

Article

E-Nose: Time–Frequency Attention Convolutional Neural Network for Gas Classification and Concentration Prediction

Minglv Jiang ^{1,2}, Na Li ^{3,4}, Mingyong Li ⁵, Zhou Wang ^{3,4}, Yuan Tian ⁶, Kaiyan Peng ², Haoran Sheng ², Haoyu Li ² and Qiang Li ^{1,2,*} 

¹ Key Laboratory of Physical Electronics and Devices for Ministry of Education and Shaanxi Provincial Key Laboratory of Photonics & Information Technology, Xi'an Jiaotong University, Xi'an 710049, China; jml.vito@stu.xjtu.edu.cn

² School of Electronic Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, China; pkymail@stu.xjtu.edu.cn (K.P.); baekhyun_shr_2001@stu.xjtu.edu.cn (H.S.); msn010@stu.xjtu.edu.cn (H.L.)

³ Northwest Survey & Planning Institute of National Forestry and Grassland Administration, Xi'an 710048, China; superman1232004@126.com (N.L.); xby107@163.com (Z.W.)

⁴ Key Laboratory of National Forestry and Grassland Administration on Ecological Hydrology and Disaster Prevention in Arid Regions, Xi'an 710048, China

⁵ CSSC AlphaPec Instrument (Hubei) Co., Ltd., Yichang 443005, China; limingyong0430@163.com

⁶ China National Engineering Laboratory for Coal Mining Machinery, CCTEG Taiyuan Research Institute Co., Ltd., Taiyuan 030032, China; 13643462056@139.com

* Correspondence: liqiang2014@mail.xjtu.edu.cn

Abstract: In the electronic nose (E-nose) systems, gas type recognition and accurate concentration prediction are some of the most challenging issues. This study introduced an innovative pattern recognition method of time–frequency attention convolutional neural network (TFA-CNN). A time–frequency attention block was designed in the network, aiming to excavate and effectively integrate the temporal and frequency domain information in the E-nose signals to enhance the performance of gas classification and concentration prediction tasks. Additionally, a novel data augmentation strategy was developed, manipulating the feature channels and time dimensions to reduce the interference of sensor drift and redundant information, thereby enhancing the model's robustness and adaptability. Utilizing two types of metal-oxide-semiconductor gas sensors, this research conducted qualitative and quantitative analysis on five target gases. The evaluation results showed that the classification accuracy could reach 100%, and the coefficient of the determination (R^2) score of the regression task was up to 0.99. The Pearson correlation coefficient (r) was 0.99, and the mean absolute error (MAE) was 1.54 ppm. The experimental test results were almost consistent with the system predictions, and the MAE was 1.39 ppm. This study provides a method of network learning that combines time–frequency domain information, exhibiting high performance in gas classification and concentration prediction within the E-nose system.

Keywords: electronic nose; gas sensor; time–frequency attention; convolutional neural network



Citation: Jiang, M.; Li, N.; Li, M.; Wang, Z.; Tian, Y.; Peng, K.; Sheng, H.; Li, H.; Li, Q. E-Nose: Time–Frequency Attention Convolutional Neural Network for Gas Classification and Concentration Prediction. *Sensors* **2024**, *24*, 4126. <https://doi.org/10.3390/s24134126>

Academic Editor: Michele Penza

Received: 29 May 2024

Revised: 14 June 2024

Accepted: 20 June 2024

Published: 25 June 2024



Copyright: © 2024 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/>).

1. Introduction

E-nose systems, often regarded as artificial olfaction systems, primarily consist of a series of sensors sensitive to specific gases and pattern recognition algorithms. In an E-nose, the perception of odors starts with gas sensors. Similar to how gas molecules bind to receptors on olfactory neurons in the sense of smell, gas molecules in the air interact with the sensing material within gas sensors, leading to measurable changes in electronic signals. These signals are then processed by neural network algorithms. Neural networks fragment the information through connections of neural nodes and link these fragments to extract and process more complex (abstract) information, and finally get the result of the analysis. This technology, capable of mimicking biological olfaction mechanisms, has become an economical and efficient non-destructive testing technique, widely applied in

fields such as biomedical diagnosis [1], industrial emission monitoring [2], and food quality assessment [3]. Given this, enhancing the detection capability and resolution of electronic nose systems is crucial. There are two basic ways to improve the performance of E-nose systems. One is to increase the sensitivity and selectivity of gas sensors by improving the sensitive materials, device structure, and manufacturing process [4–8]. The other approach involves employing advanced signal pre-processing methods, new software tools, and new algorithms [9–13] to improve the analytical performance of the pattern recognition algorithms.

In E-nose systems, sensor signals are typically represented as one-dimensional or multidimensional time series. These signals intuitively reflect the dynamic changes of specific attributes over time through their amplitudes, hence their widespread use. Although this effectively reveals the intuitive characteristics of sensor signals, current research on feature extraction in the frequency domain is relatively limited. Strategies primarily focused on time domain feature extraction may not fully explore the comprehensive information within signals. Frequency domain analysis offers a frequency-based perspective on signals, revealing their deeper characteristics and thus complementing the inadequacies of time domain analysis, involving a variety of methods such as Fourier transform, discrete cosine transform, and wavelet transform. The work of Dai et al. focused on extracting time domain and frequency domain features from E-nose signals [14]. Although they attempted to combine the maximum and average values from the time domain with the maximum and average energies obtained from the three-scale wavelet packet decomposition in the frequency domain, this approach was limited to selecting maximum and average energies as representative features, potentially leading to the loss of a considerable amount of useful information. Furthermore, their strategy for feature fusion, simply concatenating features linearly, lacks in-depth interaction analysis, which limits the ability to mine deep connections from limited features. Wang et al. expanded the range of feature extraction, including time domain features such as stable value and area value, coefficients in fitting equations, as well as frequency domain features like total energy and wavelet entropy extracted through wavelet packet decomposition [15]. Additionally, Kou et al. extracted a richer set of features from both the spatial and frequency domains, including characteristics of the spatial domain described by magnitude, difference, derivative, second derivative, integral, slope, and phase features, while applying Fast Fourier transform and wavelet to analyze frequency domain information [1]. Their experimental results revealed the potential of combining different features to enhance the accuracy of specific tasks, with wavelet, slope, and phase features showing significant importance. However, these research treats the extraction and analysis of time domain and frequency domain features as relatively independent, failing to delve into the interaction between time–frequency domain features, thus limiting the model’s ability to handle complex signals. Moreover, the manual feature extraction process overly relies on expert knowledge and experience, potentially overlooking key latent features.

This study proposed a time–frequency attention convolutional neural network (TFA-CNN) for deeply mining the time domain and frequency domain information in E-nose signals and achieved effective integration of these two types of information. The core of this network was that the discrete cosine transform (DCT) was used to replace the traditional Fourier transform (FT) and its inverse transform process. The results could reduce computational cost and avoid the interference of high-frequency noise. Additionally, Maxpooling operation was applied to process time domain signals to reduce time domain noise oscillation and highlight long-term trends. Finally, the model’s ability to analyze E-nose signals has been enhanced by expanding the receptive field (RF) of the convolutional layers, thereby improving the accuracy of predictions.

In deep learning, the generalization ability of a model, which is its performance on unseen data, is considered a key metric for assessing its success. Data augmentation, as a strategy to enhance model performance by increasing the diversity of training samples, has proven effective in the time series domain [16]. In facing the challenge of imbalance

among gas samples, Luo et al. adopted the synthetic minority oversampling technique (SMOTE), selecting several samples closely connected to the target sample through the nearest neighbor algorithm, and then using linear interpolation to increase the sample size [17]. Wang et al. further developed this idea by randomly selecting a sample from the training set for augmentation, calculating the Euclidean distance between this sample and other samples on similar labels, and selecting the five closest samples. These selected samples were combined with the original sample, assigning different weights and performing a weighted sum to create new samples, thus achieving the goal of expanding the training set sample size by tenfold [18]. Despite certain successes in data augmentation achieved by the strategies above, limitations remain. These methods rely on linear combinations of samples within the original data set, which may lead to generated data lacking in diversity and failing to cover a broader data distribution.

This work adopted a novel data augmentation strategy, namely signal masking (SM) which introduced changes directly in the data's feature and temporal dimensions. Masking operations on feature channels can simulate scenarios where partial sensor drift may occur, prompting the model to learn representations that remain effective despite partial information loss. Simultaneously, temporal domain masking, by randomly hiding parts of the time series, forces the model to not only reduce its dependence on irrelevant features, but also enhances its ability to adapt to incomplete or corrupted data that may be encountered in real-world applications. We further introduced a data augmentation strategy called window slicing (WS), which extracted shorter, continuous segments from the E-nose signals. Combining feature channel and temporal domain masking operations with the WS, a comprehensive data augmentation method for E-nose data analysis was proposed. This approach not only addressed the limitations posed by other data augmentation methods, but also enhances the model's ability to handle brief and incomplete data by introducing content-based variations.

In this paper, we designed and trained a TFA-CNN to achieve gas type discrimination and concentration prediction. Furthermore, we developed a data augmentation method to deal with scenarios lacking reliable and repeatable data. The fully trained model exhibited excellent generalization performance in both tasks. This work provides valuable insights for the further application of deep learning in E-nose systems.

2. Related Works

2.1. Convolutional Neural Networks

In the application of E-nose systems, numerous pattern recognition algorithms such as the K-nearest neighbors (KNN) [14], support vector machine regression (SVR) [19], and random forest [20] have been reported. These methods have demonstrated their effectiveness in specific applications. However, with the rise of deep learning technology as a new generation of pattern recognition method, its application in the field of E-nose systems has begun to attract researchers' attention [21]. Specifically, CNNs have been proven to be effective models for processing E-nose signals [22–24]. The framework for odor descriptor rating prediction proposed by Guo et al., which integrates Convolutional LSTM layers and regression layers into multiple sibling neural networks, has shown improvements in prediction accuracy [22]. Li et al. introduced an optimization method for detecting mixed gases and reported that the classification accuracy of the CNN model significantly outperforms support vector machine and multi-layer perception when the proportion of the training set becomes small [23].

The concept of the receptive field (RF) is an important tool for understanding deep CNNs [25]. Unlike fully connected networks, neurons in CNNs respond only to a localized area of the input data, which is referred to as the neuron's RF. The RF provides a mechanism that allows the network to capture both local and global information from the input data at different levels. Typically, the initial convolutional layers of the network have smaller RFs to capture local details of the input data. Subsequent convolutional layers have progressively larger RFs, allowing for the extraction of more complex and abstract features. When

processing time series data, the size of the RF determines how “far” the network can “see” in one-dimensional space, affecting its ability to capture longer temporal patterns. Fawaz et al. provided a definition of RF for one-dimensional time series data [26]. With a convolution stride of one and each layer’s filter length being equal to k_i , in a network of depth d , the RF can be represented as follows:

$$\text{RF} = \sum_{i=1}^d (k_i - 1) + 1, i \in [1, d]. \quad (1)$$

In computer vision, a larger RF is often required to capture more context for object recognition [25]. Similarly, it can be assumed that detecting a larger region in a one-dimensional time series data requires a larger RF.

Tang et al. discovered that the performance of one-dimensional CNNs is primarily determined by the optimal size of the RF, and it is relatively insensitive to the specific kernel size configurations that make up the RF size. Further findings suggest that if a model could encompass all possible RF sizes, its performance would be similar to that of a model with only the optimal RF size [27]. Therefore, they developed an Omni-scale block (OS-block) that can cover all sizes of RFs.

The architecture of OS-block is a three-layer multicore structure with the first two layers consisting of a set of prime numbers up to P_k , using \mathbb{P}_m to denote the set of kernel sizes in the m -th layer:

$$\mathbb{P}_m = \begin{cases} \{1, 2, 3, 5, \dots, p_k\}, & m \in \{1, 2\} \\ \{1, 2\}, & m = 3 \end{cases} \quad (2)$$

Among them, $\{1, 2, 3, 5, \dots, p_k\}$ is a set of prime numbers from 2 to P_k , including 1. As described by Equation (1), the set of RF sizes Ω is as follows:

$$\Omega = \{\mathbb{P}_1 + \mathbb{P}_2 + \mathbb{P}_3 - 2 | \mathbb{P}_i \in \mathbb{P}_m, i \in \{1, 2, 3\}\}. \quad (3)$$

Given that \mathbb{P}_1 and \mathbb{P}_2 are lists of prime numbers with a certain upper limit, according to the Goldbach conjecture, the set $\{\mathbb{P}_1 + \mathbb{P}_2\}$ will include all even numbers. With the addition of the third layer kernel sizes of 1 and 2, this design can cover all RF sizes for the length of the time series, thereby capturing features across different time scales.

2.2. Multi-Task Learning

In the application scenarios of E-noses, there is often a need to simultaneously determine gas types and predict concentrations. This typically requires training multiple independent models, thereby increasing the complexity of implementation. For example, to identify gas types and predict concentrations simultaneously, a classifier for gas type identification should be developed firstly, followed by training another regression model for gas concentration prediction. In recent years, the rise of multi-task learning (MTL) techniques has provided a solution to this challenge by sharing hidden layers among various tasks, while preserving independent output layers for each specific task [28]. Moreover, existing studies have shown that learning multiple related tasks simultaneously performs better than learning them independently [29]. Wang et al. have designed a multi-task learning convolutional neural network (MTL-CNN) to carry out three different classification tasks simultaneously: target identification, concentration prediction, and state judgment, achieving an identification accuracy of about 95% for 12 kinds of volatile organic compounds (VOCs) [24]. Research on MTL primarily focuses on exploring more effective mechanisms for parameter sharing, including hard parameter sharing, and soft parameter sharing. Among these, the hard parameter sharing mechanism is widely adopted due to its simplicity in implementation and suitability for handling highly correlated tasks. In practical applications of E-nose systems, where there is sometimes a need to detect both gas

components and their concentrations simultaneously, adopting the hard parameter sharing mechanism to simplify the training process is appropriate.

3. Method

In this study, we aim to tackle two tasks: classification and regression. Both tasks share the same training data $(x_i, y_i^k)_{i=1}^n$, with n representing the number of samples, and x_i and y_i^k representing the sensor's input data and its corresponding target labels in the k -th task, where $k \in \{1, 2\}$ stands for classification and regression tasks, respectively. The training objective of the TFA-CNN framework proposed in this study is to optimize model performance by simultaneously minimizing the loss functions of both tasks:

$$Loss = Loss_1(f^1(x), y^1) + Loss_2(f^2(x), y^2), \quad (4)$$

where, $f^k(x)$ is the prediction result given by the network based on the input data x , and $Loss_k$ represents the loss function for the k -th task.

For the classification task, the loss function uses cross-entropy loss, so the loss can be defined as follows:

$$Loss_1 = - \sum_{i=1}^n y_i^1 \log(f^1(x_i)). \quad (5)$$

while for the regression task, mean squared error (MSE) is typically used as the loss function, thus the loss is defined as follows:

$$Loss_2 = \frac{1}{n} \sum_{i=1}^n \left(y_i^2 - (f^2(x_i))^2 \right)^2. \quad (6)$$

Therefore, the total loss function can be represented as follows:

$$Loss = \alpha Loss_1 + (1 - \alpha) Loss_2. \quad (7)$$

where α is a weight factor, ranging from 0.0 to 1.0, used to balance the importance of the two tasks.

3.1. TF-Attention Block

In the field of frequency domain modeling, most methods rely on the process of FT and its inverse transform, which not only adds additional computational burden but also introduces high-frequency noise due to the inherent periodicity of the FT, leading to the Gibbs phenomenon and reducing the accuracy of model predictions. This issue is particularly prominent in applications requiring high-precision predictions, making accurate modeling of time series data an urgent problem to solve. Jiang et al. proposed a frequency-enhanced channel attention mechanism by utilizing the discrete cosine transform (DCT) to replace the commonly used FT-IFT process for frequency extraction. This mechanism integrates frequency domain data into the modeling process, enhancing the model's ability to extract frequency features [30]. Compared to the FT, DCT not only effectively mitigates the Gibbs phenomenon but also, due to its superior energy concentration properties over the discrete Fourier transform (DFT), becomes a better choice for extracting frequency information from time series.

In order to deeply explore the frequency information embedded in the E-nose signals and extract the implied sequence features more effectively, this study introduces the DCT to convert the time domain information to frequency domain information and proposes a time-frequency attention block (TF-attention block) with the fusion of frequency domain and time domain features. Figure 1 describes the architecture of the TF-attention block. For

a given single-channel E-nose signal sequence, the DCT of its length N one-dimensional sequence is defined as follows:

$$F(k) = c(k) \sum_{i=0}^{N-1} x_i \cos \left[\frac{(i+0.5)\pi}{N} k \right], k = 0, 1, 2, \dots, N-1, c(k) = \begin{cases} \sqrt{\frac{1}{N}}, & k = 0 \\ \sqrt{\frac{2}{N}}, & k \neq 0 \end{cases}. \quad (8)$$

where x_i is the sequence of E-nose signals for a single channel, and $F(k)$ is the transformed sequence.

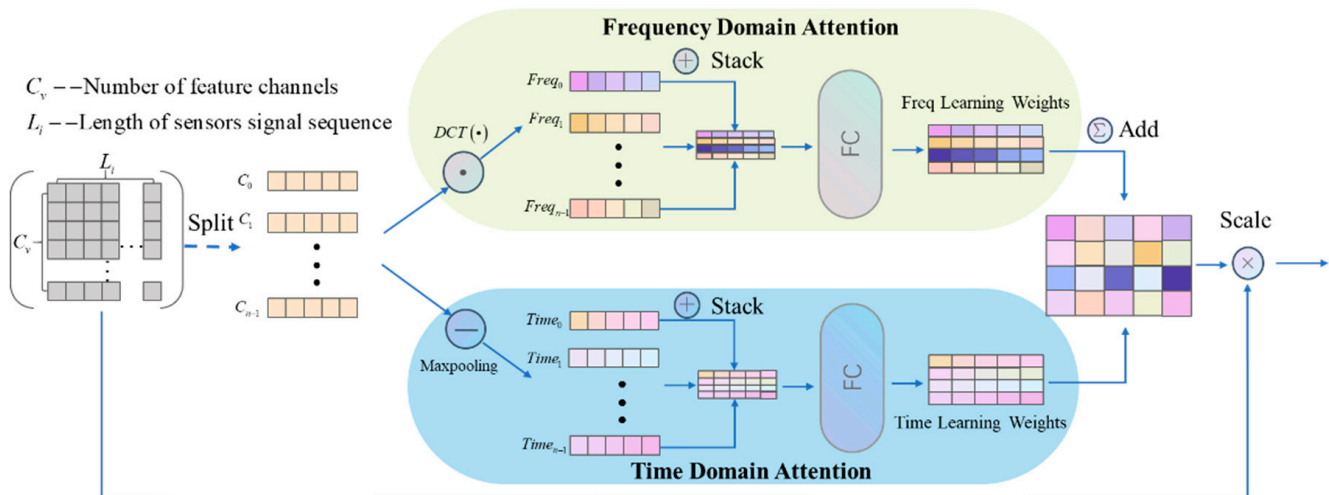


Figure 1. Schematic diagram of TF-attention block. (1) First, use the split operation to divide each multi-channel E-nose signal sequence. (2) Perform DCT and Maxpooling operations on each channel variable separately to generate $Freq_i$ and $Time_i$. (3) Stack $Freq_i$ and $Time_i$ along the channel dimension to obtain the frequency tensor and time tensor, respectively. (4) Use two fully connected layers to learn from the frequency tensor and time tensor. (5) Fuse the time domain attention and frequency domain attention to form the time–frequency attention TF_{att} , a step that facilitates the integration of adaptive features from both the temporal and frequency domains. (6) The input sequence is element-wise multiplied by the time–frequency attention to adjust the input data to better utilize its time–frequency domain features.

The process of TF-attention block is as follows:

$$c_0, c_1, \dots, c_{n-1} = \text{Split}(x), \quad (9)$$

$$Freq_i = \text{DCT}_i(c_i), \quad (10)$$

$$Time_i = \text{MaxPooling}_i(c_i), \quad (11)$$

$$Freq = \text{Stack}(Freq_0, Freq_1, \dots, Freq_{n-1}), \quad (12)$$

$$Time = \text{Stack}(Time_0, Time_1, \dots, Time_{n-1}), \quad (13)$$

$$TF_{att} = \text{Add}(\text{Fc}_1(Freq), \text{Fc}_2(Time)), \quad (14)$$

$$x_{out} = TF_{att} \cdot x. \quad (15)$$

First, the E-nose input signal $x \in R^{n \times L}$ is segmented along the channel dimension into n independent sensor channel variables $c_i \in R^{1 \times L}, i \in \{0, 1, 2, \dots, n-1\}$. This step is based on the premise that each sensor may have different characteristics as well as unique frequency and time information, thereby ensuring that individualized processing can be carried out for each channel. For each channel variable obtained from the segmentation, we first apply the DCT to convert the time domain signal into a frequency domain signal.

Subsequently, each frequency channel vector is stacked along the channel dimension to form a frequency tensor $Freq \in R^{n \times L}$. For the time domain signal, Maxpooling technology is employed to mitigate noise oscillations and highlight long-term trends, and then further stacked along the channel dimension to form a time tensor $Time \in R^{n \times L}$.

Following this, we processed the frequency tensor and time tensor separately through two independent fully connected networks to capture the time–frequency dependencies among different channels. Then, we fused the generated time domain and frequency domain attentions to form a comprehensive time–frequency attention $TF_{att} \in R^{n \times L}$. Finally, the original signal input is element-wise multiplied by the time–frequency attention, thereby adjusting the input data to better suit its time–frequency features. By introducing the time–frequency attention mechanism, our model not only achieves fine-grained adaptive feature extraction in both the time and frequency domains but also effectively enhances the model’s capability in fusing features from the frequency and time domains. Ultimately, the effectiveness of the proposed method through ablation experiments has been demonstrated.

In this study, the TF-attention block forms a base feature extraction framework with the OS-block; the OS-block is placed after the TF-attention block to obtain the optimal RF size for the time–frequency features.

3.2. Data Augmentation

In this section, the data augmentation strategies implemented for E-nose signal were described. Given that E-nose sensors may experience drift phenomena due to environmental factors in practical applications, it is important to note that drift might not occur in all sensors of the array. Firstly, considering that sensor drift could lead to more severe feature confusion than the loss of information caused by the absence of some sensor signals, a channel masking technique was adopted. Specifically, this technique randomly selects m_v times to mask $[0, m)$ channels, where $0 < m_v, m < n, n$ represents the total number of channels. Additionally, by acknowledging the potential presence of abundant redundant or irrelevant feature information in E-nose signals, such as the signals before and after adsorption is smooth so there are fewer features, a time masking strategy was proposed. This involves t consecutive time steps $[t_0, t_0 + t)$ that are masked, where t is randomly chosen from a uniform distribution ranging from 0 to the time masking parameter T , t_0 is chosen from $[0, L - t]$, $t < L$ and L represents the length of the E-nose signals.

Secondly, considering the relationship between data length and the amount of feature information, overly lengthy input data can not only extend training duration but also potentially lead the model to learn redundant or irrelevant information, thereby affecting training efficiency and the model’s generalizability. Therefore, the WS strategy was used to focus on local segments of the E-nose signals for effectively capturing short-term dependencies in the data, which involved extracting slices of a specific length from the E-nose signals and incorporating them into the training set, ensuring each slice retains the original label. To maintain consistency in the lengths of the data segments, it is necessary to upsample the sliced data to match the original sequence length. Methods similar to WS have been applied in related studies of E-nose systems [22,31].

By implementing SM to simulate sensor loss and reduce the interference of invalid information, along with adopting WS strategy to accommodate the model’s needs for analyzing local segments, our method not only enhances the model’s ability to learn hidden patterns in the data but also improves its robustness to incomplete information. Consequently, this approach enhances the diversity of the data while also boosting the model’s generalization performance. Table 1 presents the data augmentation parameters adopted in this study.

Table 1. Data augmentation parameters.

	WS			SM		
Data augmentation	Ratio 0.9	Number 3	m 2	m_v 3	t 10	m_t 3

(Where m_v and m_t , respectively, indicate the number of applications of channel masking and time masking, m denotes the number of channels masked in each application, and t specifies the duration of each masking operation along the time dimension. "Ratio" reflects the proportion of each slice relative to the original data length, while "Number" denotes the total number of slicing operations performed).

4. Experiment

To ensure consistency in the model training and testing processes, all algorithms in this study were implemented in an Integrated Development Environment (IDE) equipped with Python 3.8 (PyCharm 2023.1.3, Community Edition). The programs were run on a Windows 11 operating system (CPU: Intel Core i7-12700F (Santa Clara, CA, USA), GPU: NVIDIA RTX 3060 (Santa Clara, CA, USA), RAM: 64 GB). To accelerate the deep learning training process, GPU parallel computing was utilized, employing the GPU version of TensorFlow (V2.7.0) as the main computational tool. Regarding the model training parameter settings, a batch size of 32, a number of epochs of 200, and a learning rate of 0.001 were selected. Moreover, "Adam" was used as the optimizer, a widely adopted optimization technique in the field of deep neural networks.

4.1. Network Architecture

As shown in Figure 2, the base feature learning layer is comprised of the TF-attention block and OS-block, achieving shared feature extraction from the original data. Building on this foundation, to address specific regression and classification tasks, subsequent network blocks designed for these specific tasks were employed. These include parallel stacked Classification and Regression blocks, enabling simultaneous execution of gas type classification and gas concentration prediction tasks. Both blocks have similar structures, each processing the input tensor in parallel four times. The first three processing steps utilize convolution operations with varying numbers of layers. The first layer of convolution uses filters with a length and stride of 1, reducing the input sequence from T dimensions to m dimensions. This dimensionality reduction significantly simplifies the model structure and mitigates overfitting issues in small data sets. By employing multi-layer convolution, the RF for specific tasks is gradually expanded, while different lengths of filters are used to extract multi-resolution features. The fourth processing step involves average pooling to reduce the model's sensitivity to minor perturbations, followed by a convolution layer to further reduce data dimensions. Finally, the outputs from each parallel convolution processing are concatenated to form the final output.

At this stage, despite the processing of feature data, its dimensions remain large, which not only increases computation time but may also lead to overfitting. The introduction of a global average pooling layer effectively reduces the size of the feature data. Finally, through a fully connected layer, the extracted features are mapped to the sample label space, obtaining outputs for both classification and regression tasks. To assess network performance and optimize the model parameters, this study employed a five-fold cross validation method. By training on different subsets of the training data, this approach provides more robust evaluation results and efficient parameter selection results.

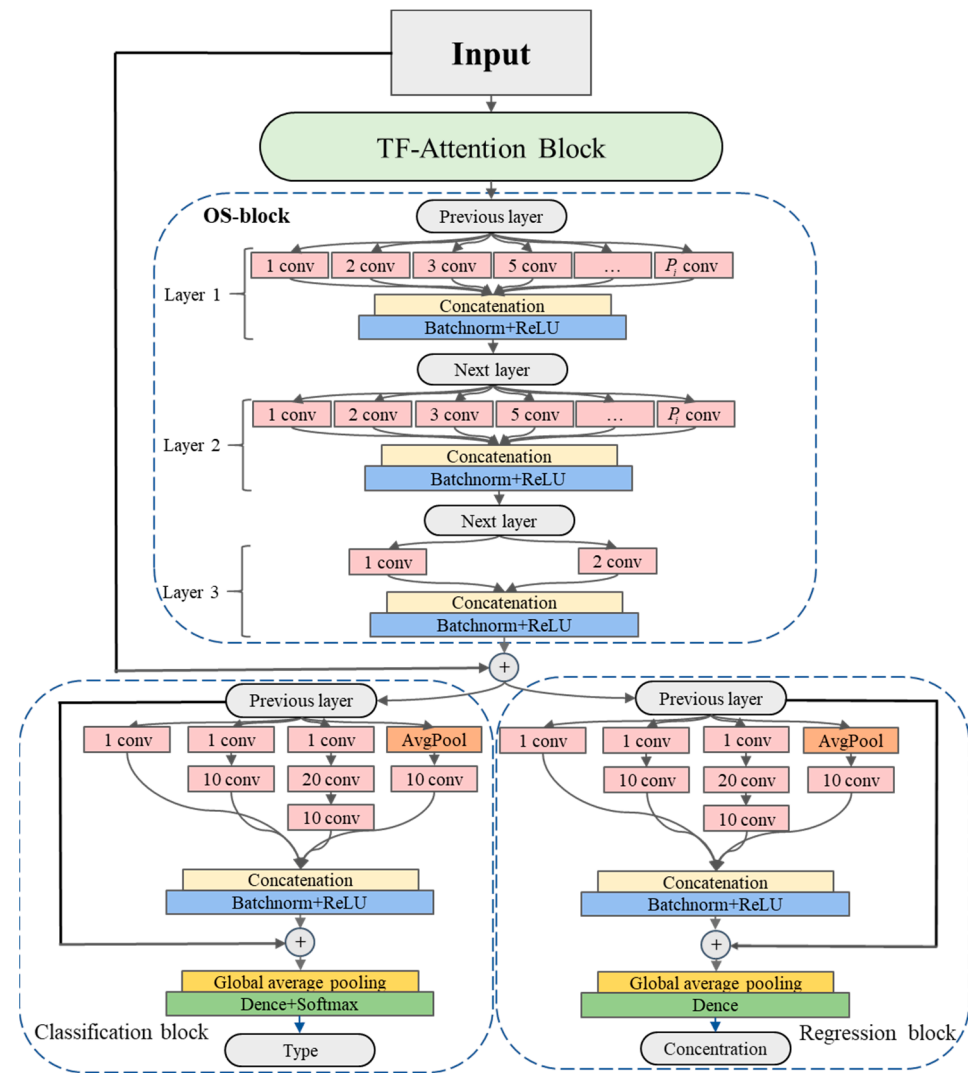


Figure 2. The TFA-CNN model based on time–frequency attention. The model is divided into three main parts: (1) The initial input E-nose signal data first flows through the TF-attention block, effectively integrating the time domain and frequency domain features of the data. (2) Subsequently, the data enters a three-layer OS-block, which promotes the capture of richer features by expanding the RF. (3) Finally, the data is transmitted to the Classification block and Regression block, respectively. These two blocks focus on the specific feature learning of their respective tasks and output results related to the tasks.

4.2. Evaluation Metrics

To evaluate the predictive performance of the model, this study uses accuracy as the assessment metric for the classification task. Meanwhile, for regression task, MAE, the r , and the R^2 are introduced to measure the model's performance:

$$Accuracy = \frac{\text{Number of correctly predicted samples}}{\text{Total number of samples}}, \quad (16)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (17)$$

$$r = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}}, \quad (18)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (19)$$

where \hat{y}_i represents the model's prediction for the i -th sample, and y_i represents the corresponding true value. \bar{y} and $\bar{\hat{y}}$, respectively, represent the mean of the true values and the mean of the predicted values. The highest value for accuracy is 100%, indicating that all type recognitions are correct. A lower MAE value indicates a lower level of error in the model, meaning higher prediction accuracy. At the same time, when the r and the R^2 values are closer to 1, it signifies a higher degree of model fit, indicating superior predictive performance.

4.3. Dataset

In this study, the four gas sensors used were independently prepared by our laboratory, manufactured using the radio frequency magnetron sputtering, with specific sputtering parameters detailed in Table 2. These sensors had different $\text{In}_2\text{O}_3/\text{SnO}_2$ ratios (wt%) set at 80:20, 85:15, and 90:10, as well as the tin oxide (SnO_2) sensor. The choice of these four different sensors aims to utilize their cross-selectivity, thereby enhancing the capability to identify specific gases.

Table 2. Parameters of radio frequency magnetron sputtering.

Labels	$\text{In}_2\text{O}_3:\text{SnO}_2$ (wt%)	Sputtering Time (min)	Sputtering Temperature ($^{\circ}\text{C}$)	RF Power (W)	Ar: O_2
0100	00:100				
9010	90:10	4	300	300	20:0
8515	85:15				
8020	80:20				

Five target gases were collected: ethanol, acetone, methanol, nitrogen dioxide and sulfur dioxide. In the setup for each individual experiment, the concentration range for each target gas was set from 7.5 ppm to 75.0 ppm, increasing in steps of 7.5 ppm, with each experiment independently repeated five times.

To accurately capture the response and recovery behavior of the gas sensors, standardized time intervals were employed. The specific process is as follows. First, the sensor array was left to stabilize in the air for 5 min after turning on the testing system. Then, the protective cover was closed within 10 s, and a specific concentration of the target gas was introduced, maintaining this state for 60 s. Subsequently, the protective cover was opened, and nitrogen gas was used to purge the inside of the protective cover for 1 min to remove residual gas. Finally, after a 4 min interval, the next round of data collection began. When completing the collection for a certain gas, the experiment was paused for over 24 h to mitigate sensor fatigue or damage caused by frequent testing; then, collection continued for the next gas type. Following this procedure, all collection work in this study lasted about half a month, ultimately accumulating over 50,000 valid data points, as detailed in Table 3.

Table 3. Gas concentration range and quantity.

Type	Days	Repetition	Concentration Range (ppm)	Number
Ethanol	1–3	5	7.5–75.0	10
Acetone	4–6	5	7.5–75.0	10
Methanol	7–9	5	7.5–75.0	10
Nitrogen dioxide	10–12	5	7.5–75.0	10
Sulfur dioxide	13–15	5	7.5–75.0	10

Considering the rapid response and recovery characteristics exhibited by sensors in gas detection, which generally allow for the detection of gas responses to be completed within about 3 and a half minutes, this study utilized only the response values within the first 210 s of each detection as the input sequence. Ultimately, a total of 250 data samples were obtained.

4.4. Discussion

The experimental data were divided into two parts: training data and testing data, with a ratio of 4:1. By using the data augmentation parameters from Table 1, the training data was expanded tenfold, from 200 training samples to 2000 samples. The training data used 5-fold cross validation to search for the optimal model parameters, while the testing data was reserved for final evaluation. To assess the performance of the base feature learning layer composed of the TF-attention block and OS-block in classification and regression tasks, two types of CNNs for comparison were introduced. These include the RTFA-CNN, which focused on the concentration prediction task by removing the classification task branch, and the CTFA-CNN, dedicated to the gas type classification task by eliminating the regression task branch. A fully convolutional network (FCN) is also introduced in this study for comparison. FCN was proposed by Wang et al. for processing time series classification tasks [32], which replaced the fully connected layer in traditional CNNs with convolutional layers. It mainly consisted of three convolutional blocks, with batch normalization and ReLU activation functions attached after each convolutional block, and a global average pooling layer applied after the last convolutional block, which is then connected to a softmax classifier for time series data classification. It was the most accurate deep learning model in the evaluation of 12 multivariate time series classification datasets [33]. In addition, Tan et al. extended the application of the FCN network to time series regression analysis by replacing the softmax function with a linear activation function, enabling the model to predict continuous values and demonstrating leading performance on multiple datasets [34]. The FCN model used the network structure and parameter configurations proposed by Tan et al. [34]. To further enrich the comparison baseline, three widely used pattern recognition algorithms were also compared, including extreme gradient boosting (XGBoost), random forest, and SVR. For these three algorithms, a grid search method was used to adjust hyperparameters, aiming to find the optimal parameter configuration. All algorithms were implemented on the same computing platform as TFA-CNN.

The final training model based on the TFA-CNN demonstrated excellent stability and generalizability. After 200 epochs of training, the model's total loss was stabilized, with no overfitting observed. The variation curves of regression and classification losses are, respectively, illustrated in Figure 3a,b.

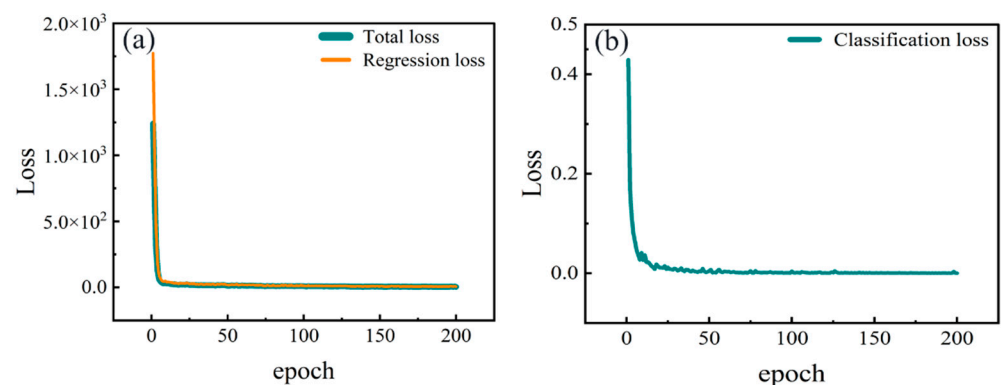


Figure 3. The decline in the loss of the TFA-CNN during the training process: (a) regression loss, (b) classification loss.

The confusion matrix for type identification, as shown in Figure 4a, indicated that the accuracy of the classification task for the five target gases reached 100%. Figure 4b

displayed the comparison between actual and predicted concentrations for 50 randomly selected test samples. Experimental results demonstrated that the model exhibited stable generalization capability in concentration prediction tasks. However, the error rate in low concentration intervals increases compared to higher concentrations. This could be due to the sensor's response in low concentration ranges not being as significant or linear as under high concentration conditions, resulting in a relatively larger variance in sensor characteristics (such as resistance or voltage changes). Such an expansion of characteristic intervals may introduce uncertainty to the model's prediction accuracy. Furthermore, sample collection at low concentrations might increase the randomness of sampling bias due to experimental errors, thereby amplifying the differences between sampling results. The results revealed that there was a good consistency between the predicted values and actual values by the TFA-CNN model, indicating that the MTL architecture was extremely effective for gas concentration prediction.

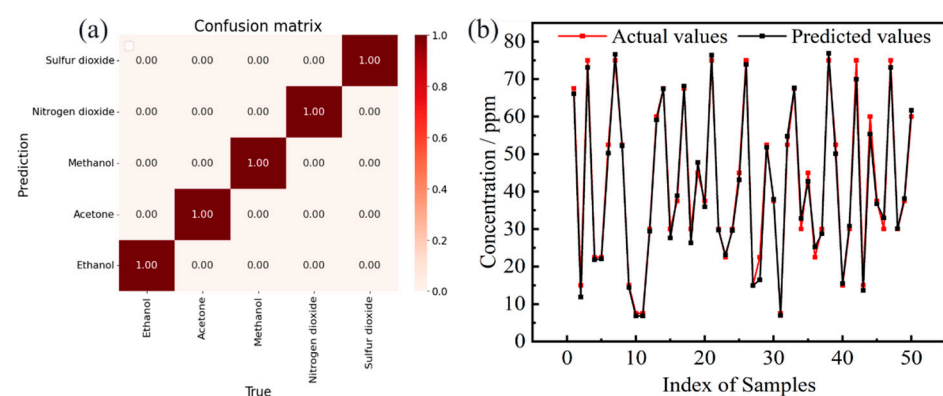


Figure 4. Results of the TFA-CNN on the testing set: (a) confusion matrix for type recognition, (b) actual and predicted values of concentration.

Table 4 provides a summary of statistical data for seven models. The results highlight that the TFA-CNN, RTFA-CNN and CTFA-CNN models exhibit excellent performance in combined classification and regression tasks, the individual regression and classification tasks, respectively. For the TFA-CNN model (The results with bold font in Table 4), the classification accuracy reached 100%, the R^2 score for regression tasks was as high as 0.99, the r was 0.99, and the MAE was 1.54 ppm.

Table 4. Performance of the seven models on the dataset.

Model	Classification Task		Regression Task	
	Accuracy	MAE	R^2	r
TFA-CNN	100%	1.542	0.990	0.996
RTFA-CNN	-	1.818	0.987	0.994
CTFA-CNN	100%	-	-	-
FCN	-	2.597	0.971	0.989
Random forest	-	3.469	0.949	0.975
SVR	-	3.438	0.956	0.979
XGBoost	-	3.727	0.943	0.972

Overall, compared to other common pattern recognition algorithms, TFA-CNN not only demonstrated more robust learning capabilities but also offered a more flexible design pattern, outperforming other models in overall performance.

4.5. Ablation Experiment

To systematically evaluate the contribution of the proposed TF-attention block and data augmentation strategies to model performance, this study designed a series of ablation experiments. These experiments aimed to verify the effectiveness of each module in enhancing the model's learning ability.

Comparative experiments were conducted to analyze the changes in the performance of the model after removing the TF-attention block, removing the frequency attention alone, or removing the time attention alone. The results of the experiments are organized in Table 5. The results showed that the absence of any form of attention mechanism leads to a decrease in model performance, with the removal of the TF-attention block having the most significant effect. The TF-attention block, by effectively weighting the input signals in both time and frequency domains, enables the model to focus more on information beneficial to the prediction task, thus improving the model's accuracy and generalization ability.

Table 5. Results of time–frequency attention ablation experiments (x represents No; ✓ represents Yes).

Frequency Attention	Time Attention	Data Augmentations	Classification Task		Regression Task		
			Accuracy	MAE	R^2	r	
x	x	✓	100%	2.222	0.979	0.993	
x	✓	✓	100%	1.875	0.984	0.994	
✓	x	✓	100%	1.767	0.986	0.994	

The specific contribution of each data augmentation method to the model performance is assessed by removing any of the three methods, channel masking, time masking and window slicing, respectively. The experimental results are shown in Table 6, where the generalization performance of the model decreased when any of the methods are absent. Using any two methods had a positive impact on the model's generalization performance with similar enhancement, and it can be hypothesized that each method can enhance the diversity of the samples in different feature dimensions and optimize the model's performance in its specific dimensions. When all three methods are applied simultaneously, the diversity of the dataset is maximized, thus achieving the best generalization effect.

Table 6. Results of data augmentation ablation experiments (x represents No; ✓ represents Yes).

Channel Masking	Time Masking	Window Slicing	Classification Task		Classification Task		
			Accuracy	MAE	R^2	r	
✓	x	✓	100%	1.853	0.983	0.992	
x	✓	✓	100%	1.803	0.985	0.993	
✓	✓	x	100%	1.837	0.985	0.993	
x	x	x	100%	3.134	0.952	0.980	

When the data augmentation strategy was not used, the model's MAE increased from 1.54 ppm to 3.13 ppm, the R^2 decreased from 0.990 to 0.952, and the r dropped from 0.996 to 0.980. This result verified the practical benefits of the proposed data augmentation strategy in improving model stability and prediction accuracy.

4.6. Independent Validation Experiment

To evaluate the model's performance on unseen data, 45 sets of data were recollected to thoroughly analyze the model's generalization ability. The specific types of gases and concentration ranges involved were detailed in Table 7. The training set comprises 30 sets of data, covering three types of gases at five different concentration levels, with each concentration level spaced by 10 ppm and sampled twice. The testing set included 15 sets of data, covering three types of gases at five different concentration levels, with each

concentration level spaced by 10 ppm and sampled once, and the concentration levels in the training and testing sets did not overlap.

Table 7. The range and quantity of gas concentrations in the independent validation experiment.

Type	Repetition	Concentration Range (ppm)	Number	Mode
Ethanol	2	10.0–50.0	5	Train
	1	15.0–55.0	5	Test
Acetone	2	10.0–50.0	5	Train
	1	15.0–55.0	5	Test
Methanol	2	10.0–50.0	5	Train
	1	15.0–55.0	5	Test

This validation experiment was conducted through fine-tuning the pre-trained TFA-CNN model, whose parameters were derived from training on five gases. Specific adjustments included removing the last classification layer in the regression task and adjusting this layer from the original five neurons to three to accommodate the classification task for three types of gases, while the configurations of other layers remained unchanged. The fine-tuning process was conducted in the same training environment as the TFA-CNN, with training parameters including a batch size of 8, a number of epochs of 200, and a learning rate set at 0.001.

Typically, a smaller data set size increases the risk of the model encountering overfitting. As shown in Figure 5, the fine-tuned TFA-CNN model's concentration prediction results on the test set effectively demonstrate the efficacy of data augmentation techniques in reducing the risk of overfitting. Even with only 30 sets of training samples collected, data augmentation methods could help the model obtain more robust prediction results.

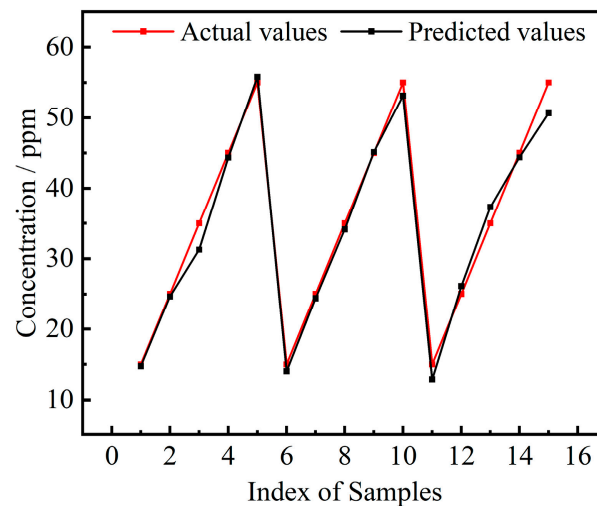


Figure 5. Actual and predicted values in the testing set of the independent validation experiment.

Table 8 summarizes the performance of the seven models in the experiment. It further validates that the TFA-CNN model (The results with bold font in Table 8), after undergoing training on a larger data set and subsequent fine-tuning, could still effectively grasp the response mechanisms of gas-sensitive sensors to various types of gases and maintain outstanding performance when faced with new data challenges.

Table 8. Performance of the seven models on the independent validation dataset.

Model	Classification Task		Regression Task	
	Accuracy	MAE	R^2	r
Fine-tuned TFA-CNN	100%	1.386	0.983	0.993
RTFA-CNN	-	2.182	0.956	0.980
CTFA-CNN	100%	-	-	-
FCN	-	2.414	0.949	0.988
Random forest	-	6.940	0.525	0.779
SVR	-	3.345	0.915	0.970
XGBoost	-	8.337	0.447	0.711

Additionally, when faced with a testing set containing unseen concentration ranges, traditional pattern recognition methods such as random forest, SVR, and XGBoost demonstrated significantly poorer generalization abilities compared to deep learning models, with MAE all exceeding 3.34 ppm. This result further validates the effectiveness of deep learning models in feature learning.

5. Conclusions

In this study, we proposed and validated a time–frequency attention convolutional neural network, specifically designed for gas type identification and concentration prediction tasks. This model integrated a time–frequency attention mechanism with deep convolutional neural networks, enabling the network to capture and merge temporal and frequency domain information in E-nose signals at the base feature learning layer. By widening the RF of convolutional layers, the model’s capability to capture gas signal features was enhanced, thereby improving gas detection and concentration estimation performance.

Experimental results indicated that the TFA-CNN model accurately predicted the gas categories and their concentrations. The accuracy of the model reached 100%, the coefficient of determination of 0.99, the Pearson correlation coefficient was also 0.99, and the MAE was only 1.54 ppm, which is significantly better than other models. Furthermore, ablation experiments validated the effectiveness of this proposed TF-attention block and data augmentation methods. The model also performed well in independent validation experiments, obtaining accurate results on the testing set of unseen concentration ranges, with a classification accuracy of 100%, the coefficient of determination of 0.98, the Pearson correlation coefficient of 0.99, and the MAE of only 1.39 ppm, further verifying the model’s ability to learn the patterns of gas responses.

Author Contributions: M.J.: conceptualization, investigation, methodology, writing—original draft. Q.L.: conceptualization, methodology, writing—original draft, review and editing, supervision, funding acquisition. N.L., M.L., Z.W. and Y.T.: investigation, methodology, technical support. K.P., H.S. and H.L.: investigation, methodology, software. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fundamental Research Funds for the Central Universities (Nos. xzy012022088, xzd012023045, xzy022024061).

Institutional Review Board Statement: This study did not require ethical approval.

Informed Consent Statement: This study did not involve humans.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: Thanks to the Instrumental Analysis Center of Xi’an Jiaotong University for helping in material analysis.

Conflicts of Interest: Author Mingyong Li was employed by the CSSC AlphaPec Instrument (Hubei) Co., Ltd. Author Yuan Tian was employed by the CCTEG Taiyuan Research Institute Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Kou, L.; Zhang, D.; Liu, D.X. A Novel Medical E-Nose Signal Analysis System. *Sensors* **2017**, *17*, 402. [[CrossRef](#)] [[PubMed](#)]
2. Li, H.; Luo, D.H.; Sun, Y.L.; GholamHosseini, H. Classification and Identification of Industrial Gases Based on Electronic Nose Technology. *Sensors* **2019**, *19*, 5033. [[CrossRef](#)] [[PubMed](#)]
3. Anwar, H.; Anwar, T.; Murtaza, S. Review on food quality assessment using machine learning and electronic nose system. *Biosens. Bioelectron. X* **2023**, *14*, 100365. [[CrossRef](#)]
4. Zhou, J.Y.; Bai, J.L.; Zhao, H.; Yang, Z.Y.; Gu, X.Y.; Huang, B.Y.; Zhao, C.H.; Cairang, L.; Sun, G.Z.; Zhang, Z.X.; et al. Gas sensing enhancing mechanism via doping-induced oxygen vacancies for gas sensors based on indium tin oxide nanotubes. *Sens. Actuators B Chem.* **2018**, *265*, 273–284. [[CrossRef](#)]
5. Hasan, B.A. SnO₂ doped In₂O₃ thin films as reducing gas sensors. In *IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Bristol, UK, 2020; p. 072011.
6. Almaev, A.V.; Kopyev, V.V.; Novikov, V.A.; Chikiryaka, A.V.; Yakovlev, N.N.; Usseinov, A.B.; Karipbayev, Z.T.; Akilbekov, A.T.; Koishybayeva, Z.K.; Popov, A.I. ITO Thin Films for Low-Resistance Gas Sensors. *Materials* **2022**, *16*, 342. [[CrossRef](#)]
7. Kononova, I.; Moshnikov, V.; Kononov, P. SnO₂-Based Porous Nanomaterials: Sol-Gel Formation and Gas-Sensing Application. *Gels* **2023**, *9*, 283. [[CrossRef](#)] [[PubMed](#)]
8. Peng, K.Y.; Li, Q.; Ma, M.W.; Li, N.; Sheng, H.R.; Li, H.Y.; Huang, Y.J.; Yun, F. Acidic Gas Determination Using Indium Tin Oxide-Based Gas Sensors. *Sensors* **2024**, *24*, 1286. [[CrossRef](#)]
9. Yi, Z.K.; Li, C. Anti-Drift in Electronic Nose via Dimensionality Reduction: A Discriminative Subspace Projection Approach. *IEEE Access* **2019**, *7*, 170087–170095. [[CrossRef](#)]
10. Liu, T.; Li, D.Q.; Chen, J.J.; Chen, Y.B.; Yang, T.; Cao, J.H. Active Learning on Dynamic Clustering for Drift Compensation in an Electronic Nose System. *Sensors* **2019**, *19*, 3601. [[CrossRef](#)]
11. Leon-Medina, J.X.; Pineda-Munoz, W.A.; Burgos, D.A.T. Joint Distribution Adaptation for Drift Correction in Electronic Nose Type Sensor Arrays. *IEEE Access* **2020**, *8*, 134413–134421. [[CrossRef](#)]
12. Zhang, W.T.; Liu, T.P.; Ye, L.; Ueland, M.; Forbes, S.L.; Su, S.W. A novel data pre-processing method for odour detection and identification system. *Sens. Actuators A Phys.* **2019**, *287*, 113–120. [[CrossRef](#)]
13. Botre, B.A.; Gharpure, D.C.; Shaligram, A.D. Embedded electronic nose and supporting software tool for its parameter optimization. *Sens. Actuators B Chem.* **2010**, *146*, 453–459. [[CrossRef](#)]
14. Dai, Y.W.; Zhi, R.C.; Zhao, L.; Gao, H.Y.; Shi, B.L.; Wang, H.Y. Longjing tea quality classification by fusion of features collected from E-nose. *Chemometr. Intell. Lab.* **2015**, *144*, 63–70. [[CrossRef](#)]
15. Wang, Z.H.; Chen, W.; Gu, S.; Wang, Y.W.; Wang, J. Evaluation of trunk borer infestation duration using MOS E-nose combined with different feature extraction methods and GS-SVM. *Comput. Electron. Agr.* **2020**, *170*, 105293. [[CrossRef](#)]
16. Iwana, B.K.; Uchida, S. An empirical survey of data augmentation for time series classification with neural networks. *PLoS ONE* **2021**, *16*, e0254841. [[CrossRef](#)] [[PubMed](#)]
17. Luo, J.; Zhu, Z.H.; Lv, W.; Wu, J.; Yang, J.H.; Zeng, M.; Hu, N.T.; Su, Y.J.; Liu, R.L.; Yang, Z. E-Nose System Based on Fourier Series for Gases Identification and Concentration Estimation From Food Spoilage. *IEEE Sens. J.* **2023**, *23*, 3342–3351. [[CrossRef](#)]
18. Wang, Y.W.; Yu, Y.; Zhao, H.J.; Yin, C.B.; Shi, Y.; Men, H. A self-developed electronic nose system combines data enhancement and multi-branch kernels channel attention to identify the gas information of industrial polypropylene. *Sens. Actuators A Phys.* **2024**, *366*, 115005. [[CrossRef](#)]
19. Bonah, E.; Huang, X.Y.; Yang, H.Y.; Aheto, J.H.; Yi, R.; Yu, S.S.; Tu, H.Y. Detection of Salmonella Typhimurium contamination levels in fresh pork samples using electronic nose smellprints in tandem with support vector machine regression and metaheuristic optimization algorithms. *J. Food Sci. Technol.* **2021**, *58*, 3861–3870. [[CrossRef](#)] [[PubMed](#)]
20. Li, Q.; Gu, Y.; Wang, N.F. Application of Random Forest Classifier by Means of a QCM-Based E-Nose in the Identification of Chinese Liquor Flavors. *IEEE Sens. J.* **2017**, *17*, 1788–1794. [[CrossRef](#)]
21. Liu, T.P.; Guo, L.H.; Wang, M.; Su, C.; Wang, D.; Dong, H.; Chen, J.D.; Wu, W.W. Review on Algorithm Design in Electronic Noses: Challenges, Status, and Trends. *Intell. Comput.* **2023**, *2*, 0012. [[CrossRef](#)]
22. Guo, J.; Cheng, Y.; Luo, D.H.; Wong, K.Y.; Hung, K.; Li, X. ODRP: A Deep Learning Framework for Odor Descriptor Rating Prediction Using Electronic Nose. *IEEE Sens. J.* **2021**, *21*, 15012–15021. [[CrossRef](#)]
23. Li, X.L.; Guo, J.Y.; Xu, W.P.; Cao, J.X. Optimization of the Mixed Gas Detection Method Based on Neural Network Algorithm. *ACS Sens.* **2023**, *8*, 822–828. [[CrossRef](#)] [[PubMed](#)]
24. Wang, T.; Zhang, H.X.; Wu, Y.; Jiang, W.K.; Chen, X.W.; Zeng, M.; Yang, J.H.; Su, Y.J.; Hu, N.T.; Yang, Z. Target discrimination, concentration prediction, and status judgment of electronic nose system based on large-scale measurement and multi-task deep learning. *Sens. Actuators B Chem.* **2022**, *351*, 130915. [[CrossRef](#)]

25. Luo, W.J.; Li, Y.J.; Urtasun, R.; Zemel, R. Understanding the effective receptive field in deep convolutional neural networks. In Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016.
26. Fawaz, H.I.; Lucas, B.; Forestier, G.; Pelletier, C.; Schmidt, D.F.; Weber, J.; Webb, G.I.; Idoumghar, L.; Muller, P.A.; Petitjean, F. InceptionTime: Finding AlexNet for time series classification. *Data Min. Knowl. Discov.* **2020**, *34*, 1936–1962. [[CrossRef](#)]
27. Tang, W.S.; Long, G.D.; Liu, L.; Zhou, T.Y.; Blumenstein, M.; Jiang, J. Omni-scale Cnns: A simple and effective kernel size configuration for time series classification. In Proceedings of the ICLR 2022-10th International Conference on Learning Representations, Virtual, 25–29 April 2022.
28. Yu, D.B.; Wang, X.R.; Liu, H.X.; Gu, Y. A Multitask Learning Framework for Multi-Property Detection of Wine. *IEEE Access* **2019**, *7*, 123151–123157. [[CrossRef](#)]
29. Chang, X.J.; Yang, Y. Semisupervised Feature Analysis by Mining Correlations Among Multiple Tasks. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 2294–2305. [[CrossRef](#)] [[PubMed](#)]
30. Jiang, M.W.; Zeng, P.Y.; Wang, K.; Liu, H.; Chen, W.B.; Liu, H.R. FECAM: Frequency enhanced channel attention mechanism for time series forecasting. *Adv. Eng. Inform.* **2023**, *58*, 102158. [[CrossRef](#)]
31. Pulluri, K.K.; Kumar, V.N. Qualitative and Quantitative Detection of Food Adulteration Using a Smart E-Nose. *Sensors* **2022**, *22*, 7789. [[CrossRef](#)]
32. Wang, Z.G.; Yan, W.Z.; Oates, T. Time series classification from scratch with deep neural networks: A strong baseline. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 1578–1585.
33. Fawaz, H.I.; Forestier, G.; Weber, J.; Idoumghar, L.; Muller, P.A. Deep learning for time series classification: A review. *Data Min. Knowl. Discov.* **2019**, *33*, 917–963. [[CrossRef](#)]
34. Tan, C.W.; Bergmeir, C.; Petitjean, F.; Webb, G.I. Time series extrinsic regression: Predicting numeric values from time series data. *Data Min. Knowl. Discov.* **2021**, *35*, 1032–1060. [[CrossRef](#)]

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.