Article

Design and Optimization of Electronic Nose Sensor Array for Real-Time and Rapid Detection of Vehicle Exhaust Pollutants

Jin Tong [1,2]**, Chengxin Song** [1,2]**, Tianjian Tong** [3]**, Xuanjie Zong** [4]**, Zhaoyang Liu** [5]**, Songyang Wang** [5]**, Lidong Tan** [6]**, Yinwu Li** [1,2,]*** and Zhiyong Chang** [1,2,7,]***

1　College of Biological and Agricultural Engineering, Jilin University, Changchun 130022, China
2　Key Laboratory of Bionic Engineering, Ministry of Education, Jilin University, Changchun 130022, China
3　Agricultural and BioSystems Engineering, Iowa State University, Ames, IA 50010, USA
4　Zibo Municipal Transport Management Service Center, Zibo 255000, China
5　Digital Intelligent Cockpit Department, Intelligent Connected Vehicle Development Institute, China FAW Group CO., Ltd., Changchun 130013, China
6　College of Transportation, Jilin University, Changchun 130022, China
7　Weihai Institute for Bionics, Jilin University, Weihai 264401, China
*　Correspondence: lyw@jlu.edu.cn (Y.L.); zychang@jlu.edu.cn (Z.C.)

Abstract: Traditional vehicle exhaust pollutant detection methods, such as bench test and remote sensing detection, have problems such as large volume, high cost, complex process, long waiting time, etc. In this paper, according to the main components of vehicle exhaust pollutants, an electronic nose with 12 gas sensors was designed independently for real-time and rapid detection of vehicle exhaust pollutants. In order to verify that the designed electronic nose based on machine learning classification method can accurately identify the exhaust pollutants from different engines or different concentration levels from the same engine. After feature extraction of the collected data, Random Forest (RF) was used as the classifier, and the average classification accuracy reached 99.92%. This result proved that the designed electronic nose combined with RF method can accurately and sensitively judge the concentration level of vehicle exhaust pollutants.. Then, in order to enable the electronic nose to be vehicle-mounted and to achieve real-time and rapid detection of vehicle exhaust pollutants. We used Recursive Feature Elimination with Cross Validation (RFECV), Random Forest Feature Selector (RFFS) and Principal Component Analysis (PCA) to optimize the sensor array. The results showed that these methods can effectively simplify the sensor array while ensuring the RF classifier's classification recognition rate. After using RFECV and RFFS to optimize the sensor array, the RF classifier's classification recognition rate of the optimized sensor arrays for vehicle exhaust pollutants reached 99.77% and 99.44%, respectively. The numbers of sensors in the optimized sensor arrays were six and eight respectively, which achieved the miniaturization and low-cost of the electronic nose. With the limitation of six sensors, RFECV is the best sensor array optimization method among the three methods.

Keywords: vehicle exhaust pollutants; electronic nose; sensor array optimization; feature extraction; feature selection

1. Introduction

Vehicle exhaust contains hundreds of harmful substances and a large number of greenhouse gases [1]. It is of great significance to strengthen the monitoring of vehicle exhaust pollutants for environmental protection. The main components of vehicle exhaust pollutants include carbon monoxide (CO), nitrogen oxides (NO_x), hydrocarbons (HC) and sulfur dioxide (SO_2) [2].

At present, the main vehicle exhaust pollutant detection methods include remote sensing detection [3], vehicle equipment detection [4] and traditional bench test [5]. However,

these methods have many problems, such as large volume, high cost, complex process, long waiting time, etc.

An electronic nose is an odor recognition system [6–8] composed of sensor arrays, which simulates the working principle of mammalian olfactory organs. Electronic nose has been widely used in various fields, such as the food industry, chemical industry, medical field, etc. [9–13]. In previous work, it has been proved that the use of gas sensors can identify CO, NO_x, HC, SO_2 and other exhaust pollutants and judge their concentration levels [14–17]. Therefore, the electronic nose with 12 gas sensors designed independently was used in this paper to carry out real-time and rapid detection of vehicle exhaust pollutants.

In order to make it possible for the vehicle-mounted electronic nose to rapidly detect vehicle exhaust pollutants in real time, it is necessary to simplify the sensor array to achieve the miniaturization and low-cost of the electronic nose. At the same time, because the sensor array in the electronic nose has cross sensitivity, there will be redundancy in the sensor array. Optimizing the sensor array can not only reduce the volume and cost [18–20], but also remove redundant information and improve the recognition rate of the electronic nose in identifying pollutants.

At present, many feature selection methods have been used to optimize the sensor array of electronic nose, and have achieved good results. Recursive Feature Elimination (RFE) is a feature selection algorithm that searches for the optimal feature subset by repeatedly constructing models, and has been widely used in the optimization of electronic nose's sensor array [21,22]. Genetic Algorithm (GA) is a randomized search method with global optimization capability, and has also been used in sensor array optimization of electronic nose [23,24]. While Random Forest (RF) can not only solve classification and regression problems, it also has certain applications in the optimization of an electronic nose's sensor array [25]. This paper used two popular feature selection methods: Recursive Feature Elimination with Cross Validation (RFECV) and Random Forest Feature Selector (RFFS). As a contrast, this paper also used the traditional Principal Component Analysis method to optimize the sensor array, and compares the optimization results of the popular feature selection methods and the traditional method.

2. Materials and Methods

2.1. Structure of the Electronic Nose

The electronic nose system designed in this paper mainly includes a sampling unit and a detection unit. The sampling unit is mainly composed of a sampling pipe, three-way valve, flowmeter and chamber. The detection unit mainly includes a sensor array, regulating circuit board, analog digital converter (16 channel 12 bits of Beijing Pop WS-5921/U60216), computer, several connecting wires and a power supply (5V DC). The sensor converts the odor information into electrical signals through the change of the internal resistance value, then transmits it to the regulating circuit board in turn, and converts the electrical signals into digital signals through the analog digital converter, finally transmits the digital signals to the Vib'SYS signal acquisition, processing and analysis software at the computer terminal.

As the main components of vehicle exhaust pollutants are carbon monoxide (CO), nitrogen oxides (NO_x), hydrocarbons (HC), Pb compounds (Pb), carbon dioxide (CO_2), particulate emissions (PM), sulfur dioxide (SO_2) and other gases [2], this paper used 12 gas sensors that are sensitive to the above gases. The details of the selected gas sensors are shown in Table 1.

The sensor array chamber is composed of an external cavity column and an octagonal plate, respectively. The gas sensor is fixed through the hole on the octagonal plate in the same direction, and the detection surface of the gas sensor is in the inner plane of the chamber. The 3D schematic diagram of the sensor cavity structure is shown in Figure 1.

Table 1. Information of the 12 gas sensors.

Serial Number	Sensor	Response Characteristics	Producer (Country)
A	GSBT11	VOCs, toluene, benzene, formaldehyde	Ogam (Korea)
B	MP135	Ethanol, cigarette smoke, air pollutants	Winsen (China)
C	MP901	Alcohol, smoke, formaldehyde	Winsen (China)
D	WSP1110	Carbon dioxide	Winsen (China)
E	WSP2110	Toluene, benzene, alcohol, acetone	Winsen (China)
F	TGS2600	Air quality gas (hydrogen sulfide, etc.)	Figaro (Japan)
G	TGS2602	Formaldehyde and VOCs	Figaro (Japan)
H	TGS2603	Sulfur odor gas	Figaro (Japan)
I	TGS2610	Butane, liquefied gas	Figaro (Japan)
J	TGS2611	Natural gas	Figaro (Japan)
K	TGS2612	Combustible gas	Figaro (Japan)
L	TGS2620	Liquor	Figaro (Japan)
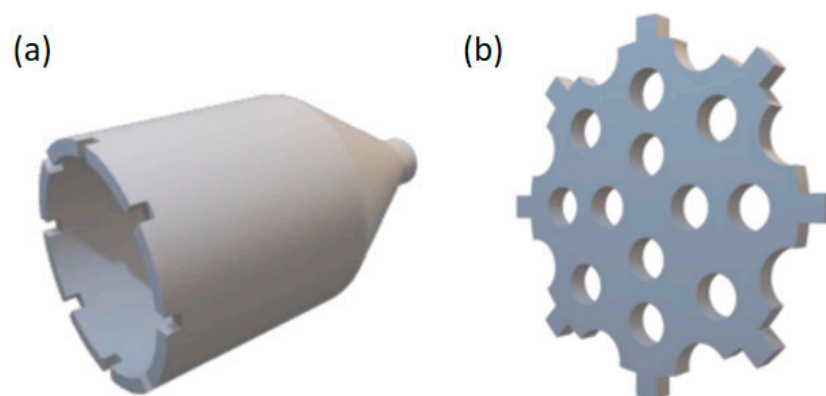
Figure 1. 3D schematic diagram of sensor cavity structure. (**a**) The schematic diagram of external cavity column. (**b**) The schematic diagram of octagonal plate.

2.2. Engine Bench Test

In order to summarize the concentration variation law of diesel engine's exhaust pollutants, we conducted the engine bench test at first. The CA4D28C5 diesel engine and the G01 gasoline engine were used in the test. The bench was set at the same rotary speed and different torques, and the exhaust pollutants concentration was measured in a continuous time. The engine's testing bench and device are shown in Figure 2.


Figure 2. Engine bench test device diagram. (**a**) Engine's testing bench. (**b**) The CA4D28C5 diesel engine testing device.

An AVL DICOM 4000 pollutant analyzer and a HORIBA MEXA-7100DEGR pollutant analyzer were used for vehicle exhaust pollutant detection, respectively. The former was

used for diesel engine pollutant detection, and the latter was used for gasoline engine pollutant detection. These two pollutant analyzers not only meet the standards of vehicle exhaust pollutants detection, but also have the advantages of moderate measurement range, high measurement accuracy, stable reading and good anti-interference performance. The two pollutant analyzers are shown in Figure 3. The changes diagrams of exhaust pollutants concentration of CA4D28C5 diesel engine measured by AVL DICOM 4000 pollutant analyzer under the operating conditions of 1600 r/min and 2200 r/min, respectively with different torques (Nm) are shown in Figure 4.
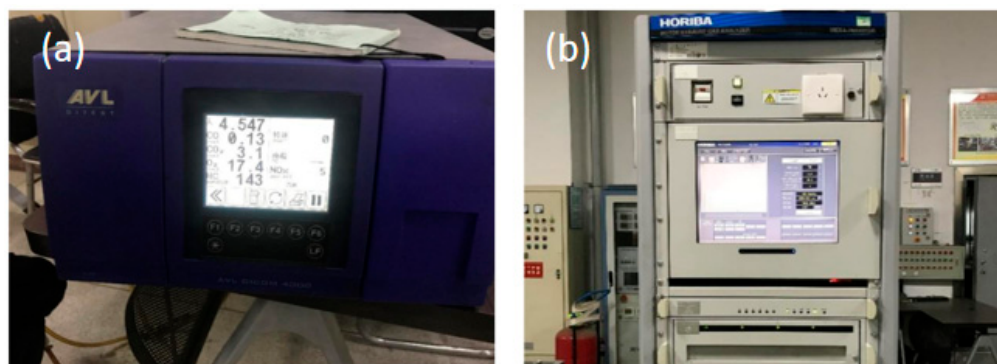
Figure 3. Physical diagram of the two pollutant analyzers. (**a**) AVL DICOM 4000 pollutant analyzer. (**b**) HORIBA MEXA-7100DEGR pollutant analyzer.
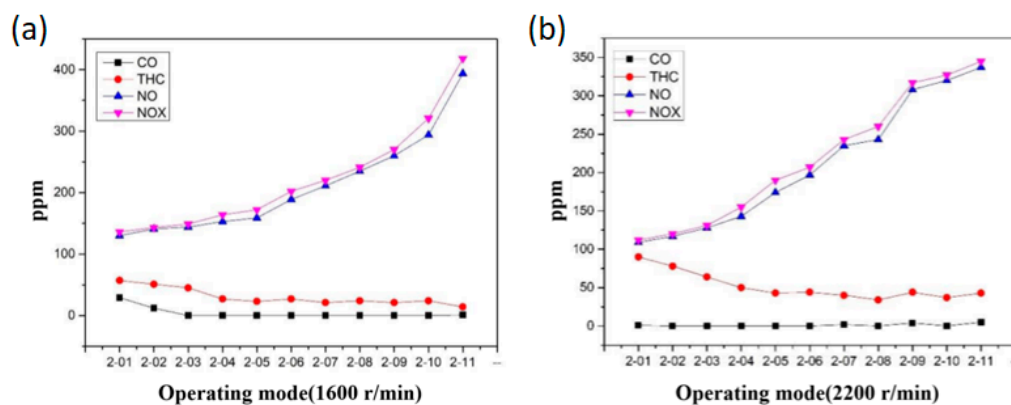
Figure 4. Exhaust pollutants concentration diagrams of CA4D28C5 diesel engine under different torques (Nm) with different rotary speeds conditions. (**a**) Variation in exhaust pollutants concentration under different torques (Nm) with 1600 r/min. (**b**) Variation in exhaust pollutants concentration under different torques (Nm) with 2200 r/min.

It can be seen from the changes in exhaust pollutants concentrations under different working conditions in Figure 4 that, at the same rotary speed, with the increase in torque, the concentrations of carbon monoxide (CO) and total hydrocarbons (THC) gradually decrease, however the concentrations of nitric oxide (NO) and nitrogen oxides (NOx) gradually increase, and their change ranges are similar; at the same torque with different rotary speeds, the change trends of CO, THC, NO and NOx are similar, but the concentrations are different.

According to the above concentration variation law of the diesel engine's exhaust pollutants, we believe that the sensor array in the electronic nose needs to have the ability to accurately identify the exhaust pollutants from different engines or different concentration levels from the same engine. The test scheme developed according to the concentration variation law of the diesel engine's exhaust pollutants obtained from the above engine bench test is shown in Section 2.3.

2.3. Experimental Setups

According to the concentration variation law of the diesel engine's exhaust pollutants in Section 2.2, we designed five groups of experiments using electronic nose to detect vehicle exhaust pollutants. The first four groups used a CA4D28C5 diesel engine, which was tested under the conditions of the same rotary speed with different torques. That is, the experiments were conducted under the conditions that the types of exhaust pollutants gases were the same, but the concentration levels of each pollutant were only slightly different. The fifth group used G01 gasoline engine, which means that the experimental conditions of different exhaust pollutants gas types and different concentration levels were taken as the control group.

The total test time of each sample was 300 s, of which 90 s is the data acquisition time, and the data acquisition frequency was 50 Hz, so that each sample contains $12 \times 90 \times 50 = 54{,}000$ data points. The other 210 s is the cleaning time of the electronic nose chamber and the zero-setting time of the resistance.

The actual experimental steps are as follows:

(1)   Check the connection tightness and safety of each component.
(2)   Power on, start the electronic nose detection system, warm up for 30 min, and expose the sensor to clean air.
(3)   The sample gas is introduced into the electronic nose's chamber through the catheter connected with the three-way valve.
(4)   Collect and store the signal of the sample gas.
(5)   After the data collection of one group of samples is completed, clean the electronic nose with clean air for about 210 s.
(6)   Perform the next set of experiments and repeat steps (2–5).

A total of 225 samples were obtained in the experiments. The concentration levels of the main exhaust pollutants under the test conditions and corresponding operating conditions of the samples are shown in Table 2.

Table 2. Composition of experimental samples.

Experimental Number	Engine	Rotary Speed (r/min)	Torque (Nm)	CO (%vol)	CO_2 (%vol)	THC (ppm)	Number of Samples
1		2000	29	0.03	9.6	96	45
2		2000	41.5	0.08	6.1	90	45
3	Diesel engine	2000	103.9	0.02	9.0	54	45
4		2000	153	0.01	9.7	42	45
5	Gasoline engine	1200	67	1200	926	1.38	45

2.4. Feature Extraction

After completing the data acquisition steps in Section 2.3, in order to reduce the data dimension and ensure the effectiveness of the subsequent pattern recognition algorithm [26], we extracted four features from each data sample obtained: Maximum Value (MAX), Average Value (Mean), Integral Value (IV) and Wavelet Transform (WT). MAX reflects the steady state information of the whole gas sensor response curve. Mean and IV combine all the information of the whole gas sensor response curve. WT can better reflect the transient information of the whole gas sensor response curve. MAX, Mean, IV and WT were extracted from each data sample obtained from 12 sensors. After feature extraction, each data sample changes from a data sample containing 54,000 data points to a feature sample containing only 12 data points. The feature samples extracted from 225 samples were spliced together to obtain a feature matrix containing 12 feature vectors and each feature vector contains 225 feature values.

2.5. Sensor Array Optimization

In order to enable the electronic nose to realize real-time and rapid detection of vehicle exhaust pollutants, it is necessary to make the electronic nose more miniaturized and low-cost. On the one hand, the sensor array optimization method can simultaneously realize the miniaturization and low-cost of the electronic nose. On the other hand, due to the cross sensitivity of gas sensors, sensor array optimization can reduce the training time of classification models, improve the recognition rate of classification models, and avoid the occurrence of over fitting problems. Three sensor array optimization (i.e., feature selection) methods based on different principles and optimization strategies were used in this paper: Recursive Feature Elimination with Cross Validation (RFECV) based on the packaging method, Random Forest Feature Selector (RFFS) based on an embedding method, and traditional Principal Component Analysis (PCA) as the comparison. They are briefly introduced below.

2.5.1. Sensor Array Optimization Based on RFECV

Recursive Feature Elimination (RFE) was proposed by Guyon [27] and has been widely used in solving feature selection problems. RFECV is a feature selection process for recursive feature elimination in the cross-validation cycle [28], which can automatically find the feature subset with the optimal number of features to obtain feature selection results. In this paper, Random Forest (RF) was used as the classification model in the process of recursive feature elimination. In the recursive step of each iteration, remove the last feature according to feature ranking, retrain the RF model with the retained features, and cross verify the performance of the RF model until there is only one feature left. Finally, according to the performance of RF model in different feature numbers and feature subsets composed of different features, the optimal number of features and the optimal feature subset can be obtained.

2.5.2. Sensor Array Optimization Based on RFFS

Random forests can not only deal with classification and regression problems, but also can be used to evaluate and select features because they can estimate the importance of features [29]. The principle of using random forests to evaluate and select features is based on the difference between the classification performance of random forest on the original dataset and the randomly extracted dataset. By calculating the classification performance difference of each decision tree in the random forest on different randomly extracted datasets, the importance of features can be estimated and the feature ranking can be obtained. The importance of the features is estimated by Equations (1) and (2):

The importance of feature A_j is estimated as:

$$I\left(A_j\right) = \sum \frac{d_i}{n \times SEd_i},\tag{1}$$

where d_i represents the performance difference of decision tree i and SEd_i represents the standard error of all decision trees:

$$SEd_i = \frac{SDd_i}{\sqrt{n}},\tag{2}$$

where SDd_i is the standard deviation of d_i and n is the number of elements in the dataset.

2.5.3. Sensor Array Optimization Based on PCA

Principal Component Analysis is a dimension reduction method, which transforms the original multivariable in high-dimensional space into a set of linear independent comprehensive indexes in low-dimensional space through orthogonal transformation [30]. The eigenvalues obtained through Principal Component Analysis are sorted from large to small to measure the importance of features. Finally, features are selected according to the importance of features [18].

Assuming that $X = (x_1, x_2, \ldots, x_{12})$ is the original variable (the feature extracted from 12 gas sensors), $Z = (z_1, z_2, \ldots, z_p)$, $(p \leq 12)$ as the comprehensive indexes in the low dimensional space. The transformation process from the original variable matrix X to the comprehensive index matrix Z can be expressed as:

$$
\begin{aligned}
z_1 &= a_{1,1}x_1 + a_{1,2}x_2 + \ldots + a_{1,12}x_{12} \\
z_2 &= a_{2,1}x_1 + a_{2,2}x_2 + \ldots + a_{2,12}x_{12} \\
&\vdots \\
z_p &= a_{p,1}x_1 + a_{p,2}x_2 + \ldots + a_{p,12}x_{12}
\end{aligned}
\tag{3}
$$

where $a_{p,12}$ represents the 12th coefficient in the p-th comprehensive index.

Calculate the absolute value of the sum of the coefficients of the original variable corresponding to each feature in all comprehensive indexes, such as the absolute value of the sum of coefficients of the feature values extracted from the 12th sensor in all comprehensive indexes A_{12}:

$$
z_2 = |a_{1,12}| + |a_{2,12}| + \ldots + |a_{p,12}|.
\tag{4}
$$

The absolute value of the sum of the coefficients of each feature is used to represent the contribution degree of each feature in the comprehensive indexes, and then the feature selection can be obtained by descending order.

3. Results and Discussion

Different sensor array optimization methods will result in different sensor array optimization results, in which the number and combination of sensors in the sensor array will be different. The experimental results of this paper were obtained by taking the original data and the data after sensor array optimization as the input data of the Random Forest (RF) classifier. The classification recognition rate of the test set is the main index to evaluate the effectiveness of the original sensor array and the optimized sensor array. Therefore, this paper used the classification recognition rate of the test set to evaluate the results of sensor array optimization. In order to make the classification results more reliable and credible, the stratified sampling strategy was used to conduct 3-fold cross-validation 100 times, and the average of the 300-test set classification recognition rate was calculated as the final classification recognition rate of the test set. In the sensor array optimization stage, the original data set was divided into 2/3 training set and 1/3 test set; the test set was not used in the sensor array optimization stage.

The classification recognition rate obtained using RF without sensor array optimization is shown in Figure 5a. It can be seen that the four feature extraction methods have achieved high classification recognition rate when using RF as the classifier. The highest MAX has an average classification recognition rate of 99.92%, and the WT with the lowest average classification recognition rate has also reached 98.16%. This shows that the original sensor array is effective and has the ability to accurately identify exhaust pollutants from different engines or different concentration levels from the same engine.

The results of sensor array optimization of the four extracted eigenvalues based on RF model and RFECV method are shown in Table 3. The optimized sensor array includes six gas sensors: MP135, TGS2600, TGS2610, TGS2611, TGS2612, and TGS2620. The optimization of sensor array based on RFECV method has achieved good results. Figure 5b shows that Mean and IV, which have the lowest average classification recognition rate after the optimization of sensor array, still reached 97.94%, almost without any loss. This shows that the RFECV method is very effective for sensor array optimization.

The classification recognition rate of sensor array optimization using RFFS is shown in Figure 5c. Compared with the RFECV method, RFFS retains a total of eight gas sensors. In addition to the six same gas sensors selected in the RFECV method, it also retains two gas sensors, GSBT11 and TGS2602. It can be seen from the comparison of average classification recognition rate between Figure 5b,c that the classification recognition rate of RFFS with eight gas sensors is slightly improved in Mean and IV compared with RFECV with six gas

sensors, but it is decreased in MAX with the highest classification recognition rate, and it also need to bear the cost of increasing the development of two gas sensors.
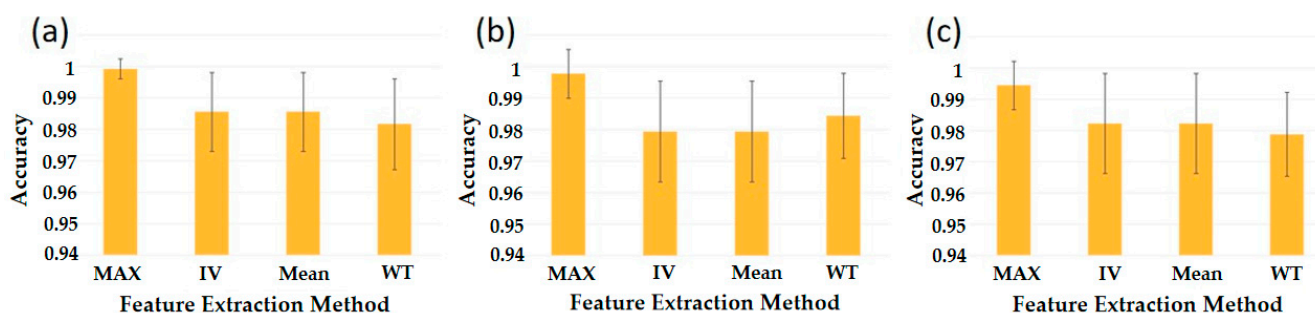
Figure 5. (**a**) The RF classifier's average classification recognition rate of the four feature extraction methods without sensor array optimization. (**b**) The RF classifier's average classification recognition rate of the four feature extraction methods after using RFECV as the sensor array optimization method. (**c**) The RF classifier's average classification recognition rate of the four feature extraction methods after using RFFS as the sensor array optimization method. (The error bars represent the standard deviations of the recognition rates, and the calculation formula is $\sigma = \sqrt{\left((x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2\right)/n}$, where σ represents the standard deviation, x_1, $x_2, \ldots, x_n$ represent recognition rates, $\overline{x}$ represents the average recognition rate, n is the number of recognition rates. In this paper, we used 3-fold cross validation 100 times, so the value of n is 300.)

Table 3. The sensor array optimization results of RFECV based on RF.

Feature Extraction Method	Number of Sensors after Optimization	Sensor Serial Number Retained after Optimization
MAX Mean IV WT	6	B, E, H I, K, L

Using PCA to optimize sensor arrays requires setting corresponding thresholds to limit the absolute value of the sum of coefficients of each feature. In this paper, in order to compare with RFECV and RFFS, the threshold values are set at the values required when six and eight gas sensors were reserved. The sensor array optimization results and the corresponding RF classification recognition rate is shown in Table 4. Obviously, when the number of optimized sensors is limited, the classification recognition rate of sensor array optimized by PCA is worse than that of other sensor array optimization methods. When eight sensors are retained, the classification recognition rate of MAX decreased by more than 2% compared with RFFS. When six sensors are retained, the classification recognition rate of MAX, Mean and IV declined compared with RFECV.

When PCA was used as the sensor array optimization method, TGS2600, TGS2603, TGS2610, TGS2611, TGS2612 and TGS2620 were selected to be retained in almost every feature extraction method, regardless of whether six or eight gas sensors were retained, which indicates that they have a good response to vehicle exhaust pollutants. However, MP135 and MP901, which were retained for many times when eight gas sensors were retained, were rarely selected when six gas sensors were retained. This may be because the target gases detected by them overlap with the six gas sensors frequently selected above. When different feature extraction methods retained the same number of gas sensors, the main reason why the gas sensors selected for retention were different and the main reason why the sensor array optimization using PCA method was not as effective as the other two sensor array optimization methods may be because it is an unsupervised dimension reduction method, and its realization method is to maximize the variance in the projection direction, so the category information is not fully utilized.

Table 4. The sensor array optimization results and the classification recognition rate based on RF after using PCA.

Feature Extraction	Number of Sensors after Optimization	Sensor Serial Number Retained after Optimization	Recognition Rate (%)
MAX		A, C, F, H, I, J, K, L	97.12
Mean	8	B, C, F, H, I, J, K, L	98.34
IV		B, C, F, H, I, J, K, L	98.34
WT		B, F, G, H, I, J, K, L	98.59
MAX		A, F, I, J, K, L	97.44
Mean	6	F, H, I, J, K, L	96.26
IV		F, H, I, J, K, L	96.26
WT		B, F, H, J, K, L	98.63

When the number of sensors retained after using different methods to optimize the sensor array is the same, the sensor array with higher classification recognition rate is better. When the classification recognition rate of sensor arrays is the same, the sensor array with fewer sensors is better. A good sensor array needs to achieve the highest recognition rate when the number of sensors is as small as possible. Considering that the recognition rate of the sensor array composed of six gas sensors is almost no lower than that of the original sensor array composed of 12 gas sensors, and the sensor array composed of six gas sensors can reduce volume and save the cost to make it possible for the vehicle-mounted electronic nose to rapidly detect vehicle exhaust pollutants in real time. We believe that the optimal number of sensors is to retain six gas sensors. After limiting the number of sensors in the sensor array to six, the recognition rate of each sensor array optimization method is shown in Table 5.

Table 5. The classification recognition rate based on RF after using different methods (limited the number of sensors in the sensor array to six).

Feature Extraction	Sensor Array Optimization Method	Recognition Rate (%)
MAX	RFECV	99.77
	RFFS	99.77
	PCA	97.44
Mean	RFECV	97.94
	RFFS	97.94
	PCA	96.26
IV	RFECV	97.94
	RFFS	97.94
	PCA	96.26
WT	RFECV	98.44
	RFFS	98.33
	PCA	98.63

In this case, the average RF classification recognition rate of MAX after using RFECV and RFFS for sensor array optimization has both reached 99.77%. The classification recognition rate of WT after using RFECV was higher than using RFFS. It means that RFECV is a

better sensor array optimization method, and MAX is a better feature extraction method than the other three. In addition, the sensor array optimized by RFECV method and using MAX as the feature extraction method includes sensors: MP135, TGS2600, TGS2610, TGS2611, TGS2612 and TGS2620.

The cost of using only the above six gas sensors is \$32, which can save about 56% of the cost compared with the original 12 sensor arrays. It can also greatly reduce the volume of the electronic nose to achieve the purpose of miniaturization. The average time of using MAX as the feature extraction after using RFECV to test a new real sample was 0.021 s (using Python 3.10.5 and Visual Studio Code 2022). The miniaturization and the rapid detection time make it possible for the vehicle-mounted electronic nose to rapidly detect vehicle exhaust pollutants in real time.

4. Conclusions

In this paper, a self-designed electronic nose composed of 12 gas sensors was used to detect vehicle exhaust pollutants from different engines or the same engine at different concentration levels. Firstly, we conducted an engine bench test to summarize the concentration variation law vehicle exhaust pollutants. After analyzing the experimental data and extracting the features, the highest RF classification recognition rate was up to 99.92% without optimizing the sensor array. In order to enable the vehicle-mounted electronic nose to quickly detect vehicle exhaust pollutants in real time, reduce the volume, save development cost, and save detection time, we used RFECV, RFFS, and PCA to optimize the sensor array. When the number of sensors was not fixed, the classification recognition rate of MAX after using RFECV and RFFS methods reached 99.77% and 99.44% respectively, while RFECV retained less sensors. When the number of sensors was limited to six, the classification recognition rate of MAX after using RFFS was up to 99.77%, as high as that of RFECV. The classification recognition rate of WT after using RFECV was higher than using RFFS, which means RFECV is a better sensor array optimization method in this case. The cost of the sensor array optimized by RFECV method and using MAX as the feature extraction method is only 32\$ with almost no loss of recognition rate. In addition, its average detection time of a new sample was 0.021 s.

In summary, through the research in this paper, we found that it is feasible to use electronic nose for real-time and rapid detection of vehicle exhaust pollutants, and electronic nose has a good application prospect in this area. At the same time, combined with the sensor array optimization methods, the electronic nose can be miniaturized and low-cost, which makes it possible to detect vehicle exhaust pollutants in real time. In the future, we will use the vehicle-mounted electronic nose in combination with the developing edge computing and cloud computing technologies to develop corresponding cloud calculating platforms. We can monitor vehicle exhaust pollutants in real time, and make corresponding predictions regarding air quality.

Author Contributions: Conceptualization, Z.C.; methodology, J.T., Z.C., Y.L. and L.T.; software, C.S., X.Z., Z.L. and S.W.; validation, C.S., X.Z., Z.L. and S.W.; formal analysis, J.T., Y.L. and L.T.; investigation, Z.C., Y.L. and L.T.; resources, X.Z., Z.L. and S.W.; data curation, Y.L., L.T., X.Z., Z.L. and S.W.; writing—original draft preparation, C.S., T.T., X.Z., Z.L. and S.W.; writing—review and editing, J.T., Y.L. and L.T.; visualization, C.S., T.T. and X.Z.; supervision, J.T. and Z.C.; project administration, X.Z., Z.L. and S.W.; funding acquisition, X.Z., Z.L. and S.W. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

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

References

1. Twigg, M. Progress and future challenges in controlling automotive exhaust gas emissions. *Appl. Catal. B* **2007**, *70*, 3523. [CrossRef]
2. Šarkan, B.; Kuranc, A.; Kučera, Ľ. Calculations of exhaust emissions produced by vehicle with petrol engine in urban area. In Proceedings of the 4th International Conference of Computational Methods in Engineering Science (CMES), Kazimierz Dolny, Poland, 21–23 November 2019.
3. Zhang, Y.; Stedman, D.H.; Bishop, G.A.; Guenther, P.L.; Beaton, S.P.; Peterson, J.E. On-road hydrocarbon remote sensing in the Denver area. *Environ. Sci. Technol.* **1993**, *27*, 1885–1891. [CrossRef]
4. Tutuianu, M.; Bonnel, P.; Ciuffo, B.; Haniu, T.; Ichikawa, N.; Marotta, A.; Pavlovic, J.; Steven, H. Development of the World-wide harmonized Light duty Test Cycle (WLTC) and a possible pathway for its introduction in the European legislation. *Transp. Res. D Transp. Environ.* **2015**, *40*, 61–75. [CrossRef]
5. Guo, W.; Qiong, W.; Zhao, M. Exhaust emissions of diesel engines with nano-copper additives. *Appl. Nanosci.* **2020**, *10*, 1045–1052. [CrossRef]
6. Jeong, S.Y.; Kim, J.S.; Li, J.H. Rational design of semiconductor-based chemiresistors and their libraries for next-generation artificial olfaction. *Adv. Mater.* **2020**, *32*, 2002075. [CrossRef]
7. Lee, J.M.; Devaraj, V.; Jeong, N.N.; Lee, Y.; Kim, Y.J.; Kim, T.; Yi, S.H.; Kim, W.G.; Choi, E.J.; Kim, H.M.; et al. Neural mechanism mimetic selective electronic nose based on programmed M13 bacteriophage. *Biosens. Bioelectron.* **2022**, *196*, 113693. [CrossRef] [PubMed]
8. Machungo, C.; Berna, A.Z.; McNevin, D.; Wang, R.; Trowell, S. Comparison of the performance of metal oxide and conducting polymer electronic noses for detection of aflatoxin using artificially contaminated maize. *Sens. Actuators B.* **2022**, *360*, 131681. [CrossRef]
9. Chang, Z.Y.; Li, J.H.; Qi, H.Y.; Ma, Y.H.; Chen, D.H.; Xie, J.; Sun, Y.H. Bacterial infection potato tuber soft rot disease detection based on electronic nose. *Open Life Sci.* **2017**, *12*, 379–385. [CrossRef]
10. Zhao, R.S.; Kong, C.; Ren, L.Q.; Sun, Y.H.; Chang, Z.Y. Real-time monitoring of the oil shale pyrolysis process using a bionic electronic nose. *Fuel* **2022**, *313*, 122672. [CrossRef]
11. Weng, X.H.; Sun, Y.H.; Xie, J.; Deng, S.H.; Chang, Z.Y. Bionic Layout Optimization of Sensor Array in Electronic Nose for Oil Shale Pyrolysis Process Detection. *J. Bionic Eng.* **2021**, *18*, 441–452. [CrossRef]
12. Tirzïte, M.; Bukovskis, M.; Strazda, G.; Jurka, N.; Taivans, I. Detection of lung cancer with electronic nose and logistic regression analysis. *J. Breath Res.* **2018**, *13*, 016006. [CrossRef] [PubMed]
13. Dragonieri, S.; Pennazza, G.; Carratu, P.; Resta, O. Electronic nose technology in respiratory diseases. *Lung* **2017**, *195*, 157–165. [CrossRef] [PubMed]
14. Kang, M.; Cho, I.; Park, J.; Jeong, J.; Lee, K.; Lee, B.; Henriquez, D.D.; Yoon, K.; Park, I. High Accuracy Real-Time Multi-Gas Identification by a Batch-Uniform Gas Sensor Array and Deep Learning Algorithm. *ACS Sens.* **2022**, *7*, 430–440. [CrossRef] [PubMed]
15. Arroyo, P.; Meléndez, F.; Suárez, J.I.; Herrero, J.L.; Rodríguez, S.; Lozano, J. Electronic nose with digital gas sensors connected via bluetooth to a smartphone for air quality measurements. *Sensors* **2020**, *20*, 786. [CrossRef] [PubMed]
16. Kun, C.; He, Y.; Li, Y.; Ng, A.; Yu, J. A Room Temperature Hydrocarbon Electronic Nose Gas Sensor Based on Schottky and Heterojunction Diode Structures. *IEEE Electron Device Lett.* **2019**, *41*, 163–166. [CrossRef]
17. Zhang, W.; Tian, F.; Song, A.; Hu, Y. Research on electronic nose system based on continuous wide spectral gas sensing. *Microchem. J.* **2018**, *140*, 1–7. [CrossRef]
18. Sun, H.; Tian, F.C.; Liang, Z.F.; Sun, T.; Yu, B.; Yang, S.X.; He, Q.H.; Zhang, L.L.; Liu, X.M. Sensor array optimization of electronic nose for detection of bacteria in wound infection. *IEEE Trans. Ind. Electron.* **2017**, *64*, 7350–7358. [CrossRef]
19. Yin, Y.; Yu, H.C.; Chu, B.; Xiao, Y.J. A sensor array optimization method of electronic nose based on elimination transform of Wilks statistic for discrimination of three kinds of vinegars. *J. Food Eng.* **2014**, *127*, 43–48. [CrossRef]
20. Bag, A.K.; Tudu, B.; Roy, J.; Bhattacharyya, N.; Bandyopadhyay, R. Optimization of sensor array in electronic nose: A rough set-based approach. *IEEE Sens. J.* **2011**, *11*, 3001–3008. [CrossRef]
21. Yan, K.; Zhang, D. Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sens. Actuators B* **2015**, *212*, 353–363. [CrossRef]
22. Li, P.; Niu, Z.; Shao, K.; Wu, Z. Quantitative analysis of fish meal freshness using an electronic nose combined with chemometric methods. *Measorement* **2021**, *179*, 109484. [CrossRef]
23. Xu, Z.; Shi, X.; Lu, S. Integrated sensor array optimization with statistical evaluation. *Sens. Actuators B* **2010**, *149*, 239–244. [CrossRef]
24. Jia, P.; Tian, F.; Fan, S.; He, Q.; Feng, J.; Yang, S.X. A novel sensor array and classifier optimization method of electronic nose based on enhanced quantum-behaved particle swarm optimization. *Sens. Rev.* **2014**, *34*, 304–311. [CrossRef]

25. Wei, G.; Zhao, J.; Yu, Z.; Feng, Y.; Li, G.; Sun, X. An effective gas sensor array optimization method based on random forest. In Proceedings of the 17th IEEE SENSORS Conference, New Delhi, India, 28–31 October 2018.

26. Yan, J.; Guo, X.; Duan, S.; Jia, P.; Wang, L.; Peng, C.; Zhang, S. Electronic nose feature extraction methods: A review. *Sensors* **2015**, *15*, 27804–27831. [CrossRef] [PubMed]

27. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **2002**, *46*, 389–422. [CrossRef]

28. Wang, S.; Chen, S. Insights to fracture stimulation design in unconventional reservoirs based on machine learning modeling. *J. Pet. Sci. Eng.* **2019**, *174*, 682–695. [CrossRef]

29. Čehovin, L.; Bosnić, Z. Empirical evaluation of feature selection methods in classification. *Intell. Data. Anal.* **2010**, *14*, 265–281. [CrossRef]

30. Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. A* **2016**, *374*, 20150202. [CrossRef]