



Article An Integrated Handheld Electronic Nose for Identifying Liquid Volatile Chemicals Using Improved Gradient-Boosting Decision Tree Methods

Mengli Cao and Xiong Hu *

Logistic Engineering College, Shanghai Maritime University; Shanghai 201306, China * Correspondence: huxiong@shmtu.edu.cn

Abstract: The main ingredients of various odorous products are liquid volatile chemicals (LVC). In human society, identifying the type of LVC is the inner logic of many applications, such as exposing counterfeit products, grading food quality, diagnosing interior environments, and so on. The electronic nose (EN) can serve as a cost-effective, time-efficient, and safe solution to LVC identification. In this paper, we present the design and evaluation of an integrated handheld EN, namely SMUENOSEv2, which employs the NVIDIA Jetson Nano module for running the LVC identification method. All components of SMUENOSEv2 are enclosed in a handheld case. This all-in-one structure makes it convenient to use SMUENOSEv2 for quick on-site LVC identification. To evaluate the performance of SMUENOSEv2, two common odorous products, i.e., perfumes and liquors, were used as the samples to be identified. After sampling data preprocessing and feature generation, two improved gradient-boosting decision tree (GBDT) methods were used for feature classification. Extensive experimental results show that SMUENOSEv2 is capable of identifying LVC with considerably high accuracies. With previously trained GBDT models, the time spent for identifying the LVC type is less than 1 s.

Keywords: liquid volatile chemicals; electronic nose; handheld device; gradient-boosting decision tree; feature generation

1. Introduction

Liquid volatile chemicals (LVC) are chemicals that can be volatilized from the original liquid form to gaseous form by themselves or external blowing. In daily life, the main ingredients of multiple beverages, flavorings, and cosmetics are LVC. For example, we can smell the fragrance when getting close to a woman wearing perfume, since the fragrant constituents of perfume have been volatilized into fragrant gas and carried into our nose. As the electronic counterpart of the human nose, many types of chemical sensors have been developed, including metal oxide semiconductor (MOS) [1], quartz crystal microbalance [2], electrochemical [3], optical [4], catalytic combustion [5], gravimetric [6], and carbon nanotubes [7] sensors. All these types of chemical sensors can be used to detect the existence of LVC by measuring the chemical concentration. However, it is hard to discriminate between different types of LVC or to identify the LVC type solely using an individual chemical sensor. As being defined in [1], a single chemical sensor is an element of a basic electronic circuit that senses the chemical concentration fluctuations and outputs valid values for further processing.

Identifying the LVC type can help in distinguishing counterfeit products [8], protecting environments [9], monitoring food quality [10], and so on. Electronic techniques for LVC identification include gas chromatography (GC) [11], GC combined with mass spectrometry (GC-MS) [12,13], and EN [14]. GC comprises the sample inlet, chromatographic column, thermostat, and detector. The sample inlet and detector are mounted at the start and end of the chromatographic column, respectively. Being driven by the standard carrier gas, which



Citation: Cao, M.; Hu, X. An Integrated Handheld Electronic Nose for Identifying Liquid Volatile Chemicals Using Improved Gradient-Boosting Decision Tree Methods. *Electronics* **2023**, *12*, 79. https://doi.org/10.3390/ electronics12010079

Received: 22 November 2022 Revised: 21 December 2022 Accepted: 22 December 2022 Published: 25 December 2022



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). is commonly preserved in a gas cylinder, the volatized LVC gas is steered from the sample inlet to the detector. Owing to their different velocities through the column, different LVC constituents reach the destination detector at different times, which enables the separation of them. To provide more details about the arriving constituents, GC-MS employs the mass spectrometry as the detector of GC. Although GC and GC-MS can separate and identify most constituents of LVC, they are faced with large space occupancy, low time efficiency, and high-cost problems. The large space occupancy caused by the bulk of the thermostat and gas cylinder especially stagnates the portability and on-site usage of GC and GC-MS.

As an important supplementary technique, ENs can mitigate the above-mentioned problems. Typical EN systems comprise two main sub-systems: the perception and pattern recognition sub-systems [15]. The perception sub-system consists of a sensor array, an interfacing circuit, and a microprocessor. Based on the sensors' raw measurements, the pattern recognition subsystem extracts and processes the corresponding features. According to the integration level, existing EN designs can be tentatively categorized into divide-body and integrated designs. In divide-body ENs [14–26], the perception and pattern recognition works are running on two spatially separated systems, which can exchange data with each other through serial port, network, or trans-flash cards. Commonly, the perception can be realized using an embedded system with ARM micro-processors, while computationally intensive pattern-recognition tasks are conducted on a lab computer or portable laptop. The advantage of divide-body designs is cost-effectiveness, since the pattern recognition work can be running on an old computer bought for common use. However, the construction of communication between perception and pattern recognition sub-systems increases the operation complexity of divide-body designs. Integrated ENs are capable of enclosing all of their components in a spatially integrated case [27–29]. The defining characteristic of integrated designs is that the pattern recognition work is also conducted on an embedded system. The works in [27–29] all employed the Field Programmable Gate Array (FPGA) for pattern recognition, which means that the perception and pattern recognition sub-systems can be integrated together into a compact embedded system. Nevertheless, the employed FPGA processor brings about noteworthy programming difficulties, which hinders the implementation of state-of-the-art machine learning (ML) methods. As an alternative compact embedded system, the NVIDIA Jetson Nano module provides a friendly platform for running Python-based ML methods. However, to our knowledge, the NVIDIA Jetson Nano module has never been used for designing EN.

In this paper, we present the design and evaluation of an integrated handheld EN, which utilizes the NVIDIA Jetson Nano module as the computing kernel. The newly designed EN system, namely, SMUENOSEv2, is compact enough to be held by hand to complete the LVC identification task. SMUENOSEv2 is composed of gas transportation and electronic hardware parts. In the gas transportation part, a small air pump is used for generating fast airflow to accelerate the volatilization of LVC samples in the volatilization pot, and to carry the gaseous LVC towards the sensor array. In the electronic hardware part, a STM32 processor and a NVIDIA Jetson Nano module are used for perception and pattern recognition, respectively. The communication between STM32 and NVIDIA Jetson Nano was fixed inside the case of SMUENOSEv2, which enables a quick and straightforward operation for LVC identification. Owing to the Linux operation system on NVIDIA Jetson Nano, it is convenient to program using the famous C/C++ and Python languages. Moreover, to evaluate the performance of SMUENOSEv2, perfume and liquor were used as two types of LVC samples to be identified. After sampling data preprocessing and feature generation, two improved gradient-boosting decision tree (GBDT) [30] methods, i.e., XGBoost [31] and LightGBM [32], were used for pattern recognition. According to the experimental comparison results in our previous work [14], the advantages in using XGBoost rather than other traditional chemometric methods, such as principal component analysis [33], support vector machine [34], artificial neural network [35], and so on, are higher identification accuracies in terms of mean and variance. For feature generation, a

novel statistics combination is proposed. Extensive real experiments were conducted to verify the applicability of SMUENOSEv2.

Compared with our previous divide-body EN used in the work [14], which does not contain the detail about EN design, the novelties of SMUENOSEv2 mainly comprise the integrated EN design scheme and new gas route. The MOS sensors in SMUENOSEv2 remained the same as those used in the previous version. However, these sensors are readily replaceable with other voltage-output MOS chemical sensors. The reason for selecting MOS sensors is that, according to the authors' experimental experience, MOS sensors have the merit of fast responsiveness. Moreover, compared with other sensor types, there are more cost-effective MOS sensors available in commercial markets. To our knowledge, SMUENOSEv2 is the first integrated EN using the NVIDIA Jetson Nano as the computing kernel, which demonstrates the originality of this paper. The main contribution of this paper is three-fold. First, the presented integrated EN design can provide important guidance for on-site quick LVC identification, e.g., paroxysmal inspection of counterfeit perfumes at the cosmetics counter in shopping arcades. Second, the experimental quantitative comparison results of the two improved GBDT methods are valuable references for further related investigation. Third, the presented statistics combination for feature generation can be considered as a time-efficient solution to EN-based LVC identification.

The rest of this paper is organized as follows: Section 2 details the design of our new EN system, and introduces the LVC identification methods and experimental setups. Section 3 presents the experimental results and discussions. Section 4 concludes the whole paper.

2. Materials and Methods

2.1. Integrated Handheld EN

The operation scene of using SMUENOSEv2 for LVC identification is shown in Figure 1. All components of SMUENOSEv2, such as volatilization, sampling, powering, and computing units, were enclosed in an 18.5 cm \times 12.0 cm \times 11.0 cm cuboid case. The net weight of SMUENOSEv2 is about 1.0 kg. Thus, it is convenient to hold SMUENOSEv2 by hand and conduct the LVC identification operation.



Figure 1. The scene of using SMUENOSEv2 for LVC identification.

The structure block diagram of SMUENOSEv2 is shown in Figure 2. The physical blocks of SMUENOSEv2 comprise gas transportation and electronic hardware blocks, which were denoted in Figure 2 as dashed and solid boxes, respectively. The gas route components are used to compose the gas routes needed during the PI process. Electronic

hardware blocks are responsible for all electronic functions. Moreover, apart from the physical blocks, the normal operation of our EN also relies on the software components. Sections 2.1.1–2.1.3 detail the design of gas routes, electronic hardware, and software components, respectively.

Three-way valve	Gas chamber	Touch-sensitive monitor	
Volatilization pot	Sensing voltage sampling board	NVIDIA Jetson Nano	
Air pump	Voltage conversion board		
Lithium battery			

Figure 2. The structure block diagram of SMUENOSEv2.

2.1.1. Gas Route Design

Figure 3 shows the gas route of SMUENOSEv2. At the beginning of LVC identification, the LVC sample was injected into the volatilization pot. Then, fast air flow is generated using the air pump to accelerate the volatilization of LVC sample in the volatilization pot. The volatilized gaseous LVC sample is then carried to pass through the three-way valve and gas chamber, which are linked by silicone rubber tubes. In the meantime, by controlling the three-way valve, the LVC flow can be switched between gas route branches 1 and 2, which can in turn change the EN's working mode between sampling and washing modes.



Figure 3. The gas route of SMUENOSEv2. Red bold numbers **1** and **2** indicate the two gas route branches in the downstream side of the three-way valve.

SMUENOSEv2 can work in two different modes: sampling mode and washing mode. It is readily seen that SMUENOSEv2 only employs one valve. The gas routes with respect to the two modes can be described as follows:

- Gas route in the sampling mode: As preparation for sampling, a fixed volume of LVC sample was dripped into the volatilization pot in advance. The fast airflow can accelerate the LVC volatilization process and drive the gaseous LVC towards the three-way valve. Through gas route branch 1, gaseous LVC can get into contact with the gas sensors mounted in the gas chamber and stimulate the sensor responses.
- Gas route in the washing mode: After each sampling time, some LVC residuals still clung to the gas route components, including volatilization pot, three-way valve, gas chamber, and the silicon tubes. To reduce the negative influence among different types of LVC samples, the EN should be thoroughly washed between successive sampling spans. The washing process comprises two steps: First, by switching to gas route branch 2, the EN enters the washing mode. The volatilization pot, three-way valve, and upstream silicon tubes can be washed by clean air flow. Second, by switching back to gas route branch 1, the upstream clean air is driven to wash the gas chamber and downstream silicon tubes.

2.1.2. Electronic Hardware Design

The electronic hardware structure of SMUENOSEv2 is shown in Figure 4. In terms of functionality, the components in Figure 3 can be categorized into sensing, sampling, computing, displaying, and powering components. The design of these components is detailed as follows.



Figure 4. Structure of the electronic hardware in SMUENOSEv2. The blocks inside the dashed rectangle belong to the sensing voltage sampling board.

The eight different gas sensors mounted in SMUENOSEv2 are MiCS-6814, MiCS-5914, and MiCS-5521 from SGX SENSORTECH Corporation, as well as TGS-2620, TGS-2602, TGS-2600, TGS-2611, and TGS-8100 from Figaro Engineering Inc. All these sensors are MOS sensors, which enable fast response to the contact with the gaseous LVC. The criteria of selecting these sensors are two-fold: reactivity to more target gases, and selective sensitivities for different gases. According to the sensors' datasheets, typical compositions of common volatile chemicals, such as ethanol, carbon monoxide, and iso-butane, can be detected by the selected sensors. In addition, the sensitivity of different selected sensors to an individual target gase are different. For example, the sensitivity coefficient of six selected sensors to the target gases are shown in Figure 5. It is readily seen that the vertexes of each sensor's radar plot differ significantly. The largest sensitivity coefficients of different. Moreover, the different sensitivity coefficients with respect to an individual gas are different. It is expected that the sensing voltages detected by the selected sensors provide sufficient useful distinction information for subsequent identification sub-processes.



Figure 5. Radar plot of the sensitivity coefficients looked up from the sensors' datasheets.

The voltage sampling front-end circuit shown in Figure 6 is used in SMUENOSEv2. As shown in Figure 6, a load resistor R_L is connected with the sensor's sensing resistor R_s in series. All the power supply voltage values V_{sup} are 5 V. By measuring the voltage u_L , the sensing resistance value can be calculated as follows:

$$R_s = -R_L + V_{sup} \times R_L \frac{1}{u_{iL}} \tag{1}$$



Figure 6. The front-end circuit for sampling the *i*-th sensor's sensing voltage. *IOA* is an integrated operation amplifier. V_{sup} is the power supply voltage of the sensors.

Thus, the gas concentration measurement problem becomes the problem of sampling the voltage u_L of the load resistor. To amplify the voltage variation, improve the load capacity, and filter the measurement noises, an integrated operation amplifier is used to form a low-pass filtering amplifier circuit. As shown in Figure 6, the voltage u_L is connected to a typical *RC* low-pass filter [36] before being passed to the in-phase end of the integrated operation amplifier *A*. The cut-off frequency of the low-pass filter composed of R_2 and *C* is

$$f_0 = \frac{1}{2\pi R_2 C} \tag{2}$$

which means the input voltage's constituents with a frequency higher than f_0 will be significantly suppressed. In SMUENOSEv2, the value of *C* and R_2 are 2.2 uF and 16 k Ω , respectively. Afterwards, the filtered voltage u'_L is transmitted into an in-phase proportion amplifying circuit composed of *A*, R_3 , and R_4 . The amplified voltage u_o can be calculated as

j

$$u_L = \left(1 + \frac{R_4}{R_3}\right) u'_L \tag{3}$$

Apart from the gas sensors, the parameters of other components in the circuit shown in Figure 6 were empirically determined to keep the value of u_0 within the valid range of AD7606. As an 8-channel analog-to-digital data acquisition chip, a piece of AD7606 was used to simultaneously sample the output of the 8 sensors' front-end circuits. The sampling data was sent through the parallel interface to STM32F407VGT6, which was also used to control the three-way valve, air pump, and the timer chip DS-1338. The timer chip was used to provide accurate time information for controlling the sampling duration.

The sampling data was then sent to NVIDIA Jetson Nano. A touch-sensitive monitor was used to interact with Jetson Nano. The identification results of Jetson Nano were displayed in the monitor. By touching the monitor's display screen, user commands can be sent to Jetson Nano. Last but not the least, a voltage converter component, which was bought from the commercial market, is incorporated to convert the 12 V voltage from the battery to stable 5 V and 3.3 V voltages for the above-mentioned components.

2.1.3. Software Design

The software of SMUENOSEv2 comprises three software components: perception, pattern recognition, and human–machine-interface (HMI) components. The perception component is composed of codes in C programming language running on the STM32, while the other two components were codes in Python programming language running on the NVIDIA Jetson Nano board.

Figure 7 shows the flow charts of the perception and pattern recognition components, which interact with each other through the serial port. The SAMPLING command is originally sent from the HMI component to the pattern recognition component, and then sent to the perception component. Once the SAMPLING command is received, the ADC chip AD7606 begins to sample the sensing voltages of the eight sensors with a sampling frequency of 150 Hz. A single sampling cycle lasts for a predefined time span. The sampling duration is determined by periodically reading the current time output by DS1338. After an individual sampling cycle, the received sampling data is saved in a matrix. If the sampling data is valid, a SAVING command can be sent from the HMI component, which can trigger the storage of the data matrix in an Excel file. In real applications, to identify a certain LVC sample, the sampling subprocess is followed by an identification subprocess. The data obtained in an individual sampling cycle is used as the test dataset. For model learning, the training subprocess loads multiple previously saved Excel data files to form the training dataset. The detail of training and identification subprocesses is related to the specific employed ML methods.



Figure 7. Flow charts of the perception and pattern recognition components. The dashed arrows stand for the interaction between the two components.

Figure 8a,b show the HMI surface of identifying an individual liquor sample and the model training process, respectively. The graphical user interface of the HMI component comprises three sectors:

1. Sector I: the real-time sensing voltage displaying sector. During the sampling process, the real-time sensing voltages received from the sampling component are plotted in this sector. The curves in different colors correspond with the data of different sensors.



Figure 8. The human–machine interfaces (HMI) of SMUENOSEv2. (a) HMI for sampling and identification process; (b) HMI for model-training process.

- 2. Sector II: the sampling operation sector. By pushing the "Sampling" and "Save" buttons in sector II, the SAMPLING and SAVING commands can be sent to the pattern recognition components, respectively. As samples to be saved for forming the training set, the sample type is previously known. For example, to form the training sets in real applications, multiple samples with known type should be collected using certified LVC in advance. This prerequisite can be satisfied by selecting the right type in the ComboBox next to the "Save" button. Moreover, since SMUENOSEv2 was designed to be capable of identifying different liquid or gaseous volatile organic compounds, it is important to set the sample classes (e.g., perfume, wine, herb) in the left ComboBox, before sending the SAVING command.
- 3. Sector III: the identification and training operation sector. Empirically, the identification and training operations are mutual exclusive, and, thus, are placed in two different tabs sharing sector III. The operation processes for training and identification are detailed as follows:
 - By pushing the "Training" button in the bottom-right corner of Figure 8b, an individual training process can be activated. The training dataset and ML method should be selected before pushing the "Training" button. In the central area of this sector, a table lists all previously saved sensing voltage samples. By clicking the table items, multiple samples can be selected to form the training dataset. The ML method used for training can be selected in the top-right ComboBox. Moreover, the hyper-parameters of the employed method can be set by pasting the specially formatted parameter values in the TextBox above the "Training" button. Finally, the resulting trained ML model, which can be used for identification, is saved in the external memory of NVIDIA Jetson Nano.
 - By pushing the "Identification" button in the bottom-right corner of Figure 8a, an individual identification process can be activated. As mentioned in the previous paragraph, the identification process follows the sampling process, and uses the newly collected samples as input. Moreover, as the identification tool, a

previously trained and saved ML model should be selected from the model lists before pushing the "Identification" button. During the identification process, the process status is shown in the blank space at the left of the "Identification" button. Finally, the identification result is displayed in the blank space above the model lists.

2.2. Improved GBDT

To evaluate the performance of SMUENOSEv2, two improved GBDT methods, i.e., XG-Boost [31] and LightGBM [32], were implemented in SMUENOSE2 to realize LVC identification. For coherence and clarity, XGBoost and LightGBM are in turn introduced. Since the two improved GBDT methods were developed based on GBDT, GBDT is also sketched as follows.

2.2.1. GBDT

GBDT [30] is a framework originally designed for function regression. It begins with a set of data {(x_i , y_i)} and an unsatisfactory model f(x), which cannot accurately repeat the relationship between the data. In the *k*-th iteration of GBDT, an additional model $h_k(x)$ with the form of a regression tree is constructed by fitting the residuals $y_i - f_{k-1}(x_i)$. Inspired by the fact that the true value equals the sum of residual and current model output, the new model fitting the residuals is expected to compensate for the inaccuracy of current model. However, the new model $h_k(x)$ can only approximate the theoretical residuals, which means the new additive model $f_{k-1}(x_i) + h_k(x)$ is still not satisfactory. Thus, the sub-sequent iterations are sequentially conducted to further improve the accuracy of the additive model. Moreover, in order to avoid overfitting, a learning rate η , $0 < \eta < 1$, is incorporated in the additive model to prevent full optimization in each step. The additive mode in GBDT turns into:

$$f(x) = \sum_{k} \eta \cdot h_k(x) \tag{4}$$

The above residual-fitting process is correlated with gradients by defining the square loss:

$$L(y, f(x)) = (y - f(x))^2 / 2$$
(5)

Then, the problem transfers to minimization of $J = \sum_{i} L(y_i, f(x_i))$ by adjusting $f(x_1)$, $f(x_2), \ldots, f(x_n)$. Taking $f(x_i)$ as parameters, the gradient can be represented as follows:

$$\frac{\partial J}{\partial F(x_i)} = f(x_i) - y_i \tag{6}$$

According to Equation (4), the residuals defined in GBDT can be interpreted as negative gradients. Thus, the model in GBDT is actually updated using the gradient descent method. In general cases, the residuals and square loss are replaced with negative gradients and any differentiable loss function, respectively.

For solving multi-class classification problems using GDBT, the label information is turned into a true probability distribution. For example, if the label of the *i*-th instance is *C*, then the probability distribution $Y_L(x_i)$ can be represented as follows:

$$Y_L(x_i) = \begin{cases} 1, & L \text{ equals } C\\ 0, & L \neq C \end{cases}$$
(7)

Afterwards, the problem turns into the regression of the true probability distribution function. The KL-divergence between the true and predicted probability distributions is considered as the loss function in GBDT.

2.2.2. XGBoost

In XGBoost [31], the implementation of GBDT is mainly increased in three aspects: incorporating a regularization term in the objective function, usage of second-order gradient statistics in the loss function approximation, and providing an approximation algorithm for split finding in the decision tree construction. The three aspects are sketched as follows:

. Incorporating a regularization term in the objective function.

The incorporated regularization term serves as a measure of the model complexity. Instead of directly minimizing the loss function, XGBoost utilizes the sum of loss function and the regularization term as the objective function for minimization. As a measure of model complexity, the minimization of regularization term can help avoid the over-fitting problem, which means the learned model fits well on the training data but badly on the test data. Empirically, simpler models are prone to demonstrating smaller performance variance between training and test data sets. The regularization term with l_2 norm in XGBoost can be represented as follows:

$$\Omega = \gamma T + \frac{1}{2}\lambda \|\omega\|^2 \tag{8}$$

where γ and λ are two constant coefficients, and w and T are the leaf score and the number of tree leaves, respectively.

2. Usage of second-order gradient statistics in the loss function approximation.

Inspired by the Taylor expansion, the first-order and second-order gradients are used to construct a simplified objective function as follows:

$$\widetilde{L}_k = \sum_{i=1}^n \left[g_i f_k(x_i) + \frac{1}{2} h_i f_k^2(x_i) \right] + \Omega_t$$
(9)

where g_i and h_i are the first-order and second-order partial derivative of the loss function to $f_k(x_i)$. Compared with only using the negative first-order gradient in GBDT, the second-order gradients used in XGBoost contain more information about the loss function, and thus, enable more accurate approximation of the loss function.

3. The approximation algorithm for split finding in the decision tree construction.

Traditional exact greedy algorithm used for split finding in the decision tree construction should enumerate all the possible splits, which is time-consuming. To adapt to large training datasets and resource-constrained applications, XGBoost provides an approximation algorithm for split finding. Main steps of the approximation algorithm can be summarized as follows:

- (1) For an individual feature, all the possible split points are mapped into buckets according to the percentile of feature distribution.
- (2) Calculate the cumulative first-order and second-order gradients for each bucket.
- (3) Consider the bucket with the largest cumulative gradient statistics as the optimal bucket. Then, the split point with the largest gradient statistics in the optimal bucket is selected as the final split point.

Apart from the above three aspects, XGBoost also additionally utilizes the column subsampling technique in Random Forest to further prevent over-fitting and accelerating the computation.

2.2.3. LightGBM

Aiming at increasing the efficiency of GBDT, LightGBM [32] reduces the volume of employed training sample data with two innovation techniques: gradient-based one-side sampling (GOSS), and exclusive feature bundling (EFB). In GOSS, the under-trained samples are paid more attention, on the premise of only slightly changing the original sample data distribution. First, the sample data are sorted in a descending order of the absolute gradients. Then, the sample dataset is divided into two parts: the part of *a* × *n*, 0 < a < 1, samples with the largest absolute gradients; and the part of other samples. All

samples in the first part, and $b \times 100\%$, 0 < b < 1 of other samples are used in the training of next decision tree.

EFB reduces the training data volume by decreasing the number of features engaged in the training process. This is realized by bundling the exclusive features, which seldom get non-zero values at the same time, into a single feature. The recognition of exclusive features is modeled as a reduced graph-coloring problem, which is then solved using a greedy algorithm. To merge the exclusive features in the same bundle, the feature bundle is created by adding offsets to the original feature values of exclusive features to make sure they are placed in different bins.

Moreover, LightGBM utilized a histogram-based algorithm to improve the efficiency in both running speed and storage consumption. For split finding in the decision tree construction, histogram-based algorithm transfers continuous feature values into discrete values, and places them in different bins. Then, the bins are used to construct feature histograms in the training process. In LightGBM, the combination of GOSS and EFB with the histogram-based split finding serves as an efficient solution to ignore the features with zero values.

2.3. Experimental Setup

In our experiments, we tested two common sets of LVC: perfume and liquor. Six different perfumes of the same brand "Scent Library" and four liquors of different brands were used as samples in the series of perfume and liquor experiments, respectively. The model names, abbreviates, and main compositions of these samples are listed in Table 1. The two series of experiments were both conducted on our newly designed E-nose platform SMUENOSEv2. Thus, the sampling and identification processes for them are the same.

Table 1. Ranges of main tunable hyper-parameters of the tested methods.

	Abbr.	Model Name	Main Compositions
Perfumes	GO	Golden Osmanthus	Ethanol, Essence, Propylene glycol, Dipropylene glycol, Tridecanol polyether-9
	MR	Misty Rainbow	Ethanol, Essence, Propylene glycol, Dipropylene glycol, PPG-26-Butanol polyester-26
	LBK	L.B.K. Water	Denatured ethanol, Essence, Tocopherol acetate, Butyl hydroxytoluene
	RR	Rose Rose I Love You	Ethanol, Essence, Propylene glycol, Dipropylene glycol, Oaklirin
	THY	Tao Hua Yun	Denatured ethanol, Essence, Tocopherol acetate, Butyl hydroxytoluene
	WR	White Rabbit	Ethanol, Essence, Propylene glycol, Tertiary butanol, Ethylhexyl salicylate
Liquors	DHX	Dao Hua Xiang	Edible alcohol, Red sorghum, Wheat, Rice, Maize, Edible essence
	JLZ	Jin Lu Zhou	Sorghum, Wheat, Rice, Glutinous rice, Maize
	NLS	Niu Lan Shan	Sorghum, Liquid-state liquor, Edible essence
	QJ	Qing Jiu	Sorghum, Wheat, Glutinous rice, Maize

To obtain the raw samples, each of the specific perfume and liquor samples was sampled for 50 cycles, which means a total of 500 sampling cycles were conducted. In a single sampling cycle, 1 uL of the LVC material was injected into the E-nose's volatilization pot. Then, the air pump was activated to accelerate the volatilization of the LVC material and carry the volatilized gases through the gas chamber. Stimulated by the gaseous LVC, the fast-fluctuating sensing voltages of the eight sensors were sampled and recorded. The time span of each sampling cycle was set as 100 s and 60 s for the perfume and liquor experiments, respectively. The sampling cycle for perfume was longer, since we found the sensors' recovering time in perfume experiments was longer than that in liquor experiments. This phenomenon is probably because the perfume samples can stick to the gas route for longer time.

The obtained raw measurements were then preprocessed. Afterwards, multiple features were generated based on the preprocessed measurements. Finally, the generated features were used as the training and test data to evaluate the LVC identification performance of SMUENOSEv2 via the two ensemble learning methods. Specifically, 10-fold cross-validation experiments were conducted: The feature set was randomly divided into 10 equal parts. Each part was in turn used as the test set while the rest of the parts were used as the training set. Thus, in each group of 10-fold cross-validation experiments, we conducted 10 experiments with different training and test data sets. With respect to each training set, the optimal combination of model parameters was selected during the model training process. The preprocessing, feature generation, and parameter selection processes are detailed as follows.

2.3.1. Preprocessing

First, the outliers in the raw measurements, which were mainly caused by the transmission errors in the high-speed serial communication, were removed based on the mean and standard deviation. If the difference between a voltage measurement u and the mean voltage is bigger than 5 times of the standard deviation, then all the values obtained at the same time as u were deleted.

Second, the high-frequency measurement noises, which were mainly introduced by the circuits, were removed using a low-pass filtering algorithm. As an important supplement of the hardware low-pass filter shown in Figure 6, a second-order Butterworth low-pass software filter with a cut-off frequency of 0.016 half-cycles/sample was constructed.

Third, the base-line voltages, which means the voltages obtained right before the sensors got into contact with the gaseous LVC material, were subtracted from all the 15,000 voltage values in the corresponding sampling cycle. The base-line removement operation can help reduce the negative influence of background gas residuals.

2.3.2. Feature Generation

As mentioned in Section 2.1, the sampling frequency of AD7606 in SMUENOSEv2 was set as 150 Hz. Therefore, in an individual sampling cycle, 15,000 voltage values were obtained for each sensor, which means a total of 120,000 sensing voltage values were collected by the eight sensors in our E-nose. Due to the curse of dimensionality, it is not feasible to directly train the ML models with the raw sampling data, especially on the resource-limited NVIDIA Jetson Nano in SMUENOSEv2. Therefore, it is necessary to reduce the dimension of the training data. For the sake of efficiency, we chose to combine multiple statistics of the sampling data as the features, rather than rely on feature extraction algorithms, e.g., principal component analysis, to determine the features used for training.

As listed in Table 2, the statistics used in our experiments are the characteristics of the voltage curves in three different period: peak, rising, falling. The statistics were calculated based on the voltage $u_L(t)$, since it can be mapped into the sensing resistance in a one-to-one manner according to Equation (1). The characteristics in Table 2 were empirically selected, since they are capable of characterizing the voltage dynamics in a single sampling cycle. Based on the sampling data of a single sensor in each cycle, five feature values can be calculated according to the equations in Table 2. Forty features were obtained based on the measurements in each sampling cycle, since there are 8 different gas sensors in our E-nose platform. The forty feature values with respect to all 8 sensors were then combined and indexed in an end-to-end manner to form the feature vector, which means the feature dimension in our experiments is forty.

Periods	Characteristic Meaning	Calculation Equations	Feature Index
Peak	Normalized voltage peak	$\widetilde{u}_m = \operatorname{norm}\left[\max_t u_L(t)\right]$	1, 2, 3,, 8
	Normalized time of voltage peak	$\widetilde{t}_m = \operatorname{norm}\left[\operatorname{argmax}_t u_L(t)\right]$	9, 10, 11, , 16
Falling	Final falling proportion	$p_1 = u_f / u_m^{-1}$	17, 18, 19, , 24
	Falling proportion at 10 s after the voltage peak	$p_2 = u_L(t_m + 10)/u_m$	25, 26, 27,, 32
Rising	Normalized time of the maximal first-order gradient of $u_L(t)$	$\tilde{t}_r = \operatorname{norm}\left[\operatorname{argmax}_t \frac{du_L(t)}{dt}\right]$	33, 34, 35, , 40

Table 2. The statistics used as features for classification.

¹ u_f means the final value of $u_L(t)$ at the end of sampling cycle.

2.3.3. Parameter Tuning

Before training the models based on a new dataset, it is necessary to tune the parameters of the two tested methods, which are listed in Table 3. The *hyperopt* Python module [37] was used for parameter tuning. In the *hyperopt* module, the parameter search problem was modeled as a multi-variable function optimization problem, which takes the identification accuracy as the objective function. Then, the Tree of Parzen Estimators algorithm [38] was used to solve the optimization problem and output the optimal parameter value combination. Note that not all parameter values should be tuned. For both methods, the "*objective*" and "*num_class*" parameters were fixed as "*multi-class*" and the actual class numbers (i.e., 6 for perfumes, and 4 for liquors), respectively. It is applicable to provide the problem type and total number of classes for the methods, since they are both supervised ML methods. Apart from the influential parameters listed in Table 3, other parameter values were fixed as their default values in *hyperopt*.

Table 3. Ranges of main tunable hyper-parameters of the tested methods.

	Parameter Meaning	Range
	Learning rate used in the additive model	uniform(0.7, 1) 1
	Maximum depth of the constructed decision trees	$\{3, 4, \ldots, 10\}$
	Ratio of data selected to construct decision trees	uniform(0.4, 0.6)
XGBoost	Minimum loss reduction required to split a leaf node	uniform(0, 0.4)
	L1 regularization term on weights	uniform(0.2, 0.7)
	L2 regularization term on weights	uniform(0.2, 0.7)
	Minimum sample weights in each leaf node of decision tree	uniform(3, 7)
LightGBM	Learning rate used in the additive model	uniform(0.7, 1)
	Maximum depth of the constructed decision trees	$\{3, \ldots, 10\}$
	Number of leaves in a decision tree	$\{15, 4, \ldots, 100\}$
	Minimum sample numbers in each leaf node of decision tree	{3, 4, 5, 6, 7}
	L1 regularization term on weights	uniform(0, 1)
	L2 regularization term on weights	uniform(0, 1)
	Ratio of data randomly selected without resampling	uniform(0.4, 1)
	Frequency for randomly selecting data without sampling	$\{1, 2, 3, \dots, 20\}$

¹ *uniform*(*, Δ) means generate a value according to the uniform distribution within the range (*, Δ).

3. Results and Discussions

Based on the data obtained in 500 sampling cycles, we have conducted five groups of ten-fold cross-validation experiments for each combination of LVC type and identification method. Since we tested two LVC types (i.e., perfume and liquor) and two identification methods (i.e., XGBoost and LightGBM), we have conducted a total of 200 individual experiments with different training and test data sets.

To demonstrate the validation of using SMUENOSEv2 for sampling the transient voltage values stimulated by the volatized LVC, Figure 9 shows the value and first-order

gradients of preprocessed u_L . All curves in Figure 9a demonstrate fast-rising and slowfalling dynamics. This phenomenon coincides with the fast reaction and slow recovery characteristic of MOS sensors. The peak values and time, falling speed, and the maximum first-order gradients are distinctive for different sensors. Thus, it is expected that the statistics in Table 2 contain specific "fingerprint" information about the corresponding LVC sample.



Figure 9. Typical values obtained in an individual cycle of liquor sampling. (a) The value of preprocessed u_L ; (b) The first-order gradient of preprocessed u_L in the first 10 s of a liquor sampling cycle.

The statistical results obtained in the perfume and liquor experiments are presented as follows.

3.1. Experiments in the Perfume Group

Figure 10a shows the boxplot of identification accuracies in the perfume experiments. The identification accuracy is defined as the proportion of correct identification instances. As mentioned in Section 2.1.3, in an identification instance, a previously trained model took the sample with the true type t as input, and then output a predicted type t' as the identification result. If the predicted type t' equals to the true type t, then the identification instance is considered as correct. Otherwise, the identification is considered as incorrect. As shown in Figure 10a, apart from the mean values, other boxplot elements are similar for the two tested methods. The identification accuracy means of XGBoost and LightGBM are 94.8% and 95.5%, respectively. All obtained identification accuracies are higher than 90%. The high identification accuracies verify that it is feasible to conduct perfume identification using our newly designed E-nose platform.



Figure 10. Statistical results obtained in perfume experiments. (**a**) Boxplot of the identification accuracies. (**b**) Boxplot of the time spent for training. (**c**) Boxplot of the time spent for identification.

Figure 10b,c show the boxplot of time spent for training and identification, respectively. The considerably short training and identification time also verified the feasibility of using our resource-constrained integrated E-nose for perfume identification. The training times spent in all trials are shorter than 45 s. Compared with the identification times, which are shorter than 1 s, the training times are much longer. Fortunately, for on-site usage in real applications, the training component is less frequently used than the identification component. The well-trained model can be repeatedly used for identification. The much longer training time can be attributed to the time spent for parameter selection, which was reckoned with in the training time since parameters are also part of the trained model. During the training process, we found the parameter selection time dominate the training time. As mentioned in Section 2.3.3, the parameter selection was realized by solving function optimization problems, which involve a large number of model training with different parameter combinations. Moreover, the time-spans spent by XGBoost are generally longer, but more concentrated, than those spent by LightGBM. The generally shorter but more divergent training and identification time of LightGBM is mainly attributed to its GOSS and EFB mechanisms, which reduced the number of engaged features and incorporated more random operations.

To thoroughly compare the identification accuracy with respect to different perfume types, the confusion matrices obtained in the perfume experiments are shown in Figure 11. The elements of confusion matrices are the times of experiments. As mentioned, to test each method based on the perfume samples, five groups of ten-fold cross-validation experiments were conducted, which means a total of 50 individual perfume experiments for each of XGBoost and LightGBM. Moreover, as mentioned in Section 2.3, each perfume type was sampled for 50 cycles. The 50 samples were divided into 10 equal parts, and an individual part with five samples was used as the test set in each perfume experiment. Thus, for each method, each perfume type was tested 250 times, which coincides with the fact that the elements in each row of the confusion matrices sum up to 250. The auxiliary diagonal elements stand for the times of correctly identified experiments.



Figure 11. Confusion matrices in perfume experiments. (**a**) The confusion matrix obtained by XGBoost; (**b**) The confusion matrix obtained by LightGBM.

In both confusion matrices, the LBK perfume samples were all successfully identified. Compared with other perfume types, the MR perfume samples were incorrectly identified the most times. The differences among the identification accuracies of different perfume types mainly came from the different chemical constituents of the perfume samples, which are not the main concern of this paper. The numbers in both confusion matrices demonstrate the best and worst performance of using our E-nose platform with the two ensemble learning methods for perfume identification. Even in the worst case, an identification accuracy of 88.4% was obtained in the perfume experiments, which further validates the reliability of our E-nose platform SMUENOSEv2.

Figure 12 shows the feature importance obtained in perfume experiments. The feature importance is defined as the proportion of times the feature was used in a model, which can be represented as follows:

$$f_i = \frac{e_i}{\sum_{j=1}^{40} e_j}, i = 1, 2, \dots, 40$$
(10)

where e_i denotes the number of times the *i*-th feature was used in a model. As mentioned in Section 2.3.2, the feature dimension of our problem is forty. The more times a feature was used to construct the decision trees means the feature has taken more effect in the ensemble learning method. Thus, the feature importance measures the contribution degree of each feature. According to Figure 12 and the feature indexes listed in Table 2, at least two conclusions can be summarized as follows:

- 1. For both methods, the normalized voltage peaks were generally more important than other characteristics. On the contrary, the normalized times of the maximum first-order gradient were slightly less important than other characteristics, although the importance of feature 38 in Figure 10b existed as an exception. It can be concluded that the peak values contain more distinguishing information about the different perfume constituents. The maximum first-order gradient, which characterizes the time of maximum rising velocity, is less correlated with the stimulant perfumes.
- 2. XGBoost obviously focused more on features 1, 8, and 18 than other features. In comparison, LightGBM concentrated more on almost a half of the features while less on the other features. The decentralized attention of LightGBM could be attributed to its EFB mechanism.



Figure 12. The boxplot of feature importance in perfume experiments. (**a**) The feature importance obtained by XGBoost; (**b**) The feature importance obtained by LightGBM.

3.2. Experiments in the Liquor Group

With respect to the liquor experiments, Figure 13 shows that the identification accuracies and time-efficiencies are also considerably high, which validates the feasibility of using our E-nose for liquor identification. Correspondingly, Figures 14 and 15 demonstrate the confusion matrices and feature importance of the liquor experiments, respectively. The meaning of the factors in Figures 13–15 are the same as those in Figures 10–12, respectively. The dimension of confusion matrices in liquor experiments is 4, since only four liquor types were tested in our experiments. By comparing the results obtained by the two tested methods shown in Figures 13–15, we found that some conclusions similar to those in the perfume experiments can also be deduced in the liquor experiments. For example, in the liquor experiments, the training and test time spent by XGBoost are generally longer, but more concentrated, than those spent by LightGBM.



Figure 13. Statistical results obtained in liquor experiments. (a) Boxplot of the identification accuracies.(b) Boxplot of the time spent for training. (c) Boxplot of the time spent for prediction.



Figure 14. Confusion matrices in liquor experiments. (**a**) The confusion matrix obtained by XGBoost; (**b**) The confusion matrix obtained by LightGBM.



Figure 15. The boxplot of feature importance in liquor experiments. (**a**) The feature importance obtained by XGBoost; (**b**) The feature importance obtained by LightGBM.

Moreover, compared with the perfume experiments, the identification accuracies obtained in the liquor experiments are more divergent. The ranges with $1.5 \times IQR$ of Figure 13a are 15%, which are 5% higher than those of Figure 10a. This phenomenon coincides with the relatively smaller elements of the confusion matrices in Figure 14. The relatively lower identification accuracies obtained in liquor experiments are mainly because the constituents in different liquor types are closer with each other than those in different perfume types. The increased difficulty of liquor identification brings about more challenges for the E-nose platform and identification methods. It is noteworthy that the mean identification accuracies in the liquor experiments are 92.3% and 94.6, which are still considerably high.

4. Conclusions and Future Works

In this paper, an integrated handheld EN was designed for LVC identification. By LVC, we mean the chemical that can be volatilized from original liquid form to gaseous form by themselves or external blowing. The computing core of our newly designed EN is an NVIDIA Jetson Nano module. Owing to its miniature volume, the NVIDIA Jetson Nano module can be mounted together with other EN components in a handheld case. The newly designed EN consists of gas transportation and electronic hardware components. A small air pump was used to accelerate the volatilization of LVC samples and transport their gaseous form towards the sensor array. In the meantime, a STM32 processor was used to acquire the sensing voltages of the sensor array and transmit the acquired data to NVIDIA

Jetson Nano, on which two improved GBDT methods (i.e., XGBoost and LightGBM) were separately employed for LVC identification. Compared with common divide-body EN designs, our integrated EN is more suitable for on-site quick LVC identification. With GBDT models previously trained by both methods, our EN can realize highly accurate identification of perfumes and liquors in less than one second. In comparison, LightGBM spent generally less time for the model training and identification processes.

Despite its considerably high performance on LVC identification, our newly designed EN could be further improved in the following aspects:

- (1) Adapting the EN for identifying original gaseous chemicals. In the current EN design, it is supposed that an LVC in its original liquid form was dripped into the EN and then volatilized to the gas form. The scheme for adding original gaseous chemicals into the EN can be further investigated.
- (2) Adding a temperature and humidity control module to the EN. Controlling the temperature and humidity around the sensor array at a fixed value was not considered in the current EN design. The sensing data acquired at significantly different temperature and humidity values could influence the identification results. Future research works should cover adding a temperature and humidity control module to our EN.

Author Contributions: Conceptualization, methodology, software, investigation, writing—original draft preparation, writing—review and editing, M.C.; writing—review and editing, supervision, project administration, X.H.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant no. 61801287).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Acronyms and Abbreviations

LVC	Liquid Volatile Chemicals
EN	Electronic Nose
GBDT	Gradient-Boosting Decision Tree
MOS	Metal Oxide Semiconductor
GC	Gas Chromatography
GC-MS	GC combined with Mass Spectrometry
FPGA	Field Programmable Gate Array
ML	Machine Learning
XGBoost	Extreme Gradient Boosting
LightGBM	Light Gradient-Boosting Machine
HMI	Human–Machine-Interface
GOSS	Gradient-based One-Side Sampling
EFB	Exclusive Feature Bundling

References

- 1. Patial, P.; Deshwal, M. Selectivity and Sensitivity Property of Metal Oxide Semiconductor Based Gas Sensor with Dopants Variation: A Review. *Trans. Electr. Electron. Mater.* **2021**, *23*, 6–18. [CrossRef]
- Sharma, P.; Tudu, B.; Bhuyan, L.P.; Tamuly, P.; Bhattacharyya, N.; Bandyopadhyay, R. Detection of Methyl Salicylate in Black Tea Using a Quartz Crystal Microbalance Sensor. *IEEE Sens. J.* 2016, 16, 5160–5166. [CrossRef]
- Puttasakul, T.; Pintavirooj, C.; Sangma, C.; Sukjee, W. Hydrogel Based-Electrochemical Gas Sensor for Explosive Material Detection. *IEEE Sens. J.* 2019, 19, 8556–8562. [CrossRef]
- Li, Z.; Gui, X.; Hu, C.; Zheng, L.; Wang, H.; Gong, J. Optical Gas Sensor Based on Gas Conjugated Interference Light Source. *IEEE Photon-Technol. Lett.* 2015, 27, 1550–1552. [CrossRef]
- Liu, X.; Dong, H.; Xia, S. Micromachined catalytic combustion hydrogen gas sensor. In Proceedings of the 8th Annual IEEE International Conference on Nano/Micro Engineered and Molecular Systems, Suzhou, China, 7–10 April 2013; pp. 282–285.
- 6. Khatib, M.; Haick, H. Sensors for Volatile Organic Compounds. ACS Nano 2022, 16, 7080–7115. [CrossRef]

- Freddi, S.; Sangaletti, L. Trends in the Development of Electronic Noses Based on Carbon Nanotubes Chemiresistors for Breathomics. *Nanomaterials* 2022, 12, 2992. [CrossRef]
- 8. Roy, M.; Yadav, B.K. Electronic nose for detection of food adulteration: A review. J. Food Sci. Technol. 2021, 59, 846–858. [CrossRef]
- 9. Arnold, C.; Harms, M.; Goschnick, J. Air quality monitoring and fire detection with the Karlsruhe electronic micronose KAMINA. *IEEE Sens. J.* **2002**, *2*, 179–188. [CrossRef]
- Ali, M.M.; Hashim, N.; Aziz, S.A.; Lasekan, O. Principles and recent advances in electronic nose for quality inspection of agricultural and food products. *Trends Food Sci. Technol.* 2020, 99, 1–10. [CrossRef]
- Ju, X.; Lian, F.; Ge, H.; Jiang, Y.; Zhang, Y.; Xu, D. Identification of Rice Varieties and Adulteration Using Gas Chromatography-Ion Mobility Spectrometry. *IEEE Access* 2021, *9*, 18222–18234. [CrossRef]
- 12. Feizi, N.; Hashemi-Nasab, F.S.; Golpelichi, F.; Sabouruh, N.; Parastar, H. Recent trends in application of chemometric methods for GC-MS and GCGC-MS-based metabolomic studies. *TrAC-Trends Anal. Chem.* **2021**, *138*, 116239. [CrossRef]
- Li, J.-J.; Song, C.-X.; Hou, C.-J.; Huo, D.-Q.; Shen, C.-H.; Luo, X.-G.; Yang, M.; Fa, H.-B. Development of a Colorimetric Sensor Array for the Discrimination of Chinese Liquors Based on Selected Volatile Markers Determined by GC-MS. *J. Agric. Food Chem.* 2014, 62, 10422–10430. [CrossRef]
- 14. Cao, M.; Ling, X. Quantitative Comparison of Tree Ensemble Learning Methods for Perfume Identification Using a Portable Electronic Nose. *Appl. Sci.* **2022**, *12*, 9716. [CrossRef]
- 15. Uçar, A.; Özalp, R. Efficient android electronic nose design for recognition and perception of fruit odors using Kernel Extreme Learning Machines. *Chemom. Intell. Lab. Syst.* 2017, *166*, 69–80. [CrossRef]
- Kim, H.; Konnanath, B.; Sattigeri, P.; Wang, J.; Mulchandani, A.; Myung, N.; Deshusses, M.A.; Spanias, A.; Bakkaloglu, B. Electronic-nose for detecting environmental pollutants: Signal processing and analog front-end design. *Analog. Integr. Circuits Signal Process.* 2011, *70*, 15–32. [CrossRef]
- 17. Li, D.; Lei, T.; Zhang, S.; Shao, X.; Xie, C. A novel headspace integrated E-nose and its application in discrimination of Chinese medical herbs. *Sens. Actuators B Chem.* 2015, 221, 556–563. [CrossRef]
- 18. Meléndez, F.; Arroyo, P.; Gómez-Suárez, J.; Palomeque-Mangut, S.; Suárez, J.I.; Lozano, J. Portable Electronic Nose Based on Digital and Analog Chemical Sensors for 2,4,6-Trichloroanisole Discrimination. *Sensors* **2022**, 22, 3453. [CrossRef]
- 19. Wojnowski, W.; Majchrzak, T.; Dymerski, T.; Gębicki, J.; Namieśnik, J. Portable Electronic Nose Based on Electrochemical Sensors for Food Quality Assessment. *Sensors* 2017, *17*, 2715. [CrossRef]
- 20. Haddi, Z.; Amari, A.; Alami, H.; El Bari, N.; Llobet, E.; Bouchikhi, B. A portable electronic nose system for the identification of cannabis-based drugs. *Sens. Actuators B Chem.* 2010, 155, 456–463. [CrossRef]
- 21. Tang, K.-T.; Chiu, S.-W.; Pan, C.-H.; Hsieh, H.-Y.; Liang, Y.-S.; Liu, S.-C. Development of a Portable Electronic Nose System for the Detection and Classification of Fruity Odors. *Sensors* **2010**, *10*, 9179–9193. [CrossRef]
- 22. Huang, Y.; Doh, I.-J.; Bae, E. Design and Validation of a Portable Machine Learning-Based Electronic Nose. *Sensors* **2021**, *21*, 3923. [CrossRef] [PubMed]
- 23. Branca, A.; Simonian, P.; Ferrante, M.; Novas, E.; Negri, R. Electronic nose based discrimination of a perfumery compound in a fragrance. *Sens. Actuators B Chem.* 2003, 92, 222–227. [CrossRef]
- 24. Kim, S.-T.; Choi, I.-H.; Li, H. Identification of multi-concentration aromatic fragrances with electronic nose technology using a support vector machine. *Anal. Methods* **2021**, *13*, 4710–4717. [CrossRef] [PubMed]
- 25. Penza, M.; Cassano, G.; Tortorella, F.; Zaccaria, G. Classification of food, beverages and perfumes by WO3 thin-film sensors array and pattern recognition techniques. *Sens. Actuators B Chem.* **2001**, *73*, 76–87. [CrossRef]
- 26. Yang, Y.; Liu, H.; Gu, Y. A Model Transfer Learning Framework with Back-Propagation Neural Network for Wine and Chinese Liquor Detection by Electronic Nose. *IEEE Access* 2020, *8*, 105278–105285. [CrossRef]
- 27. Benrekia, F.; Attari, M.; Bouhedda, M. Gas Sensors Characterization and Multilayer Perceptron (MLP) Hardware Implementation for Gas Identification Using a Field Programmable Gate Array (FPGA). *Sensors* **2013**, *13*, 2967–2985. [CrossRef]
- Jia, T.; Guo, T.; Wang, X.; Zhao, D.; Wang, C.; Zhang, Z.; Lei, S.; Liu, W.; Liu, H.; Li, X. Mixed Natural Gas Online Recognition Device Based on a Neural Network Algorithm Implemented by an FPGA. *Sensors* 2019, *19*, 2090. [CrossRef]
- Zhai, X.; Ali, A.A.S.; Amira, A.; Bensaali, F. MLP Neural Network Based Gas Classification System on Zynq SoC. *IEEE Access* 2016, 4, 8138–8146. [CrossRef]
- 30. Friedman, J.H. Greedy function approximation: A gradient boosting machine. Ann. Stat. 2001, 29, 1189–1232. [CrossRef]
- Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
- Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A highly efficient gradient boosting decision tree. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 3149–3157.
- 33. Gewers, F.L.; Ferreira, G.R.; Arruda, H.F.D.; Silva, F.N.; Comin, C.H.; Amancio, D.R.; Costa, L.D.F. Principal Component Analysis: A Natural Approach to Data Exploration. *ACM Comput. Surv.* **2022**, *54*, 1–34. [CrossRef]
- Mei, X.; Wang, B.; Z, Z.; Zhao, P.; Hu, X.; Lu, G. Design of electronic nose system for perfume recognition based on support vector machine. J. Jilin Univ. Inf. Sci. Ed. 2014, 32, 355–360.
- Nakamoto, T.; Fukuda, A.; Moriizumi, T. Perfume and flavour identification by odour-sensing system using quartz-resonator sensor array and neural-network pattern recognition. *Sens. Actuators B Chem.* 1993, 10, 85–90. [CrossRef]

- 36. Hao, X.; Du, H.; Dai, X. Signal filtering. In *Measuring Control. Circuit Design and Application*; Publishing House of Electronics Industry: Beijing, China, 2018.
- Bergstra, J.; Yamins, D.; Cox, D.D. Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. In Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, USA, 17–19 June 2013; pp. 115–124.
- 38. Bergstra, J.; Bardenet, R.; Bengio, Y.; Kégl, B. Algorithms for Hyper-Parameter Optimization. In Proceedings of the 24th International Conference on Neural Information Processing Systems, Granada, Spain, 12–15 December 2011; pp. 2546–2554.

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