve high predictive accuracy and reliability, especially in tasks

Table 2
Evaluation results of vegetable oils' recognition efficacy for various machine learning algorithms (Best values are in bold letters).

ML Model	Accuracy	F1 score	Recall	Precision	Kappa	Hamming loss
MLP	0.4166	0.2677	0.3769	0.2717	0.1507	0.6471
AdaBoost	0.6788	0.6476	0.6839	0.7438	0.5759	0.3212
k-NN	0.6002	0.6020	0.6074	0.6000	0.4669	0.3998
RF	0.9633	0.9619	0.9621	0.9618	0.9510	0.0367
SVM	0.3331	0.2807	0.3526	0.4367	0.1295	0.6669
GBDT	0.9400	0.9375	0.9383	0.9373	0.9199	0.0600
XGBoost	**0.9670**	**0.9656**	**0.9659**	**0.9654**	**0.9560**	**0.0330**

Table 3
Performance Metrics for ML Models with Mixed Oil Samples.

ML Model	Purity Level	Accuracy	F1 Score	Kappa	Hamming Loss
XGBoost	90% Pure	94-95%	91-94%	0.93-0.94	0.05
	80% Pure	90-92%	87-91%	0.89-0.91	0.08
	70% Pure	85-87%	82-86%	0.84-0.86	0.12
Random Forest (RF)	90% Pure	91-93%	89-92%	0.91	0.06
	80% Pure	88-90%	84-89%	0.88-0.89	0.09
	70% Pure	82-85%	78-83%	0.80-0.82	0.14
Gradient Boosting Decision Trees (GBDT)	90% Pure	89-91%	87-90%	0.88	0.07
	80% Pure	85-88%	82-86%	0.85	0.11
	70% Pure	80-83%	76-81%	0.78	0.15

characterized by complex and nuanced data. Notably, the success of Random Forests and GBDT can be attributed to their robustness against overfitting and their capacity to model complex, nonlinear relationships. XGBoost, in particular, distinguishes itself through its advanced regularization techniques and efficient handling of sparse data, making it exceedingly effective for this classification challenge. The algorithm's flexibility in model tuning and its ability to incrementally improve weak learners underscore its superiority, positioning XGBoost as the most reliable choice among the evaluated models. These findings highlight the varying adaptability and efficiency of machine learning algorithms in classifying complex datasets, such as different types of oils. The stark contrasts in performance further emphasize the importance of meticulous model selection and parameter optimization, tailored to the unique characteristics of the data at hand.

The final evaluation metrics for each oil type in this study are meticulously presented in Table 2. This table encompasses crucial performance indicators, including accuracy, and macro averages of F1 score, precision, and recall, as well as the Kappa coefficient and Hamming loss. Out of all the machine learning models that were rigorously tested, the XGBoost method stood out as the most effective. This model not only achieved a remarkable accuracy rate of 97% but also boasted an F1 score ranging between 93% and 98%, making it highly reliable for both precision and recall. In terms of Hamming Loss, which measures the fraction of incorrect classifications, the XGBoost model recorded a low score of 0.041, suggesting a high rate of accurate label predictions. Furthermore, the Kappa statistic, a measure that is especially useful for gauging model performance on unbalanced datasets, was also notably high at 0.97 for the XGBoost algorithm. Given these robust and compelling performance metrics, it becomes evident that the XGBoost algorithm outperforms its counterparts. Therefore, it has been selected as the most effective and reliable model for the classification tasks involved in this particular research endeavor.

XGBoost has more likely outperformed due to the inherent capabilities such as capturing non-linear relationships, implicit feature selection by prioritizing the most informative features, and built-in regularization.

6.2. Performance evaluation with mixed oil samples

To comprehensively evaluate the robustness and adaptability of the proposed oil recognition system, two distinct experiments were conducted: one using pure oil samples and the other using mixed oil samples with varying purity levels (90%, 80%, and 70%). The results for the mixed oil samples across different machine learning models are summarized in Table 3.

6.2.1. Results and analysis

1. **90% Pure Samples:** At 90% purity, the XGBoost model demonstrated its robustness with an accuracy rate of 94-95%, a minor decrease from its performance with pure samples. The F1 score ranged between 91-94%, reflecting the model's consistency in maintaining precision and recall even when slight variations were present. The Kappa coefficient was high, between 0.93-0.94, indicating a strong agreement between predicted and actual labels. The Hamming Loss remained low at 0.05, suggesting a minimal impact on the system's reliability when the oil purity is only slightly compromised. The Random Forest (RF) and Gradient Boosting Decision Tree (GBDT) models also performed well but with slightly lower metrics than XGBoost. RF achieved an accuracy range of 91-93%, while GBDT recorded 89-91%. Their F1 scores remained consistent, indicating that tree-based ensemble methods are suitable for managing small impurities in the samples. These results confirm that PIORS, utilizing these models, can reliably classify oils even when minor impurities are present.

2. **80% Pure Samples:** As the oil purity dropped to 80%, the models exhibited a more noticeable decline in performance. The XGBoost model maintained an accuracy rate between 90-92%, with F1 scores ranging from 87-91%. The Kappa coefficient decreased slightly to 0.89-0.91, demonstrating the model's adaptability but also indicating the challenges posed by increased impurity levels. The Hamming Loss increased to 0.08, reflecting a moderate rise in the fraction of misclassified samples. RF and GBDT models followed this trend, with RF's accuracy dropping to 88-90% and GBDT's to 85-88%. Their F1 scores showed variability, with RF ranging from 84-89% and GBDT from 82-86%. These results suggest that while the models remain effective, their performance begins to degrade more significantly as impurity levels increase.

3. **70% Pure Samples:** At 70% purity, the models encountered significant challenges. The XGBoost model's accuracy declined to 85-87%, and its F1 score dropped to 82-86%. The Kappa coefficient further reduced to 0.84-0.86, indicating that, although XGBoost still outperformed the other models, its ability to accurately classify the oils was

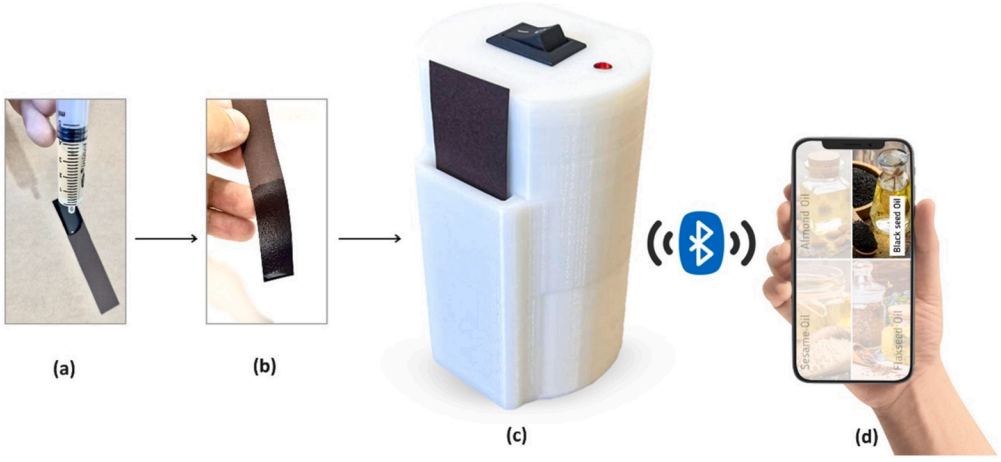


Fig. 12. AI-powered oil recognition prototype: (a) dropping oil on tested paper. (b) tested paper is saturated with oil before testing. (c) placing the tested paper within the testing slot (d) display the results via a mobile app.

impacted by the higher level of impurity. The Hamming Loss value rose to 0.12, showing an increased rate of incorrect classifications. The RF and GBDT models experienced even greater performance reductions, with RF achieving 82-85% accuracy and GBDT reaching 80-83%. Their F1 scores reflected these challenges, ranging between 78-83% for RF and 76-81% for GBDT. These findings highlight that while PIORS can still classify oils at 70% purity, the effectiveness of the system is considerably diminished when dealing with such high levels of impurity.

The results from the mixed oil sample tests reveal that XGBoost remains the most reliable and accurate model across varying purity levels. However, the system's performance does decrease as impurity levels increase, indicating the challenges associated with distinguishing oils that are not entirely pure. Ensemble methods like RF and GBDT also follow this trend, but with a more significant impact, demonstrating that while they are effective with minor impurities, their reliability diminishes more rapidly as purity levels decrease. These insights are crucial for further development of PIORS, as they indicate the need for continued model refinement and optimization to enhance the system's robustness in handling various levels of oil purity. By incorporating models that adapt well to such variations, PIORS can achieve a higher level of accuracy and reliability in real-world scenarios, ensuring its effectiveness for food quality assessment and safety applications.

6.3. Model deployment with embedded microcontroller

After a comprehensive evaluation of multiple machine learning algorithms to pinpoint the most proficient model, the next key phase involves deploying the XGBoost model into a real-world production setting. The best-performing model is then deployed in the testbed using a microcontroller, and the programming language of choice for this environment is C++. There are three primary methods for adapting the model for C++ implementation including OpenCV, TensorFlow, or ONNX Runtime [45]. Among these, TensorFlow offers more features. It offers a C++ API to facilitate seamless integration into microcontroller-based applications [46]. This API is constructed using a set of low-level programming paradigms, including both class structures and functions native to C++. Once the machine learning model is successfully converted through this API to get the C++ source files. These source files are not just transpositions of the original model but are optimized to run efficiently on microcontroller units (MCUs). This process effectively streamlines the transition from a theoretical model to a deployable application, thereby enabling the introduction of the machine-learning solution into real-world scenarios. The availability of these C++ source files marks a crucial step towards operationalizing MCU applications, facilitating both immediate and future deployments of the machine learning model across various applications.

6.4. Prototype performance

Within the framework of this research endeavor, a cutting-edge prototype has been meticulously designed and fabricated to cater to the increasingly complex demands of oil identification and classification. Utilizing advanced 3D printing techniques, a multi-dimensional enclosure was crafted to securely house the pivotal components of the system, notably the sensor-laden hardware testing board and an ESP32 microcontroller. This integrated setup is energized by a lithium battery, a layout explicitly detailed in Fig. 12. The operational procedure for oil sample evaluation has been rigorously delineated to ensure accurate and reliable results. Initial steps involve depositing oil droplets onto a specifically engineered absorbent medium, as visualized in Fig. 12. (a). This absorbent paper (Blulu-Tissues Oil -67) was selected after comprehensive testing for its unparalleled ability to imbibe oil, surpassing the performance of other commercially available alternatives. Once the paper is saturated with the oil sample as shown in Fig. 12. (b), it is carefully placed within the device's designated testing chamber, as captured in Fig. 12. (c). An additional salient feature of this prototype is the integration of a built-in Bluetooth module within the ESP32 microcontroller. This integration facilitates seamless communication with mobile applications, thereby enabling real-time display of test results as shown in Fig. 12. (d). Moreover, the system is characterized by its remarkable energy efficiency, enabled by a 400-mA lithium battery that permits uninterrupted 24-hour operation. This efficiency is further exemplified by the device's rapid predictive analytics capabilities, delivering classification results within an impressive five-second timeframe.

7. Conclusion

In conclusion, this study introduces the groundbreaking Portable Intelligent Oil Recognition System (PIORS), marking a significant advancement in the agri-food technology sector. As the first device of its kind, PIORS revolutionizes oil quality assessment by combining portability, rapid analysis, and AI-driven precision. This innovative system is capable of identifying oil types within an unprecedented five-second timeframe, setting a new benchmark for efficiency and convenience. At the core of PIORS is a sophisticated array of machine learning models, with XGBoost demonstrating remarkable accuracy, achieving a total accuracy rate of 97%, and F1 scores ranging between 93% and 98%. This integration of advanced machine learning not only bolsters the system's reliability but also transforms quality control processes, offering real-time, actionable insights. Furthermore, PIORS's user-friendly design, featuring seamless Bluetooth connectivity and real-time synchronization with mobile applications, ensures that users have immediate access to critical data. The introduction of PIORS represents a paradigm shift

in the agri-food industry, highlighting the transformative potential of AI-powered solutions in enhancing food safety, environmental monitoring, and quality assurance practices. As a trailblazer in oil recognition technology, PIORS is poised to redefine industry standards and catalyze further innovation in the field.

CRediT authorship contribution statement

Montaser N.A. Ramadan: Writing – original draft, Validation, Software, Methodology, Conceptualization, Formal analysis. **Mohammed A.H. Ali:** Writing – review & editing. **Shin Yee Khoo:** Writing – review & editing. **Layth Hamad:** Writing – review & editing. **Mohammad Alkhedher:** Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Montaser Ramadan reports writing assistance was provided by University of Malaya. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

ADU author acknowledges financial support from Abu Dhabi University's Office of Research and Sponsored Programs, Project No. 19300889.

Data availability

The authors do not have permission to share data.

References

[1] F.D. Gunstone, Vegetable oils in food technology: composition, properties, and uses, https://api.semanticscholar.org/CorpusID:109100776, 2002.
[2] M. Lees, J.-F. Morin, FoodIntegrity Handbook: A Guide to Food Authenticity Issues and Analytical Solutions, Eurofins Analytics France, 2018.
[3] Y. Aghoutane, M. Brebu, M. Moufid, R. Ionescu, B. Bouchikhi, N. El Bari, Detection of counterfeit perfumes by using gc-ms technique and electronic nose system combined with chemometric tools, Micromachines 14 (3) (2023) 524.
[4] M.F.S. Mota, H.D. Waktola, Y. Nolvachai, P.J. Marriott, Gas chromatography–mass spectrometry for characterisation, assessment of quality and authentication of seed and vegetable oils, TrAC, Trends Anal. Chem. 138 (2021) 116238.
[5] M. Beneito-Cambra, D. Moreno-González, J.F. García-Reyes, M. Bouza, B. Gilbert-López, A. Molina-Díaz, Direct analysis of olive oil and other vegetable oils by mass spectrometry: a review, TrAC, Trends Anal. Chem. 132 (2020) 116046.
[6] A.D. Wilson, M. Baietto, Applications and advances in electronic-nose technologies, Sensors 9 (7) (2009) 5099–5148.
[7] A. Dey, Semiconductor metal oxide gas sensors: a review, Mater. Sci. Eng. B 229 (2018) 206–217.
[8] I.H. Sarker, Machine learning: algorithms, real-world applications and research directions, SN Comput. Sci. 2 (3) (2021) 160.
[9] A. Cataldo, E. Piuzzi, G. Cannazza, E. De Benedetto, L. Tarricone, Quality and anti-adulteration control of vegetable oils through microwave dielectric spectroscopy, Measurement 43 (8) (2010) 1031–1039.
[10] A.D. Wilson, M. Baietto, Applications and advances in electronic-nose technologies, Sensors 9 (7) (2009) 5099–5148.
[11] S. Freddi, L. Sangaletti, Trends in the development of electronic noses based on carbon nanotubes chemiresistors for breathomics, Nanomaterials 12 (17) (2022) 2992.
[12] J. Jońca, M. Pawnuk, A. Arsen, I. Sówka, Electronic noses and their applications for sensory and analytical measurements in the waste management plants—a review, Sensors 22 (4) (2022) 1510.
[13] B. Plutowska, W. Wardencki, Application of gas chromatography–olfactometry (gc–o) in analysis and quality assessment of alcoholic beverages–a review, Food Chem. 107 (1) (2008) 449–463.
[14] M. Moufid, C. Tiebe, N. El Bari, D.A.H. Fakra, M. Bartholmai, B. Bouchikhi, Pollution parameters evaluation of wastewater collected at different treatment stages from wastewater treatment plant based on e-nose and e-tongue systems combined with chemometric techniques, Chemom. Intell. Lab. Syst. 227 (2022) 104593.
[15] T.E. Burch, C.W. Hanson III, E.R. Thaler, Detection, diagnosis, and monitoring of a medical condition or disease with artificial olfactometry, US Patent 7,255,677, Aug. 14 2007.
[16] M. Ghasemi-Varnamkhasti, M. Aghbashlo, Electronic nose and electronic mucosa as innovative instruments for real-time monitoring of food dryers, Trends Food Sci. Technol. 38 (2) (2014) 158–166.
[17] R. Mariotti, E. Núñez-Carmona, D. Genzardi, S. Pandolfi, V. Sberveglieri, S. Mousavi, Volatile olfactory profiles of umbrian extra virgin olive oils and their discrimination through mox chemical sensors, Sensors 22 (19) (2022), https://doi.org/10.3390/s22197164, https://www.mdpi.com/1424-8220/22/19/7164.
[18] D. Rusak, L. Brown, S. Martin, Classification of vegetable oils by principal component analysis of ftir spectra, J. Chem. Educ. 80 (05 2003), https://doi.org/10.1021/ed080p541.
[19] Y.G. Mbesse Kongbonga, H. Ghalila, M.B. Onana, Z. Ben Lakhdar, Classification of vegetable oils based on their concentration of saturated fatty acids using laser induced breakdown spectroscopy (libs), Food Chem. 147 (2014) 327–331, https://doi.org/10.1016/j.foodchem.2013.09.145, https://www.sciencedirect.com/science/article/pii/S0308814613014106.
[20] D.B. Dahlberg, S.M. Lee, S.J. Wenger, J.A. Vargo, Classification of vegetable oils by ft-ir, Appl. Spectrosc. 51 (8) (1997) 1118–1124, https://doi.org/10.1366/0003702971941935.
[21] C.E. Tanajura da Silva, V.L. Filardi, I.M. Pepe, M.A. Chaves, C.M.S. Santos, Classification of food vegetable oils by fluorimetry and artificial neural networks, Food Control 47 (2015) 86–91, https://doi.org/10.1016/j.foodcont.2014.06.030, https://www.sciencedirect.com/science/article/pii/S0956713514003594.
[22] L. Zhang, P. Li, X. Sun, X. Wang, W. Zhang, B. Xu, X. Wang, F. Ma, Q. Zhang, X. Ding, Classification and adulteration detection of vegetable oils based on fatty acid profiles, J. Agric. Food Chem. 62 (07 2014), https://doi.org/10.1021/jf501097c.
[23] R. Al-Ruzouq, M.B.A. Gibril, A. Shanableh, A. Kais, O. Hamed, S. Al-Mansoori, M.A. Khalil, Sensors, features, and machine learning for oil spill detection and monitoring: a review, Remote Sens. 12 (20) (2020) 3338.
[24] W.J. Pranoto, S.G. Al-Shawi, P. Chetthamrongchai, T.-C. Chen, E. Petukhova, N. Nikolaeva, W.K. Abdelbasset, N.A. Yushchenko, S. Aravindhan, Employing artificial neural networks and fluorescence spectrum for food vegetable oils identification, Food Sci. Technol. 42 (2021) e80921.
[25] A. Manoharan, K. Begam, V.R. Aparow, D. Sooriamoorthy, Artificial neural networks, gradient boosting and support vector machines for electric vehicle battery state estimation: a review, J. Energy Storage 55 (2022) 105384.
[26] M.N. Ramadan, M. Alkhedher, B.T. Akgün, S. Alp, Portable ai-powered spice recognition system using an enose based on metal oxide gas sensors, in: 2023 International Conference on Smart Applications, Communications and Networking (SmartNets), IEEE, 2023, pp. 1–6.
[27] D. Hercog, T. Lerher, M. Truntič, O. Težak, Design and implementation of esp32-based iot devices, Sensors 23 (15) (2023) 6739.
[28] B. Sensortec, Gas sensor bme688, https://www.bosch-sensortec.com/products/environmental-sensors/gassensors/bme688, 2021.
[29] V. Chaturvedi, Data communication using mqtt with bme688 sensor, 2023.
[30] M.F.R. Al-Okby, T. Roddelkopf, H. Fleischer, K. Thurow, Evaluating a novel gas sensor for ambient monitoring in automated life science laboratories, Sensors 22 (21) (2022) 8161.
[31] A.H. Fath, F. Madanifar, M. Abbasi, Implementation of multilayer perceptron (mlp) and radial basis function (rbf) neural networks to predict solution gas-oil ratio of crude oil systems, Petroleum 6 (1) (2020) 80–91.
[32] S. Mathanker, P. Weckler, T. Bowser, N. Wang, N. Maness, Adaboost classifiers for pecan defect classification, Comput. Electron. Agric. 77 (1) (2011) 60–68.
[33] R.E. Schapire, Explaining adaboost, in: Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik, Springer, 2013, pp. 37–52.
[34] N. Ali, D. Neagu, P. Trundle, Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets, SN Appl. Sci. 1 (2019) 1–15.
[35] A. Parmar, R. Katariya, V. Patel, A review on random forest: an ensemble classifier, in: International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI), 2018, Springer, 2019, pp. 758–763.
[36] E. Mayoraz, E. Alpaydin, Support vector machines for multi-class classification, in: International Work-Conference on Artificial Neural Networks, Springer, 1999, pp. 833–842.
[37] E. Osuna, R. Freund, F. Girosi, An improved training algorithm for support vector machines, in: Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop, IEEE, 1997, pp. 276–285.
[38] L. Jinshu, W. Yijiang, W. Ganjun, P. Xiaoasheng, L. Taiwei, J. Yuhang, Gradient boosting decision tree and random forest based partial discharge pattern recognition of hv cable, in: 2018 China International Conference on Electricity Distribution (CICED), IEEE, 2018, pp. 327–331.
[39] Z. Zhang, C. Jung, Gbdt-mo: gradient-boosted decision trees for multiple outputs, IEEE Trans. Neural Netw. Learn. Syst. 32 (7) (2020) 3156–3167.
[40] X. Yu, Z. Xu, X. Zhou, J. Zheng, Y. Xia, L. Lin, S.-H. Fang, Load forecasting based on smart meter data and gradient boosting decision tree, in: 2019 Chinese Automation Congress (CAC), IEEE, 2019, pp. 4438–4442.
[41] Y. Fan, W. Lei, Wind speed prediction based on gradient boosting decision tree, in: 2022 International Conference on Big Data, Information and Computer Network (BDICN), IEEE, 2022, pp. 93–97.
[42] J. Song, L. Jin, Y. Xie, C. Wei, Optimized xgboost based sparrow search algorithm for short-term load forecasting, in: 2021 IEEE International Conference on Computer Science, Artificial Intelligence and Electronic Engineering (CSAIEE), IEEE, 2021, pp. 213–217.

[43] W. Liang, S. Luo, G. Zhao, H. Wu, Predicting hard rock pillar stability using gbdt, xgboost, and lightgbm algorithms, Mathematics 8 (5) (2020) 765.
[44] T. Merembayev, D. Kurmangaliyev, B. Bekbauov, Y. Amanbek, A comparison of machine learning algorithms in predicting lithofacies: case studies from Norway and Kazakhstan, Energies 14 (7) (2021), https://doi.org/10.3390/en14071896, https://www.mdpi.com/1996-1073/14/7/1896.
[45] N. Schizas, A. Karras, C. Karras, S. Sioutas, Tinyml for ultra-low power ai and large scale iot deployments: a systematic review, Future Internet 14 (12) (2022) 363.
[46] C.A. Estrebou, M. Fleming, M.D. Saavedra, F. Adra, A.E. De Giusti, Lightweight convolutional neural networks framework for really small tinyml devices, in: International Conference on Smart Technologies, Systems and Applications, Springer, 2021, pp. 3–16.