*Journal of Network Intelligence*
**Taiwan Ubiquitous Information**
©2025 ISSN 2414-8105
Volume **10**, Number **1**, February **2025**

# Remote Vehicle Exhaust Detection Based on Integrated Deep Learning Models for Environmental Monitoring

Lu Xiao*

Jiangxi Environmental Engineering Vocational College, Ganzhou 341000, P. R. China
xiaolugz@126.com

Fang Deng

Jiangxi College of Applied Technology, Ganzhou 341000, P. R. China
Universiti Teknologi MARA, Shah Alam 40450, Malaysia
tabee1500@163.com

*Corresponding author: Lu Xiao
Received March 25, 2024, revised July 11, 2024, accepted September 12, 2024.

Abstract. *The traditional equipment for tailpipe emission monitoring basically adopts the static fixed-point detection mode. Not only does this detection mode fail to ensure accurate analysis of vehicle exhaust emissions during dynamic operation, but its measurement results are also susceptible to interference from human intervention or the detection environment. Therefore, a remote vehicle exhaust gas detection system based on an integrated deep learning model is proposed. Firstly, the TLZ7x-EasyEVM development board designed based on the SOM-TLZ7x core board is used as the on-board hardware platform, and the FS704UM network module is interconnected with the development board through the GPIO interface to realise wireless remote data transmission. The NHA-502 exhaust gas analyser was used to detect the components of CO, CO2, O2 and NOx in the exhaust gas, and the performance parameters of the vehicle were monitored in real time through the OBD-II interface. Then, in order to further improve the efficiency of feature information extraction by sliding window, an incremental computation-based feature information calculation method is proposed to extract feature information from preprocessed vehicle exhaust data. Secondly, using the extracted seven types of feature information as inputs in the tailpipe monitoring centre, the LSTM neural network, which has the ability to remember long-term temporal information, is used as the weak predictor of integrated learning, and the weak predictor is weighted and combined with the strong predictor using the AdaBoost integrated learning algorithm. The effectiveness of the proposed system is verified by experiments on an Audi A4 vehicle. The results show that the proposed system achieves an accuracy of 91.33% for the detection of abnormal exhaust conditions.*
**Keywords:** Environmental monitoring; Vehicle exhaust; Remote detection; Integrated learning; Deep learning; LSTM; AdaBoost

1. **Introduction.** Currently, the most commonly used means of vehicle exhaust gas detection is the static fixed-point detection method [1, 2]. However, due to the actual operating conditions of the engine is complex and variable, coupled with the influence of ambient temperature, fuel quality, wear and tear, component corrosion aging and other factors. The traditional static fixed-point detection mode not only can not ensure accurate analysis of vehicle exhaust emissions in the dynamic operation process [3], but also its

measurement results are susceptible to human interference or the interference of the detection environment, resulting in part of the emissions exceeding the standards of the vehicle through the detection of driving on the road. In the long run, the strength of automobile exhaust emission control will inevitably become stronger and stronger, therefore, in order to effectively make up for the shortcomings of the static fixed-point detection method, the research and development of remote dynamic on-board automobile exhaust emission detection technology that can detect the exhaust emission status of the automobile when it is running has become very urgent. However, due to the use of cost, technology and other aspects of the reasons [4, 5], the current market related to the study of practical technology is still in the primary stage, there is still a large distance from the market application.

The OBD (On-Board Diagnostics) system determines whether the current vehicle emissions are normal or not based on the electrical parameters of various types of sensors and actuators obtained from the vehicle's electronic control system [6, 7], and will immediately illuminate the malfunction light once an abnormality is detected. However, the OBD system essentially adopts the method of indirect measurement of engine exhaust emissions. This method is relatively inexpensive and easy to install, but the accuracy and reliability of its measurement results are difficult to accurately assess. Today, when the global vehicle ownership reaches hundreds of millions of scale, the tailpipe emission is a very large value, so it is very necessary to continue to improve the accuracy of tailpipe emission measurement [8]. Currently, fixed-point static analyses are mainly used to improve the measurement accuracy of tailpipe emission pollutants. Commonly used methods include electrical and electrochemical methods. For example, thermal conductivity chromatography [9], layer analysis [10], and optical absorption spectroscopy [11].

With the rapid development of computer technology, machine learning algorithms are widely used in various fields, and exhaust gas detection is no exception. Traditional exhaust gas detection methods usually rely on chemical analysis and manual operation, which is time-consuming and labour-intensive, and the accuracy of the results needs to be improved. Machine learning algorithms, however, can automatically learn the rules from a large amount of data, so as to detect and analyse the composition of vehicle exhaust in real time and accurately. Based on the labelled exhaust data, supervised learning algorithms (e.g., support vector machine, decision tree, etc.) or deep learning algorithms (e.g., convolutional neural network) can be used to train a model that can accurately identify the components of exhaust gas. The purpose of this work is to build a vehicle hardware device with 4G network module for data acquisition and remote transmission, and deploy the trained model in the vehicle or roadside exhaust monitoring centre to achieve real-time and accurate monitoring of vehicle exhaust, which provides the basis for subsequent environmental management and decision-making.

1.1. **Related Work.** Currently, the main task of existing exhaust gas analysis techniques is the quantitative measurement of pollutants emitted by engines.

Li et al. [12] proposed an SVM-based algorithm for identifying abnormalities in exhaust emissions of petrol engines in response to the problem that the exhaust emissions of petrol engines change nonlinearly with the change of working conditions. The algorithm is based on a genetic algorithm to intelligently optimise the SVM parameter structure, which improves the effectiveness of the exhaust gas identification model. Neural network is also a commonly used pattern recognition method for nonlinear fitting modelling. Bu [13] proposed an automotive engine exhaust condition analysis and fault diagnosis model based on Elman neural network, which takes the relationship between the emission components of the engine and the misfire condition as the training samples, and achieves the exhaust gas identification of the engine in normal, mild misfire and severe misfire conditions. Ceviz et al. [14] found that by recording the real-time dynamic data during vehicle operation, and then applying the data flow analysis method based on the one-dimensional wavelet analysis function to comprehensively analyse the input and output signal data and the working state of the sensors and actuators in the engine control system, it is possible to effectively determine whether the engine emissions are exceeding the standard and the corresponding causes of failure. However, the engine itself in a variety of different working conditions of the emission law is different. For example, under idling conditions, CO and CO2 emissions are high and NOX emissions are low; under heavy load conditions, NOX emissions are high. Therefore, only measuring the pollutant emissions of an engine under a particular operating condition is not comprehensive. It is necessary to combine with the real-time working conditions of the engine in order to obtain a more objective and effective emission status. At present, the research work in this area is still relatively small.

In the context of deep learning, Chung and Kim [15] proposed a real-time vehicle emission monitoring model that utilises data collected by low-cost gas sensors. The model uses a Long Short-Term Memory (LSTM) network to learn patterns in time-series data to predict the concentrations of different gases. Experiments have shown that the model has high accuracy in predicting the major tailpipe components such as CO, CO2 and NO. However, the model is mainly for known types of gases and has limited ability to detect novel pollutants. In addition, the data quality may be affected to some extent due to the use of low-cost sensors. Yu et al. [16] proposed an interpretable deep learning model for estimating vehicle emissions. The researchers used an attention mechanism and visualisation techniques to enable the model not only to accurately predict emissions but also to explain the main factors affecting emissions. This provides valuable insights for developing emission reduction policies and optimising vehicle design. However, the training and deployment process of the model is relatively complex and requires significant computational resources. In addition, there may be a trade-off between interpretability and accuracy, requiring further optimisation of the model architecture.

1.2. **Motivation and contribution.** Existing exhaust gas detection methods mainly use a single deep learning method, a single integrated learning method, coupled data pre-processing methods, etc., which have poor prediction accuracy, stability and generalisation performance. However, AdaBoost [17, 18] is able to better fit complex data distributions by combining multiple weak classifiers, and the integrated model usually improves the prediction accuracy of the concentration of tailpipe components compared to a single LSTM model. Since the AdaBoost algorithm does not have many assumptions on the distribution of the training data, the integrated model may have better adaptive ability and generalisation performance when dealing with exhaust data from different scenarios and car models compared to the single LSTM model [19, 20]. The main innovations and contributions of this work include:

(1) An in-vehicle hardware platform based on the TLZ7x-EasyEVM development board was constructed, and 4G network transmission was carried out through FS704UM. The NHA-502 tail gas analyser was used as the signal input platform for the on-board tail gas platform.

(2) In order to further improve the efficiency of feature information extraction by sliding window and to solve the problem of excessive memory occupied by the device during traffic processing, a feature information calculation method based on incremental computation is proposed, which can be carried out on a dynamic number of data streams.

(3) In order to effectively improve the problems of low accuracy and insufficient generalisation ability of real-time detection of exhaust gas dynamics, an improved LSTM

model based on the integrated learning algorithm is proposed. In the exhaust gas monitoring centre, seven types of extracted feature information are used as inputs, and the LSTM neural network with the ability to remember long-term temporal information is used as the weak predictor of integrated learning, and the weak predictor is weighted and combined with the strong predictor using the AdaBoost integrated learning algorithm.

## 2. System hardware selection.

**2.1. System platform.** The on-board hardware platform is a specialised computer system, which mainly consists of a central processor, internal and external memories and input/output devices. In order to improve its development efficiency, it is generally preferable to develop based on mature embedded development boards. When the system's hardware and software are all debugged, then the hardware platform is finally customised with circuit boards.

TLZ7x-EasyEVM is a development board designed based on the SOM-TLZ7x core board [21]. The backplane is designed as a 4-layer board with a lead-free process and provides users with a test platform for the SOM-TLZ7x core board. The SOM-TLZ7x core board adopts the general-purpose MIPSII instruction set, with a main frequency of up to 266 MHz. The SOM-TLZ7x core board integrates all kinds of mainstream internal and external memory interfaces (such as SDRAM interface, NOR FLASH/ROM, and NAND FLASH interface), and supports interfaces such as UART, USB 2.0, CAN, Camera, SD, and Gigabit Ethernet.

To facilitate pre-debugging and post-customisation, the core board of the SOM-TLZ7x mainly integrates the central processor, main memory, external memory and clock source. The backplane integrates serial/parallel interfaces, USB interface, 10/100M LAN interface, four-wire resistive touch screen interface, LCD interface and other I/O resources. The appearance of the SOM-TLZ7x core board development board is shown in Figure 1.

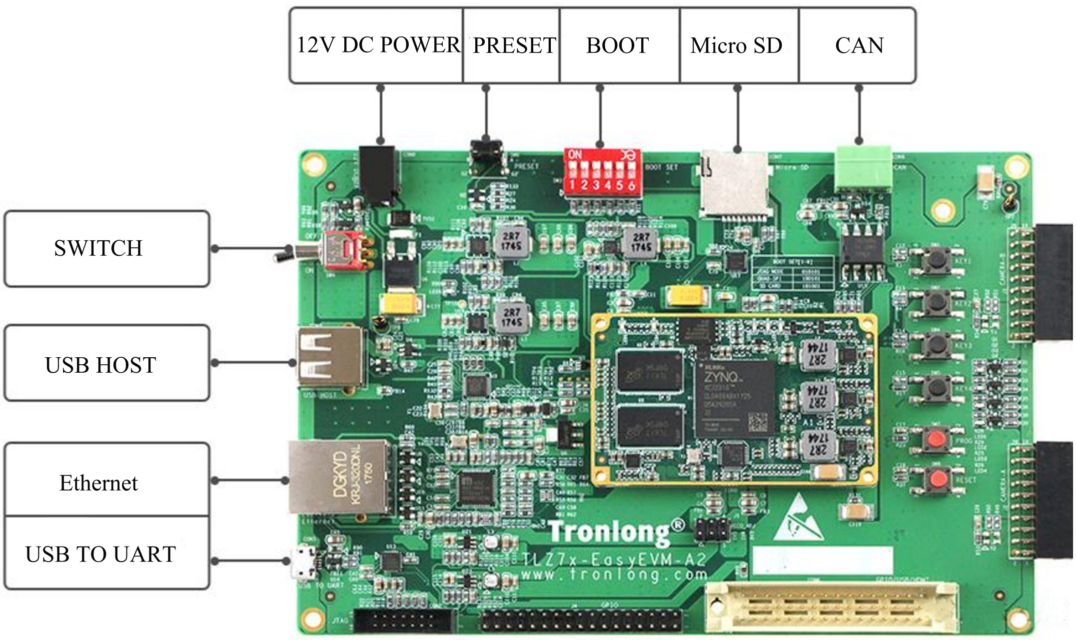


Figure 1. SOM-TLZ7x Development Board

**2.2. 4G network interface.** A large number of mature 4G network interface expansion modules have been released in the market, and in this paper, we directly use the FS704UM [22]. The FS704UM is fully compatible with the TLZ7x-EasyEVM hardware and software. The FS704UM network module is interconnected with the TLZ7x-EasyEVM development board through the GPIO interface. The FS704UM network module is packaged in a pin 7PIN (2.54mm) package as shown in Table **??**.

The power supply is 5-16V DC and the peak power needs to be more than 8W. The serial port is TTL level (default 3.3V). RDY: high level means not connected to the server, low level means connected to the server. RSP: pull down for 3∼15 seconds to restore the factory settings. Normal use only needs to connect VIN, GND, TX and RX. The principle of power supply part is shown in Figure 2.

Table 1. FS704UM Network Module Pin Definitions

| Pinout | Name | Hidden meaning |
|---|---|---|
| 1 | RSP | Restore factory settings |
| 2 | RDY | Socket connection status indication |
| 3 | PEN | Core board power enable |
| 4 | RX | Data reception |
| 5 | TX | Data transmission |
| 6 | GND | Power input negative |
| 7 | VIN | Power input positive, supports 5-16V |

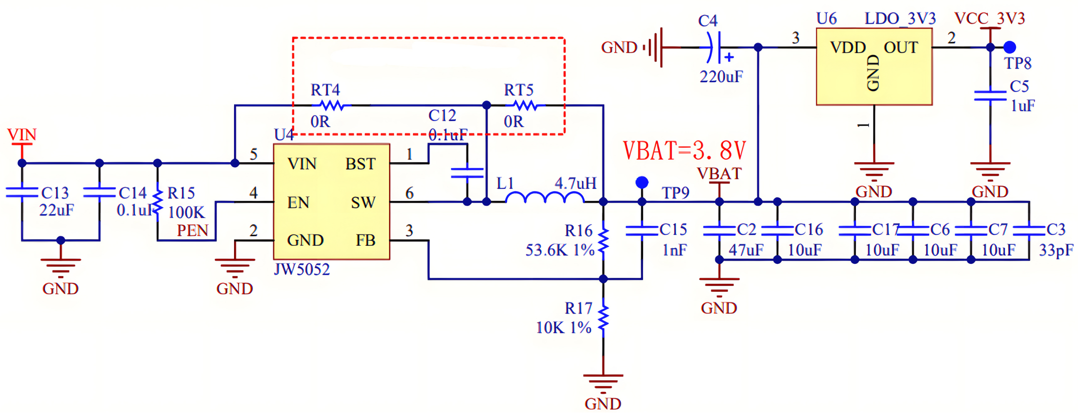


Figure 2. Principle of power supply part

**2.3. Gas analyser equipment.** In this paper, the NHA-502 model tail gas analyser [23] is used as the signal input platform for the vehicle tail gas platform. The NHA-502 model tail gas analyser is a device used for vehicle tail gas analysis, which is usually used for detecting and analysing the components of exhaust gases emitted by vehicles. This model of tailpipe analyser usually contains a variety of sensors and detection modules, which can detect CO, $CO_2$, $O_2$, $NO_x$ and other components in the exhaust gas, and evaluate the emissions of the vehicle based on these parameters. The main parameters of the NHA-502 tailpipe analyser are shown in Table 2.

**2.4. Engine operating data acquisition.** The OBD-II (On-Board Diagnostics II) interface is a standardised automotive diagnostic interface [24], which is usually located in the cabin of a vehicle and is used to connect a diagnostic tool or device in order to carry out vehicle troubleshooting, performance monitoring and data logging. Through

Table 2. Main parameters of tail gas analysing equipment

| **Parameters** | **Descriptions** |
|---|---|
| Measurement range | HC: 0 ~ 9,999 ppm; CO: 0 ~ 10%; $CO_2$: 0 ~ 18%; $O_2$: 0 ~ 15%; NO: 0 ~ 5,000 ppm |
| Measurement accuracy | HC: ±10 ppm (absolute), ±5% (relative); CO: ±0.3% (absolute), ±5% (relative); $CO_2$: ±0.5% (absolute), ±5% (relative); $O_2$: ±0.1% (absolute), ±5% (relative); NO: ±5% (absolute), ±4% (relative) |
| Response time | HC, CO, $CO_2$, $O_2$: ¡10 seconds; NO: ¡5 seconds |
| Preheating time | 8 minutes (3-minute rapid measurement) |
| Power supply | AC 220V ±10%, 50Hz ±1Hz |
| Net weight | 9 kg |
| Sizes | 300mm (W) × 200mm (H) × 250mm (D) |

the OBD-II interface, this paper can monitor the performance parameters of the vehicle in real time, such as fuel consumption rate, mileage, and emissions [25], and record and analyse these data, as shown in Table 3.

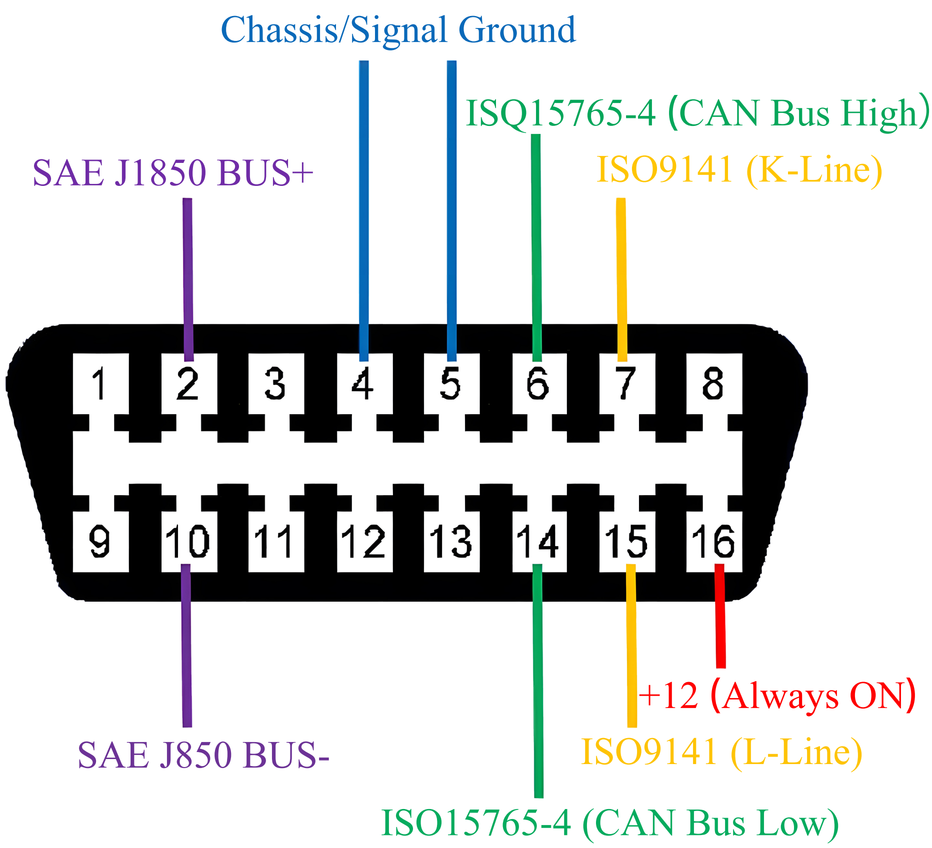


Figure 3. OBD-II Diagnostic Interface Pin Definitions

3. **Data pre-processing.** As the sensor device is easy to be interfered by the external environment in the process of collecting the automobile exhaust data, it produces a certain amount of noise or different data dimensions. In order to make the collected vehicle exhaust data easy to process and can reflect the actual results more realistically, the

PCA dimensionality reduction method [26] is applied to pre-process the collected vehicle exhaust data.

Let the collected car exhaust data set be:

$$A = \{a_1, a_2, \ldots, a_z\} \tag{1}$$

The initial dataset is decentralised as follows:

$$a' = \frac{1}{Z}\sum_{r=1}^{Z} a_r \tag{2}$$

where $a'$ denotes the mean value of the car exhaust data samples; $Z$ denotes the total number of data samples; $a_r$ denotes the $r$-th group of car exhaust data samples.

Let the covariance matrix of the vehicle exhaust sample data be $\frac{1}{Z}AA^T$, and its corresponding covariance be:

$$\mathrm{Cov} = \frac{\sum_{r=1}^{Z}(a_r - a')(a_r - a')^T}{Z-1} \tag{3}$$

The eigenvalues of the vehicle exhaust data are used to decompose the covariance matrix to obtain the eigenvalues of the covariance matrix and their corresponding eigenvectors. The eigenvalues of all the obtained data are arranged in descending order from the largest to the smallest, and the largest $k$ groups are selected. These $k$ eigenvectors form a matrix $D$, which transforms the data into a new dimensional space composed of $k$ eigenvectors. The preprocessing of the data is completed, which lays a reliable data foundation for the subsequent calculations.

## 4. Remote vehicle exhaust gas detection based on integrated deep learning.

**4.1. Feature information extraction.** This article extracts characteristics from 4G wireless flow data and studies their statistical variations, such as variance and mean, to find out how they contribute to categorization. Reducing the resource consumption of the feature information extraction process and speeding up the classification algorithm work is achieved by removing any feature that does not contribute to the classification process from the final feature set. A sliding window is a method of data sampling that involves placing a window on a portion of the data stream. This window only displays the most recent data that has arrived in the stream. The sliding window updates the oldest data with the most recent data as new data comes. In a data stream, a sliding window may be defined in two ways: first, according to sequence number; and second, according to time.

This article will examine the concept of a time-based sliding window, specifically referred to as a sliding time window [27]. The sliding time window approach does not explicitly indicate the quantity of flow packets caught inside the sliding time window, but rather focuses on the duration of the window itself [28]. Thus, in contrast to the initial sliding window based on sequence order, the quantity of flow packets collected fluctuates rather than being constant over the course of the sliding time frame. This variability accurately reflects the real-time fluctuations in flow.

To further improve the efficiency of feature information extraction by sliding window and to solve the problem of large memory occupied by the device during traffic processing, this paper adopts an incremental computation-based feature information calculation method. It is essential to compute several statistics of the flow packets, including variance, mean, correlation coefficient, and so on, in order to extract feature information. Incremental computing works on the following principle: The mean and variance of this collection of samples are shown below, supposing that $N$ packets, denoted as $x_1, x_2, \ldots, x_N$, are obtained from the network.

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{4}$$

$$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{X})^2 \tag{5}$$

In the event where there exist two sets of sample values, namely the historical value $h_1, h_2, \ldots, h_M$ and the incremental value $a_1, a_2, \ldots, a_N$, it is possible to derive the mean and variance of both sets by using the definitions of the formulas for variance and mean that were shown before, such as historical mean $\bar{H}$, the historical variance $\sigma_H^2$, the incremental mean $\bar{A}$, and the incremental variance $\sigma_A^2$, respectively.

$$\bar{H} = \frac{1}{M}\sum_{i=1}^{M} h_i \tag{6}$$

$$\sigma_H^2 = \frac{1}{M}\sum_{i=1}^{M}(h_i - \bar{X})^2 \tag{7}$$

$$\bar{A} = \frac{1}{N}\sum_{i=1}^{N} a_i \tag{8}$$

$$\sigma_A^2 = \frac{1}{N}\sum_{i=1}^{N}(a_i - \bar{A})^2 \tag{9}$$

Table 3. Feature types and calculation methods

| **Diagnostic property** | **Notation** | **Calculation method** |
|---|---|---|
| Quantities | $n$ | $n$ |
| Average value | $\mu_{s_i}$ | $\frac{LS}{n}$ |
| Mean square | $\sigma_{s_i}$ | $\sqrt{\frac{SS}{n} - (\frac{LS}{n})^2}$ |
| 2D mean | $\Vert s_i, s_j \Vert$ | $\sqrt{\mu_{s_i}^2 + \mu_{s_j}^2}$ |
| 2D variance | $R_{s_i,s_j}$ | $\sqrt{(\sigma_{s_i}^2)^2 + (\sigma_{s_j}^2)^2}$ |
| Covariance | $Cov_{s_i,s_j}$ | $\frac{SR_{i,j}}{n_i+n_j}$ |
| Correlation coefficient | $P_{s_i,s_j}$ | $\frac{Cov_{s_i,s_j}}{\sigma_{s_i}\sigma_{s_j}}$ |

The next step is to determine the average and standard deviation of the samples, $h_1, h_2, \ldots, h_M$ and $a_1, a_2, \ldots, a_N$ [29], and the results are shown in Equation (10) and Equation (11) respectively. When $N = 1$, the incremental mean becomes $\bar{A} = a_1$, and the incremental variance becomes $\sigma_A^2 = 0$, and at this point, Equation (10) and Equation (11) can be reduced to Equation (12) and Equation (13). So the incremental computation is characterised by the fact that the current statistics can be computed based on the historical statistics and the current increment. It may lessen the memory strain and increase the computational efficiency of low-performance Internet of Things gateways. The sliding time window size is 500 ms.

$$\bar{X} = \frac{1}{M+N}\left[\sum_{i=1}^{M} h_i + \sum_{i=1}^{N} a_i\right] = \frac{M\bar{H} + N\bar{A}}{M+N} \tag{10}$$

$$\sigma^2 = \frac{M\left[\sigma_H^2 + (\bar{X} - \bar{H})^2\right] + N\left[\sigma_A^2 + (\bar{X} - \bar{A})^2\right]}{M+N} \tag{11}$$

$$\bar{X} = \frac{M}{M+1}\bar{H} + \frac{M}{M+1}a_1 \tag{12}$$

$$\sigma^2 = \frac{M}{M+1}\left(\sigma_H^2 + (\bar{X} - \bar{H})^2\right) + \frac{1}{M+1}(\bar{X} - a_1)^2 \tag{13}$$

In this paper, we adopt a method based on data incremental computation of feature information, which can be extracted over a dynamic number of data streams. $S$, the mean value $\mu_s$, the variance $\sigma_s^2$, the standard deviation $\sigma_s$ of $S$ can be calculated by maintaining a ternary array $IS = (N, LS, SS)$, where $N$ is the number of items in $S$, $LS$ is the sum of items in $S$, and $SS$ is the square sum of items in $S$. There is no need to store each item $x_i$ in $S$ in memory when data $x_n$ is added to $S$; instead, $IS = (N + 1, LS + x_n, SS + x_n^2)$ is sufficient. Just $IS$ requires updating whenever fresh information comes in. The characteristics such as mean square, 2D mean, 2D variance, covariance, correlation coefficient, etc. can be calculated according to the method in Table 3.

4.2. **Emission status detection.** After the monitoring centre receives the vehicle's exhaust pollutant data and engine operating condition data, it is necessary to detect its status in order to find vehicles with abnormal emissions in a timely manner. The petrol engine control mechanism is a nonlinear time-varying closed-loop system, resulting in a complex nonlinear relationship between exhaust and operating conditions, which is difficult to be accurately identified with a linear model. Therefore, in this paper, integrated deep learning techniques are chosen to achieve the task of detecting abnormal emission patterns.

4.2.1. *AdaBoost algorithm.* AdaBoost is an integrated learning algorithm with adaptive enhancement capabilities that improves on the Boosting algorithm [30, 31]. Based on the weighting method, AdaBoost combines multiple weak learners with poor training ability on the same training set to form a strong learner with excellent training ability. AdaBoost can effectively avoid overfitting, can be used as an algorithmic framework to optimise other algorithms, and is extremely flexible.

4.2.2. *LSTM neural network.* LSTM neural network is a modified version of Recurrent Neural Network (RNN). The gating structure consists of forgetting gates, input gates and output gates, which control the information taking, inputting, updating and outputting. Therefore, LSTM neural network has a long time memory function, which can effectively solve the gradient explosion and gradient vanishing problems generated by RNN during training, and greatly enhance the accuracy of RNN [32, 33].

The calculation of the oblivion gate is shown below:

$$f_t = \sigma(W_f h_{(t-1)} + U_f x_t + b_f) \tag{14}$$

The calculation of the infeed gate is shown below:

$$i_t = \sigma(W_i h_{(t-1)} + U_i x_t + b_i) \tag{15}$$

$$C_t = \tanh(W_c h_{(t-1)} + U_c x_t + b_c) \tag{16}$$

$$C_t = C_{(t-1)} \odot f_t + i_t \odot C_t \tag{17}$$

The calculation of the output gate is shown below:

$$o_t = \sigma(W_o h_{(t-1)} + U_o x_t + b_o) \tag{18}$$

$$h_t = o_t \odot \tanh(C_t) \tag{19}$$

where $\odot$ is the product of matrix elements; $W_c$ is the weight matrix from the unit status to the input; $W_f$, $W_i$ and $W_o$ are the weight matrices of slave forgetting, input and output gates respectively; $U_c$ is the weight matrix from the unit state to the hidden layer; $U_f$, $U_i$, and $U_o$ are the weight matrices from forgetting, input, and output gates to the hidden layer respectively; $b_c$ is the deviation of the unit state; $b_f$, $b_i$, and $b_o$ are the bias vectors of forgetting, input, and output gates respectively; $\sigma$ is the sigmoid activation function; tanh is a hyperbolic tangent activation function.

4.2.3. *Improved LSTM based on integrated learning.* The improvement method proposed in this paper enhances the prediction accuracy and robustness of the LSTM neural network method by means of integrated learning. Multiple LSTM weak predictors are serially trained by AdaBoost integrated learning algorithm, and the samples and weak predictor weights are continuously adjusted during the training process, and then the weak predictors are weighted and combined to generate the strong predictors, which output the final prediction results. Improved LSTM integrates the potential of AdaBoost deep mining algorithm with the advantages of LSTM in dealing with time series problems, solves the problem of complex selection of parameters of multi-layer LSTM, and improves the defects of AdaBoost's sensitivity to outliers.

The construction process of the improved LSTM is as follows:

(1) Assign each sample data the same weight.

$$D_n = \frac{1}{M}, \quad n = 1, 2, \ldots, M \tag{20}$$

where $D_n$ is the weight of the $n$-th sample data, and $M$ is the total number of sample data.

(2) Set the network hyperparameters, and set the total number of LSTM weak predictors to $N_n$, and train the samples using LSTM neural network.

(3) For the $n$-th LSTM weak predictor, compute the maximum error of this weak predictor on the training set as:

$$E_n = \max |y_i - G_n(x_i)| \tag{21}$$

where $y_i$ is the weak predictor's prediction on the training set, and $G_n(x_i)$ is the observation on the training set.

(4) Calculate the relative error for each sample as:

$$e_{n,j} = \frac{(y_i - G_n(x_i))^2}{E_n^2} \tag{22}$$

where $e_{n,j}$ is the relative error of the $i$-th sample data for the $n$-th weak predictor.

(5) The error rate of the $n$-th LSTM weak predictor is obtained as:

$$e_n = \sum_{i=1}^{M} w_{n,i} e_{n,i} \tag{23}$$

where $e_n$ is the error rate of the $n$-th predictor, and $w_{n,i}$ is the weight of the $i$-th sample data for the $n$-th weak predictor.

(6) The weight coefficients of the $n$-th LSTM weak predictor are obtained as:

$$\alpha_n = \frac{e_n}{1 - e_n} \tag{24}$$

(7) Weight update for the $n+1$-th weak learner.

$$w_{n+1,i} = \frac{w_{n,i}\alpha_n^{1-e_{n,i}}}{Z_n} \tag{25}$$

where the normalisation factor is:

$$Z_n = \sum_{i=1}^{M} w_{n,i}\alpha_n^{1-e_{n,i}} \tag{26}$$

(8) A median-based combination method is used to fuse multiple weak learners into a strong learner.

$$f(x) = \sum_{i=1}^{N_n} \left( \ln \frac{1}{\alpha_n} \right) g(x) \tag{27}$$

where $g(x)$ is the median of $\alpha_n G_n(x)$.

## 5. Experimental results and analyses.

**5.1. Experimental setup.** In order to check the working performance of the automotive remote exhaust gas detection system designed in this paper, it needs to be verified by means of real vehicle testing. A 2022 Audi A4 sedan was chosen as the test vehicle. The basic parameters of the engine of this vehicle are shown in Table 4.

Table 4. 2022 Audi A4 Sedan Engine Parameters

| Parameters | Descriptions |
|---|---|
| Engine type | 2.0-litre turbocharged 4-cylinder engine |
| Maximum output power | 201 hp |
| Maximum torque | 236 Nm |
| Driver type | Front front-wheel drive/optional all-wheel drive system |
| Transmission | 7-speed dual-clutch automatic transmission |
| Fuel type | Premium petrol |
| Combined fuel consumption | Approx. 28mpg |

Theoretically, with the prolongation of vehicle use time, the engine exhaust pollutant emission law will change to some extent. However, considering the limitation of the experimental cost and the validity of the experiment, it is temporarily difficult for this paper to cover a variety of vehicles with different lengths of use for exhaust gas collection and analysis. The mileage of the vehicle selected in this paper is 30,000 kilometres, and its overall performance is in the normal smooth period. The mean square error (MSE) is used as the loss error during the neural network training, and the Adam optimisation algorithm is used to train the LSTM method. The learning rate is set to 0.001, the number of iterations is set to 100, the batch size is set to 32, and the number of predictors is set to 2. To avoid overfitting of the method, regularisation is used with dropout=0.2, and the ratio of the training set to the test set is 8:2.

5.2. **Analysis of test results.** The whole testing process is trained and recognised on the engine operating conditions and exhaust emission dataset. The specific parameters of the dataset are shown in Table 5.

Table 5. Experimental data sets

| Engine operating condition | Engine exhaust condition | Number (groups) | Sample size |
|---|---|---|---|
| Idle | Normal | 40 | 500 |
| | NO exceeded | 40 | 700 |
| | CO exceeded | 40 | 500 |
| | $CO_2$ exceeded | 40 | 700 |
| 1000r/min | Normal | 30 | 500 |
| | NO exceeded | 30 | 500 |
| | CO exceeded | 30 | 500 |
| | $CO_2$ exceeded | 30 | 500 |
| 2000r/min | Normal | 40 | 700 |
| | NO exceeded | 40 | 700 |
| | CO exceeded | 40 | 700 |
| | $CO_2$ exceeded | 40 | 700 |
| 3000r/min | Normal | 50 | 600 |
| | NO exceeded | 50 | 600 |
| | CO exceeded | 50 | 600 |
| | $CO_2$ exceeded | 50 | 600 |

Based on the dataset, the tailpipe identification algorithm designed in this paper is validated according to the 10-fold cross-validation method, and compared with the traditional Random Forest RF, AdaBoost and LSTM models. The recognition test results are shown in Figure 4.

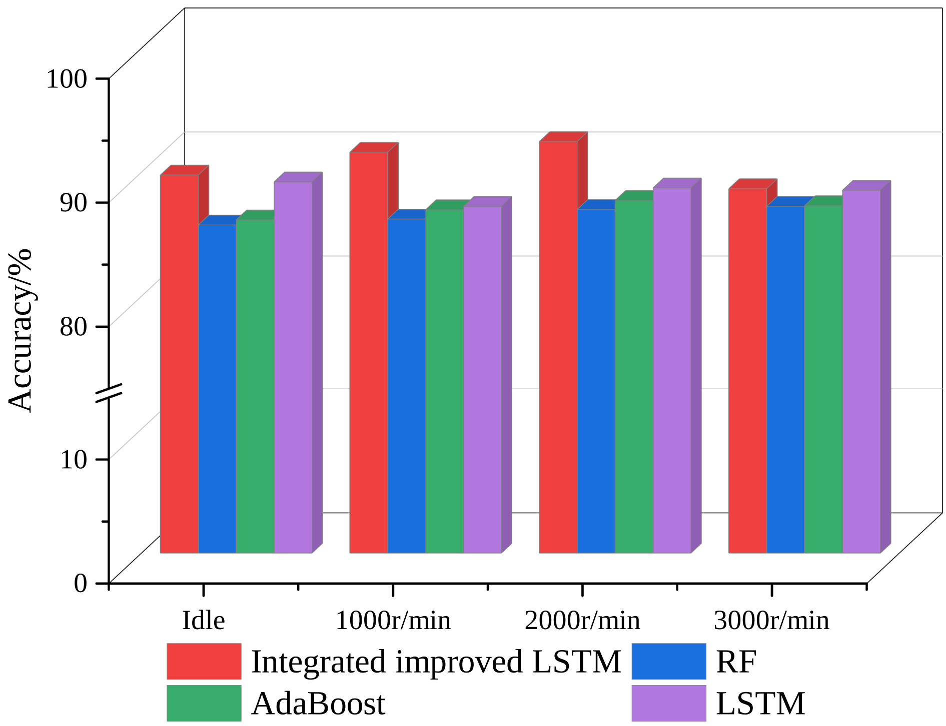


Figure 4. Accuracy of tailpipe condition detection models

The test results show that the automotive remote exhaust gas detection system designed in this paper is able to remotely monitor the engine conditions and exhaust gas data in real time. The system achieves an average accuracy of 91.33% for the abnormal tailpipe conditions, which is better than the traditional Random Forest RF, AdaBoost and LSTM models. In addition, it can be found that the integrated and improved LSTM model has a higher accuracy of exhaust gas identification when the engine is under small or medium load. Since the time period when the engine is in small and medium load is more during the actual operation of the vehicle, this system has better practicality.

6. **Conclusion.** In this work, an improved LSTM model based on AdaBoost integrated learning is proposed for implementing remote vehicle exhaust gas detection. An in-vehicle hardware platform based on the TLZ7x-EasyEVM development board was built and 4G network transmission was performed via FS704UM. The NHA-502 tailpipe analyser is used as the signal input platform of the vehicle tailpipe platform. A feature information calculation method based on incremental computation is proposed, which can be used to extract feature information on a dynamic number of data streams. In order to effectively improve the problems of low accuracy and insufficient generalisation ability of the dynamic real-time detection of exhaust gas, an improved LSTM model based on integrated learning algorithm is proposed. The extracted seven types of feature information are used as inputs in the exhaust gas monitoring centre, the LSTM neural network is used as the weak predictor of integrated learning, and the weak predictor is weighted and combined to obtain the strong predictor using the AdaBoost integrated learning algorithm. Through experimental validation on Audi A4 vehicles, the results show that the automotive remote exhaust gas detection system designed in this paper can remotely monitor the engine operating conditions and exhaust gas data in real time, and its accuracy rate of detecting abnormal exhaust gas conditions reaches 91.33%.

## REFERENCES


[1] J. Jones, E. Sckuck, R. Eldridge, N. Endow, and F. Cranz, "Measurement of automobile exhaust gas hydrocarbons," *Journal of the Air Pollution Control Association*, vol. 13, no. 2, pp. 73-77, 1963.

[2] W. J. Fleming, D. S. Howarth, and D. S. Eddy, "Sensor for on-vehicle detection of engine exhaust gas composition," *SAE Transactions*, pp. 1969-1984, 1973.

[3] M. M. Roy, "HPLC analysis of aldehydes in automobile exhaust gas: Comparison of exhaust odor and irritation in different types of gasoline and diesel engines," *Energy Conversion and Management*, vol. 49, no. 5, pp. 1111-1118, 2008.

[4] R. Moos, "A brief overview on automotive exhaust gas sensors based on electroceramics," *International Journal of Applied Ceramic Technology*, vol. 2, no. 5, pp. 401-413, 2005.

[5] G. R. Cann, W. M. Noble, and G. P. Larson, "Detection of Smog Forming Hydrocarbons in Automobile Exhaust Gases Using Plants as Indicators," *Air Repair*, vol. 4, no. 2, pp. 31-34, 1954.

[6] J. H. Visser, and R. Soltis, "Automotive exhaust gas sensing systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 6, pp. 1543-1550, 2001.

[7] M. Votsmeier, T. Kreuzer, J. Gieshoff, and G. Lepperhoff, "Automobile exhaust control," *Ullmann's Encyclopedia of Industrial Chemistry*, vol. 4, pp. 407-424, 2009.

[8] M. T. McCulloch, N. Langford, and G. Duxbury, "Real-time trace-level detection of carbon dioxide and ethylene in car exhaust gases," *Applied Optics*, vol. 44, no. 14, pp. 2887-2894, 2005.

[9] R. H. Quan, W. J. Wang, J. F. Wu, L. F. Hou, and G. F. Tang, "Study on a New Type of Automobile Exhaust Detecting System," *Applied Mechanics and Materials*, vol. 598, pp. 459-462, 2014.

[10] M. Xu, B. Peng, X. Zhu, and Y. Guo, "Multi-gas detection system based on non-dispersive infrared (NDIR) spectral technology," *Sensors*, vol. 22, no. 3, 836, 2022.

[11] O. K. Babalola, "Analysis of Automobile Exhaust Gas Effect on the Environment," *BIJOTE-Bichi Journal of Technology Education*, vol. 4, no. 1, pp. 84-91, 2016.

[12] Z. Li, C. Jin, and Y. He, "Fault diagnosis of engine based on exhaust density analysis and support vector machines," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 26, no. 4, pp. 143-146, 2010.

[13] Y. H. Bu, "Research in Elman Neural Network for AFR Model of Automotive Engine," *Advanced Materials Research*, vol. 204, pp. 755-759, 2011.

[14] M. A. Ceviz, A. K. Sen, A. K. Küleri, and İ. V. Öner, "Engine performance, exhaust emissions, and cyclic variations in a lean-burn SI engine fueled by gasoline–hydrogen blends," *Applied Thermal Engineering*, vol. 36, pp. 314-324, 2012.

[15] J.-J. Chung, and H.-J. Kim, "An automobile environment detection system based on deep neural network and its implementation using IoT-enabled in-vehicle air quality sensors," *Sustainability*, vol. 12, no. 6, 2475, 2020.

[16] Y. Yu, Y. Wang, J. Li, M. Fu, A. N. Shah, and C. He, "A novel deep learning approach to predict the instantaneous NO, emissions from diesel engine," *IEEE Access*, vol. 9, pp. 11002-11013, 2021.

[17] T.-Y. Wu, H. Li, S. Kumari, and C.-M. Chen, "A Spectral Convolutional Neural Network Model Based on Adaptive Fick's Law for Hyperspectral Image Classification," *Computers, Materials & Continua*, vol. 79, no. 1, pp. 19-46, 2024.

[18] Y. Ma, Y. Peng, and T.-Y. Wu, "Transfer learning model for false positive reduction in lymph node detection via sparse coding and deep learning," *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 2, pp. 2121-2133, 2022.

[19] F. Zhang, T.-Y. Wu, Y. Wang, R. Xiong, G. Ding, P. Mei, and L. Liu, "Application of Quantum Genetic Optimization of LVQ Neural Network in Smart City Traffic Network Prediction," *IEEE Access*, vol. 8, pp. 104555-104564, 2020.

[20] K. Smagulova, and A. P. James, "A survey on LSTM memristive neural network architectures and applications," *The European Physical Journal Special Topics*, vol. 228, no. 10, pp. 2313-2324, 2019.

[21] A. Khanna, and S. Kaur, "Internet of things (IoT), applications and challenges: a comprehensive review," *Wireless Personal Communications*, vol. 114, pp. 1687-1762, 2020.

[22] Y. B. Zikria, R. Ali, M. K. Afzal, and S. W. Kim, "Next-generation internet of things (IoT): Opportunities, challenges, and solutions," *Sensors*, vol. 21, no. 4, 1174, 2021.

[23] S. Wang, Y. Cai, H. Guo, D. Wu, and Y. Xie, "Effect of fuel concentration, inert gas dilutions, inert gas–water mist twin fluid medium dilutions, and end boundary condition on overpressure transients of premixed fuel vapor explosion," *Fuel*, vol. 309, 122083, 2022.

[24] D. Rimpas, A. Papadakis, and M. Samarakou, "OBD-II sensor diagnostics for monitoring vehicle operation and consumption," *Energy Reports*, vol. 6, pp. 55-63, 2020.

[25] M. Rykała, M. Grzelak, Ł. Rykała, D. Voicu, and R.-M. Stoica, "Modeling Vehicle Fuel Consumption Using a Low-Cost OBD-II Interface," *Energies*, vol. 16, no. 21, 7266, 2023.

[26] M. P. Uddin, M. A. Mamun, and M. A. Hossain, "PCA-based feature reduction for hyperspectral remote sensing image classification," *IETE Technical Review*, vol. 38, no. 4, pp. 377-396, 2021.

[27] Z. Zeng, L. Cui, M. Qian, Z. Zhang, and K. Wei, "A survey on sliding window sketch for network measurement," *Computer Networks*, vol. 226, 109696, 2023.

[28] A. S. Alghawli, "Complex methods detect anomalies in real time based on time series analysis," *Alexandria Engineering Journal*, vol. 61, no. 1, pp. 549-561, 2022.

[29] N. Li, D. Liu, L. Wang, H. Ye, Q. Wang, D. Yan, and S. Zhao, "Combination prediction of underground mine rock drilling time based on seasonal and trend decomposition using Loess," *Engineering Applications of Artificial Intelligence*, vol. 133, 108064, 2024.

[30] X. Huang, Z. Li, Y. Jin, and W. Zhang, "Fair-AdaBoost: Extending AdaBoost method to achieve fair classification," *Expert Systems with Applications*, vol. 202, 117240, 2022.

[31] E. Sevinç, "An empowered AdaBoost algorithm implementation: A COVID-19 dataset study," *Computers & Industrial Engineering*, vol. 165, 107912, 2022.

[32] A. Moghar, and M. Hamiche, "Stock market prediction using LSTM recurrent neural network," *Procedia Computer Science*, vol. 170, pp. 1168-1173, 2020.

[33] B. Lindemann, B. Maschler, N. Sahlab, and M. Weyrich, "A survey on anomaly detection for technical systems using LSTM networks," *Computers in Industry*, vol. 131, 103498, 2021.