An All-in-One Electronic Nose with a Multi-Scale Temporal Shift and Depth Dynamic Aggregation Network for Low-Concentration Malodorous Gas Classification

Chenlong Gu, Qianshen Wu, Nan Wang, *Member, IEEE,* Yuxuan Zhang, *Member, IEEE,* Sebastian Bader, *Senior Member, IEEE,* Xiaofeng Ling†, Yongjing Wan†, Daqi Gao*

***Abstract*—Detecting low-concentration malodorous gases of down to single-digit ppm levels remains challenging due to the weak and overlapping transient responses of metal oxide semiconductor (MOS) sensors and limited, as well as imbalanced datasets. In this work, we propose an all-in-one electronic nose (E-nose) prototype and a multi-scale, one-dimensional convolutional neural network (CNN) incorporating a temporal-shift and depth dynamic aggregation (TS-DDA) module for robust odor classification. The E-nose instrument adopts a modular design comprising: 1) a sensing module with a 16-channel MOS sensor array enclosed in an annular gas chamber validated through k-ω computational fluid dynamics (CFD) simulation for uniform flow and rapid desorption (< 40 s residual washout); and (2) a data-acquisition and control module, implemented on a single custom printed circuit board (PCB), providing precise sampling, pump and valve control. Additionally, the proposed MS-TS-DDA network enhances temporal feature density and multi-depth information fusion while maintaining low computational cost. A controlled laboratory dataset consisting of 497 samples covering eight types of low-concentration malodorous gases (0.5-60 ppm) was collected and balanced via a temporal oversampling strategy inspired by T-SMOTE. The proposed method achieves a mean classification accuracy of 95.15 % under 5-fold cross-validation, outperforming classical CNN baselines. These results indicate that the proposed framework provides a compact, cost-effective and robust solution for low-concentration odor detection under resource-constrained conditions.**

***Index Terms*—Electronic nose (E-nose), annular gas chamber, computational fluid dynamics (CFD) simulation, convolutional neural network (CNN), temporal shift-depth dynamic aggregation (TS-DDA), data augmentation, odor detection.**

I. Introduction

MALODOROUS gases refer to volatile compounds that stimulate human olfactory organs, causing unpleasant smells and environmental harm. They are widely present in chemical industries, waste-treatment and livestock facilities, and surrounding areas, posing serious environmental and health hazards [1]. Common harmful components include hydrogen sulfide (H_2S), trimethylamine, methanethiol, and styrene, which can cause respiratory irritation, neurotoxicity, and systemic organ damage even at low concentrations [2], [3]. With increasing public concern for air quality, the safety risks associated with such malodorous pollutants have attracted growing attention, making reliable and efficient odor detection an essential part of environmental monitoring and pollution control.

Olfactometry, such as the three-point comparison odor bag method, is a widely employed technique for analyzing air quality and quantifying malodorous intensity [4]. This method relies on human panels to detect and compare diluted odor samples against odor-free references, providing semi-quantitative odor concentration measurements (e.g., odor units per cubic meter, ou/m^3). However, exposure to malodorous pollutants during olfactometric testing poses health hazards, including respiratory irritation, neurotoxic effects, and long-term systemic toxicity [1].

For certain malodorous gases (e.g., hydrogen sulfide and ammonia), spectrophotometric methods are conventionally employed for qualitative and quantitative analysis by using a spectrophotometer [5]. This method achieves low detection limits (e.g., gas absorption in reagent solutions) and is widely adopted as a reference method in wastewater, industrial emissions and air quality assessments. However, it is time-consuming and labor-intensive, requiring extensive sample pre-treatment. As for some other malodorous gases (e.g., trimethylamine, methanethiol, dimethyl sulfide, dimethyl disulfide, carbon disulfide, and styrene), gas chromatography-mass spectrometry (GC-MS) is a prevalent analytical technique for both qualitative and quantitative detection [6]. GC-MS separates and identifies volatile compounds with high accuracy, yet its high cost and lack of portability limit field applications, driving the adoption of electronic noses (E-noses), which prioritize rapid detection, cost-efficiency, and field applicability [7].

Some internationally influential E-nose systems such as PEN3 and FOX series have been applied to the detection of odor gases [8]. PEN3 is composed of 10 different metal-oxide semiconductor (MOS) sensors with varying sensitivities to different volatile organic compounds (VOCs), while FOX series includes 12-18 MOS sensors. Both E-nose system are

This work was supported by the National Natural Science Fundation of China under Grant 62376096.

*Main Corresponding Author: Daqi Gao (gaodaqi@ecust.edu.cn).

†Co-Corresponding Author: Xiaofeng Ling (xfling@ecust.edu.cn) and Yongjian Wan (wanyongjing@ecust.edu.cn).

Chenlong Gu, Xiaofeng Ling, Nan Wang, Yongjing Wan and Daqi Gao are with the School of Information Science and Engineering, East China University of Science and Technology, Shanghai.

Yuxuan Zhang is with the College of Intelligent Science and Engineering, Beijing University of Agriculture, Beijing, China and the Department of Computer and Electrical Engineering, Mid Sweden University, Sundsvall, Sweden. (yuxuan.zhang@miun.se)

Sebastian Bader is with the Department of Computer and Electrical Engineering, Mid Sweden University, Sundsvall, Sweden. (sebastian.bader@miun.se)

Manuscript received xx xx, 2026.

compact and battery-powered for field deployment, enabling real-time monitoring and detection. However, the price of these commercial E-noses is prohibitively expensive (i.e., more than 50,000 dollars each), limiting their scalability in field applications.

Some recent research on E-nose design has centered on gas chambers to improve dispersion uniformity, purge efficiency, cost efficiency and measurement repeatability. For instance, Qian et al. [9] developed a multi-sensor system with a stepped chamber to classify herbal medicines, while Wang et al [10]. designed and optimized a bionic gas chamber for Chinese liquor recognition. Both chambers demonstrate the benefit of modular design in enhancing portability and system stability. However, such designs still suffer from limitations in gas-flow distribution and chamber geometry. Conventional square or stepped chambers often create turbulent airflow and stagnant regions [9], leading to uneven gas dispersion across sensor arrays and residual analyte accumulation in sharp corners. The bionic chamber [10] requires at least 60 s to reduce the residual concentration in all regions to below 10 %, indicating relatively slow desorption and non-uniform gas exchange. These effects degrade the response uniformity and recovery speed of sensors. Therefore, a gas chamber with uniform flow distribution and efficient desorption is essential to achieve stable and repeatable E-nose measurements.

While an optimized gas chamber ensures more stable and consistent sensor responses, the overall recognition accuracy of an E-nose system ultimately depends on the effectiveness of its data processing and classification algorithms. As the sensing module transforms gas concentration variations into multichannel temporal signals, robust pattern recognition models are required to extract discriminative temporal and cross-sensor features for reliable odor identification. In contrast to image or speech signal processing domains, electronic nose systems analyze multichannel time-series signals arising from the dynamic responses of gas sensors to exposed odors. For odor classification, traditional machine learning methods, such as k-nearest neighbor (KNN) [11], support vector machine (SVM) [12], and random forest (RF) [13], have been widely used for odor classification. More recently, deep-learning methods, particularly convolutional neural networks (CNNs) [14], have shown superior performance by automatically extracting discriminative features and modeling complex nonlinear relationships. Due to the one-dimensional (1D) temporal characteristics of gas sensor signals, 1D-CNN enables effective extraction of discriminative features from sequential sensor measurements while preserving spatial-temporal correlations between sensor channels. For example, in [14]–[16] lightweight 1D-CNN architectures were proposed to achieve high performance in various gas classification tasks. To further improve extraction of distinct drift-invariant features from E-nose response signals, and to enhance long-term gas-recognition stability, Guo et al. [17] proposed an anti-drift gas-detection algorithm based on a multiscale CNN, which was evaluated on public E-nose datasets. However, conventional CNNs struggle to capture long-term dependencies and to model global temporal correlations in time-series E-nose data due to the limited local receptive field. Zhu et al. [18] thus introduced dilated convolutions in CO concentration prediction to enlarge the receptive field without increasing model parameters, thereby enabling the network to establish connections with a broader range of past time steps and capture long-range temporal dependencies. MobileNet architectures [19] introduced separable convolutions, which leverage depth-wise and point-wise convolutions to reduce the parameter count and computational cost of CNNs. Recently, Chen et al [20] combined dilated and separable convolutions to extract the temporal features of each signal and the correlations between different sensors, achieving high performance on both public and private datasets. Inspired by video-understanding frameworks for temporal modeling, particularly DyFADet [21], which employs temporal-shift operations to promote inter-frame feature interaction at no extra computational cost, we incorporate a similar strategy into our model for E-nose time-series analysis. Furthermore, feature aggregation across different network depths has recently attracted attention in various time-series and lightweight vision models [22], as it enables complementary information from shallow and deep layers to be dynamically fused. Thereby, it enhances feature representation and stability. However, to the best of our knowledge, explicit depth-wise adaptive aggregation of shallow and deep features has not been reported in E-nose applications.

Motivated by the limitations of conventional odor classification approaches and E-nose systems, such as high costs, poor gas distribution, and limited performance in low-concentration scenarios, we propose an all-in-one, low-cost E-nose system, which ensures uniform flow, rapid desorption, and stable multi-channel signal collection, while the multi-scale temporal-shift and depth dynamic aggregation (MS-TS-DDA) network enhances temporal feature learning and multi-depth information fusion for accurate malodorous gas classification. The main contributions in this work are summarized as follows:

1) A compact and portable E-nose prototype is proposed, integrating a 16-channel MOS sensing module and a data acquisition and control module, which are jointly implemented on a single PCB. The system incorporates an annular gas chamber validated through k-ω computational fluid dynamics (CFD) simulation to verify uniform flow distribution and rapid gas exchange (< 40 s residual washout). The prototype has a material cost of approximately $1500.

2) A low-concentration (0.5–60 ppm) malodorous gas dataset is generated using the proposed E-nose system, comprising eight distinct malodorous gases and a total of 497 original samples under controlled laboratory conditions. The training set is further augmented using a temporal oversampling strategy to mitigate data imbalance.

3) A multi-scale 1D-CNN architecture with a temporal-shift and depth dynamic aggregation (TS-DDA) module is proposed, which addresses the limitations of conventional CNNs in capturing long-range dependencies and fusing multi-depth features. The performance of the MS-TS-DDA network is demonstrated through structural optimization, ablation experiments, and performance comparison with multiple classification methods.

II. E-Nose Instrument Design

Considering a comprehensive set of factors including system cost, physical size, gas sampling efficiency, cleaning performance, and signal acquisition precision, we built an all-in-one E-nose instrument prototype that integrates gas inlet/outlet control, odor sensing, analog-to-digital signal conversion, and data transmission to a mini industrial workstation for analysis. Fig. 1(a) shows the photograph of the E-nose instrument designed, which mainly consists of a sensing module and a data acquisition and control module implemented on a single PCB. Fig. 1(b) illustrates the schematic diagram of the sensing module, and Fig. 1(c) presents the photograph of the hardware architecture of the data acquisition and control board as well as each sub-module diagram.

A. Sensing Module

The sensing module comprises an MOS sensor array, a three-path gas flow system and a gas chamber as shown in Fig. 1(b). The sensor array in this work consists of 16 MOS sensors with partially overlapping sensitivity ranges, one temperature sensor integrated within the gas chamber, and one ambient humidity-temperature (RH-T) sensor for environmental monitoring, as shown in Table I with detailed information. The selected gas sensors synergistically work to generate unique "odor fingerprint" features for single or complex gas mixtures, enabling pattern recognition capabilities [23]. The temperature and RH-T sensors ensure stable operational conditions by monitoring real-time thermal dynamics of the gas chamber (maintained at $50 \pm 0.5\,^{\circ}\mathrm{C}$) and ambient environment ($25\,^{\circ}\mathrm{C}$ controlled by air conditioner), with chamber heating wires maintaining optimal temperature control.

The three-path gas flow system employs three solenoid valves (Valves 1-3) to control the switching of the gas flow directions. Solenoid valve 1 governs the reference air path, supplying clean calibration gas to reset sensor baselines, while Solenoid 2 regulates the sample inlet path, directing target gases into the chamber for detection. The third valve manages the exhaust pathway, expelling residual gases after measurement cycles and mitigating the sensor drift. Flow dynamics across these paths are precisely maintained through adjustable flow restrictors (the throttle valve), which collaboratively regulate and monitor gas flow rates within a defined range (e.g., 1.5-6 L/min). A vacuum pump serves as the system's primary power source, driving active gas exchange and enabling rapid chamber purging to minimize cross-contamination. Table II shows the state of the three solenoid valves under different working phases, along with their corresponding flow rates and durations.

The core innovative feature enabling rapid gas exchange lies in the annular gas chamber as shown in Fig. 2(a), which adopts a single-channel concentric configuration with an inward-axial inlet and an outward-axial outlet to form a unidirectional purge flow. This annular configuration allows for a longer gas flow path to avoid sensor drift within a compact footprint compared to linear designs, while simultaneously accommodating a larger number of gas sensors around the perimeter. To mitigate the dilution effect, our annular chamber

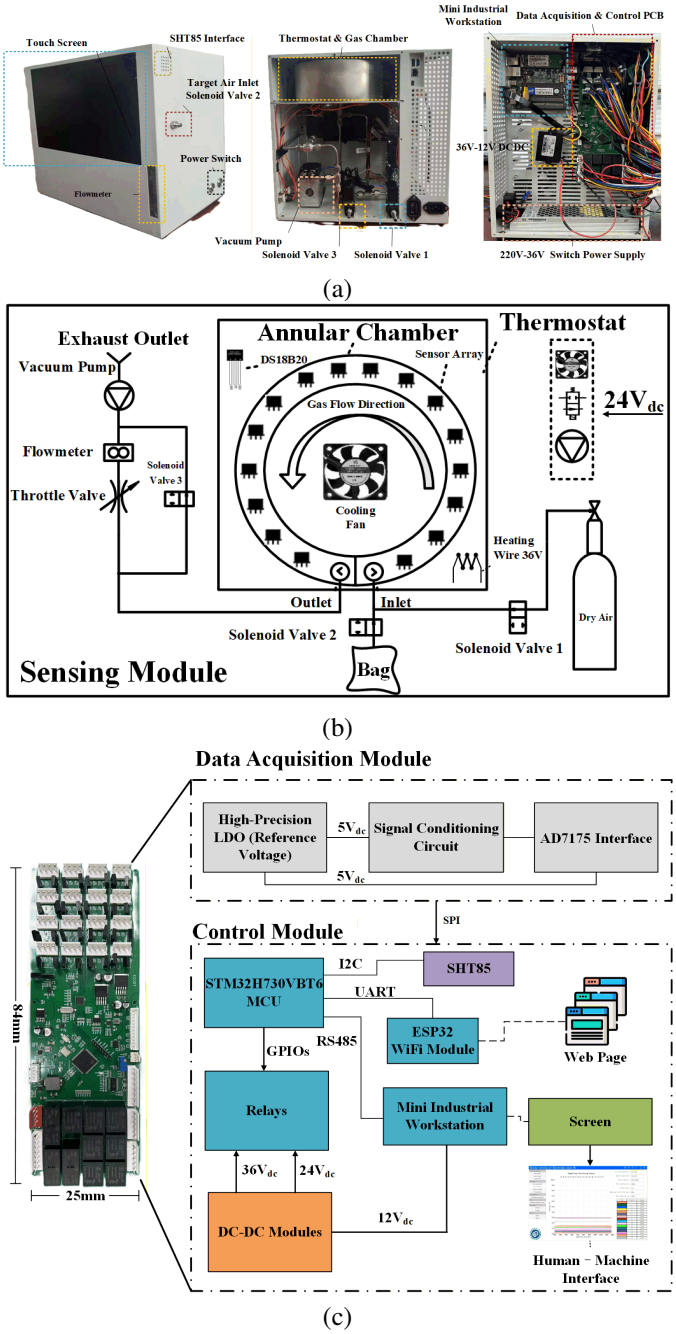


Fig. 1. (a) Photos of the proposed E-nose instrument with dimensions 420mm (width), 380mm (height) and 230mm (length). (b) Schematic diagram of the sensing module. (c) Hardware architecture and detailed information.

minimizes internal volume by restricting gas flow to confined spaces around each gas sensor. To further investigate the flow condition inside the gas chamber and consider the presence of gas sensors as the disturbance source, we employ a low-Reynolds number k-ω model based on COMSOL Multiphysics to simulate the computational fluid domain inside the gas chamber. Fig. 2(a) shows the flow field of the simulation. The governing equations are as follow:

$$\rho(u_2 \cdot \nabla)\epsilon = \nabla \cdot \left[\left(\mu + \frac{\mu_T}{\sigma_\epsilon}\right)\nabla\epsilon\right] + C_{\epsilon_1}\frac{\epsilon}{k_2}P_k - C_{\epsilon_2}\rho\frac{\epsilon^2}{k_2} \quad (1)$$

TABLE I
SENSOR SPECIFIC INFORMATION

Sensor Index	Model	Sensitivity to
1	TGS2600	Hydrogen, Carbon Monoxide
2-3	TGS2602	Ammonia, Hydrogen Sulfide, toluene
4-5	TGS2603	Amine, Sulfurous Odors
6	TGS2611	Alcohol, Methane, Natural Gas
7	TGS800	Carbon Monoxide, Methane, Isobutane
8	TGS813	Methane, Propane, Butane
9	TGS816	Methane, Propane, Butane
10	TGS821	Hydrogen
11	TGS822	Ethanol, Organic solvent vapors
12	TGS823	Ethanol, Carbon monoxide
13-14	TGS826	Ammonia
15-16	TGS832	Refrigerant gases
17	DS18B20	Temperature, $\pm 0.5\,^{\circ}$C
18	SHT85	Temperature, Humidity, 20∼50 $\pm 0.1\,^{\circ}$C, 20∼80 $\pm 1.5\,^{\circ}\%$ RH

$$\mu_T = \rho C_\mu \frac{k_2^2}{\epsilon} f_\mu(\rho, \mu, k_2, \epsilon, l_w) \tag{2}$$

$$\nabla G_2 \cdot \nabla G_2 + \sigma_w G_2(\nabla \cdot \nabla G_2) = (1 + 2\sigma_w) G_2^4 \tag{3}$$

where ρ represents the fluid density, u_2 the velocity vector, μ the molecular viscosity, μ_T the turbulent viscosity, k_2 stands for the turbulent kinetic energy, ϵ the turbulent dissipation rate, P_k the turbulent kinetic energy production term, σ_ϵ represents the turbulent Prandtl number for ϵ, C_{ϵ_1}, C_{ϵ_2} and $C_\mu = 0.09$ are the model constants. f_μ is the Low-Reynolds number correction function, and G_2 the auxiliary scalar field. Sixteen MOS sensors are uniformly arranged along the annular flow channel. The inlet and outlet pipes possess a diameter of approximately 3 mm.

Subsequently, the model was meshed using fluid dynamic grid elements, with sizes ranging from 1.2 mm to 3.9 mm. The inlet boundary condition was set to a flow rate of $2.5 \times 10^{-5}\,\mathrm{m}^3/\mathrm{s}$ for the target gases and $10 \times 10^{-5}\,\mathrm{m}^3/\mathrm{s}$ for purified dry air during the purge of the E-nose. The outlet boundary condition was defined as a free stream with a pressure of 0 Pa. To simulate the internal conditions of the gas chamber, all other surfaces of the model were designated as walls by default. The simulation time step was set to 1 s. CFD simulations were conducted to analyze the gas dynamics in the E-nose chamber, assuming an initial hydrogen sulfide (H_2S) concentration of 5 ppm. Fig. 2(b) illustrates the simulated flow velocity distributions during clean air purging (6L/min), while the simulation result of clean air purging, as shown in Fig. 2(c) indicates that after 40 seconds of air injection, the residual H_2S concentration dropped less than 2×10^{-7} ppm and was confined only to certain corners near the outlet region of the gas chamber. Our simulation result further reveals that over 90% of the hydrogen sulfide in the gas chamber was flushed out within 5 seconds. As shown in Fig. 3(a), we track the time-dependent changes in the volume fraction of H_2S at three points within the chamber, the corresponding results are shown in Fig. 3(b), it can be observed that the volume fractions at all three points rapidly decreased to values below the sensor's

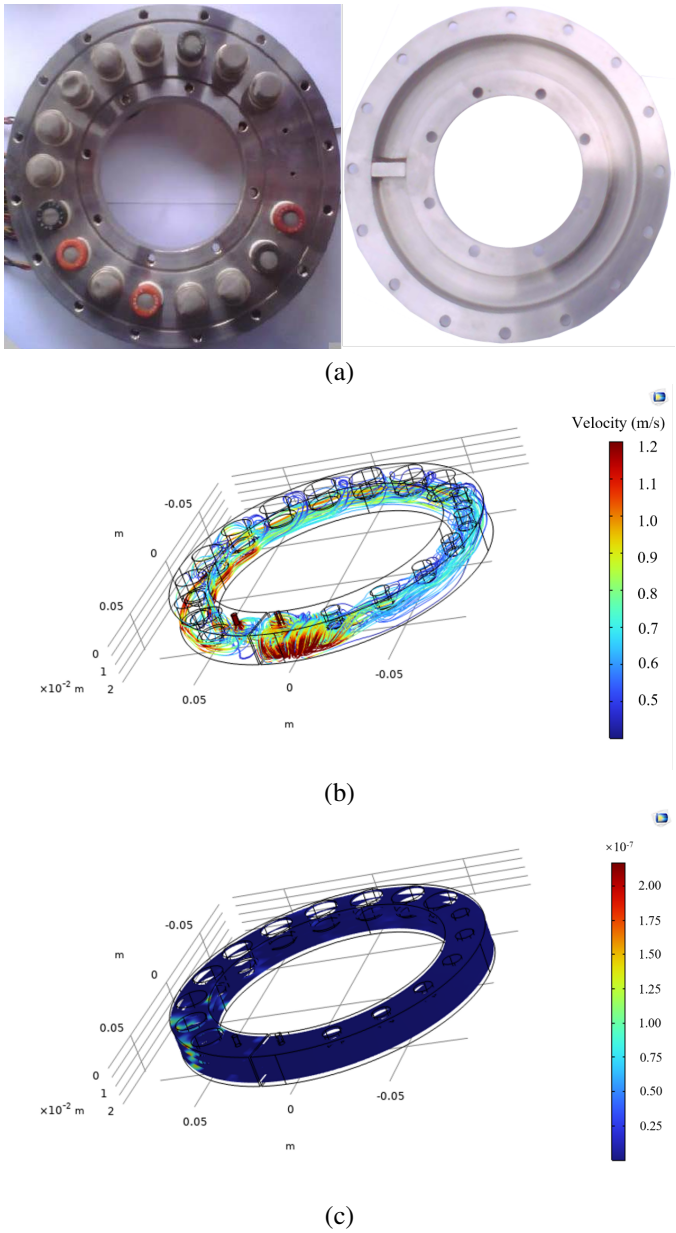
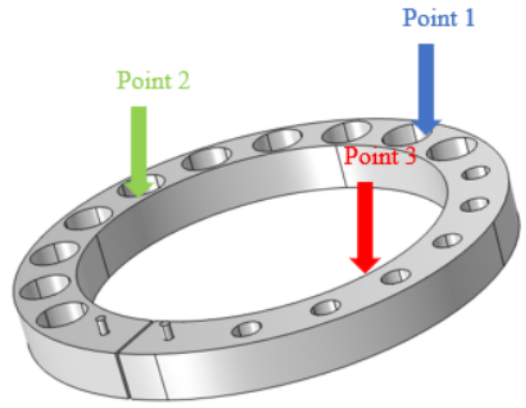


Fig. 2. (a) Photos of the annular gas chamber (with the sensor array). (b) Flow velocity distribution during clean air purging at 6L/min. (c) Distribution of residual H_2S concentration (ppm) after 40 seconds of clean air purging.

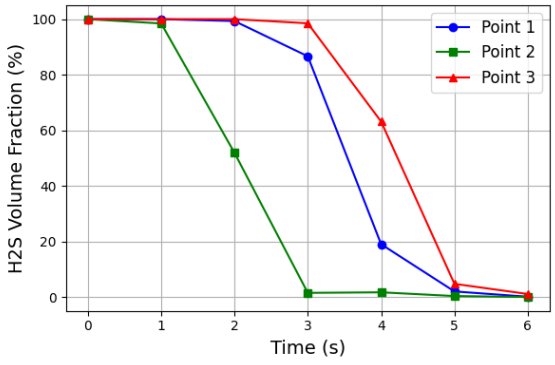
detection limit—well below 0.1 ppm, indicating that the H_2S had been completely purged.

TABLE II
THE STATE OF SOLENOID VALVES UNDER DIFFERENT WORKING PHASES OF E-NOSE

Phase	Solenoid Valve 1	Solenoid Valve 2	Solenoid Valve 3	Flow rate(l/min)	Duration(s)
Restoration	On	Off	On	6.0	120
Equilibrium	Off	Off	Off	0	5
Sampling	Off	On	Off	1.5	40
Purge	On	off	On	6.0	60



(a)



(b)

Fig. 3. (a) Flow field of the simulation and positions of the three monitoring points. (b) Time-dependent H_2S volume-fraction variations at the three monitoring points.

B. Design of the Data Acquisition and Control Module

The data acquisition and control functionalities in this work are implemented through a custom PCB, embedded software, and a user interface program as shown before in Fig. 1(c). This module handles gas sensor data collection, transmits it to the host computer, and controls relay switching to operate solenoid valves for workflow control. Finally, the system feeds the acquired gas data into the detection model, providing real-time feedback on malodorous gas detection and recognition.

1) Hardware design: The data acquisition module consists of a driver circuit and a signal conditioning circuit integrated on the designed PCB. The driver provides stable 5 V supplies for the MOS sensors, heating elements, and environmental sensors. Sixteen independent analog front-end (AFE) channels convert the resistance changes of the gas sensors into voltage signals through a divider configuration, followed by filtering and buffering to improve signal quality. These signals are digitized by a 24-bit ADC (AD7175), with single-point grounding to minimize noise.

The control module includes DC-DC converters supporting both 36V battery input and 220V AC through an external adapter. At the core of the control architecture resides an ARM STM32H730VBT6 microcontroller unit (MCU) for data transmission and controlling peripherals. Specifically, the MCU communicates with AD7175 via SPI to receive raw voltage

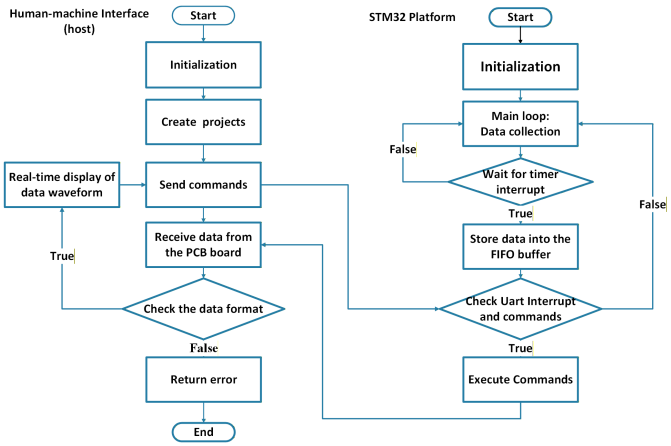


Fig. 4. Flow chart of the software.

data from the gas sensors, and interfaces with the DS18B20 temperature sensor using a one-wire protocol, as well as the SHT85 humidity and temperature sensor via I^2C. Then, all the data are collected at a 10Hz rate and packaged into a 10-depth first-in-first-out (FIFO) buffer and then transmitted to the PC and Wi-Fi module through UART using Direct Memory Access (DMA). Human-machine interaction is realized through RS-485 communication, while GPIOs drive relays for external peripherals such as fans, pumps, heaters, and solenoids.

Moreover, an ESP32 PICO-D4 is deployed as a co-processor and Wi-Fi module to enable IoT capabilities and wireless connectivity. The full circuit schematics are provided in the Supplementary Material.

2) Software design: A lightweight human–machine interface, developed with QT C++ on an industrial PC with a touch screen, enables data logging, algorithm execution, and real-time monitoring. Fig. 4 represents the flow chart of the software proposed in this work.

For system reliability, a command-response verification mechanism is implemented: When the host computer sends relay control commands through the RS-485 human-machine interface, the MCU processes these commands as external interrupts, which then return the updated GPIO status to the host for cross-checking. If a mismatch is detected, the system halts operations instantly and triggers an error alert. Simultaneously, all incoming sensor data packages undergo protocol validation and checksum verification during unpacking to ensure data integrity, maintaining reliability across both control and measurement workflows.

III. Method

The preprocessed response signals from the E-nose instrument are used as the input to the proposed MS-TS-DDA network. As illustrated in Fig. 5, the network is composed of L repeated MS-TS-DDA blocks, where each block contains a multi-scale convolutional extractor, a temporal-shift and depth dynamic aggregation (TS-DDA) module, and an anti-aliasing pooling unit. Furthermore, an initial stem layer is applied before the cascaded blocks, and a final classification head is

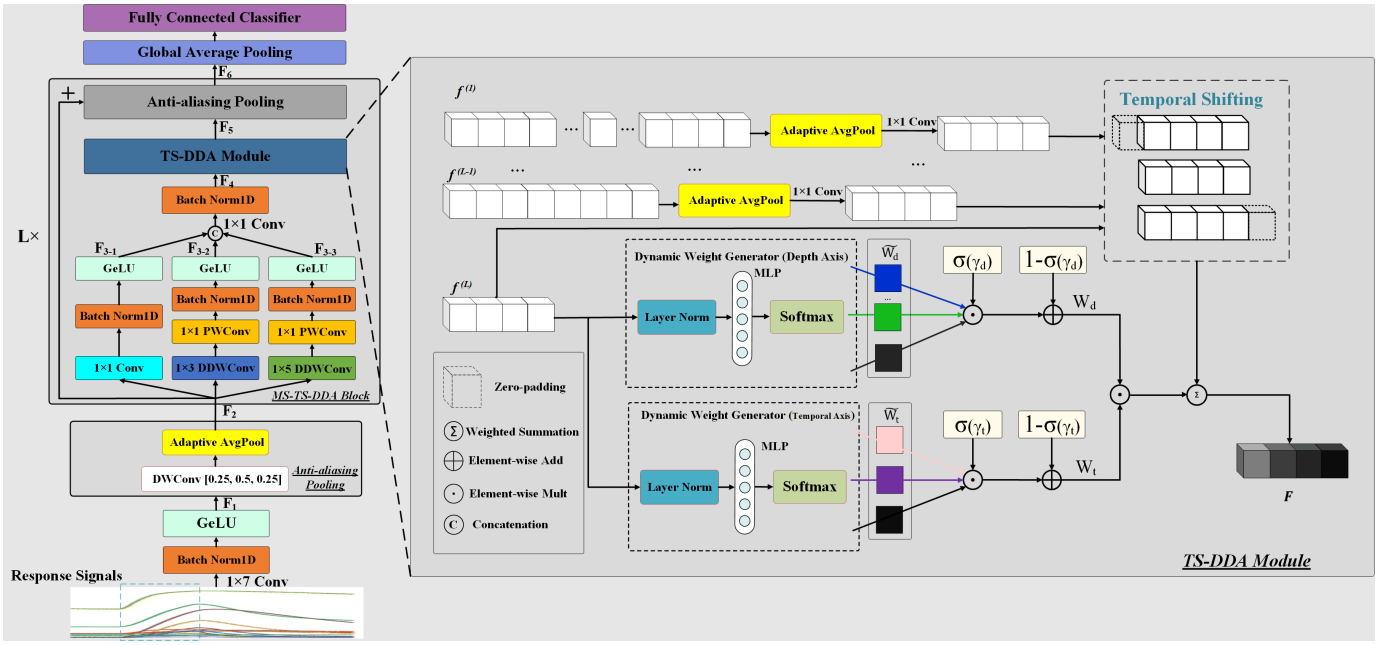


Fig. 5. The proposed MS-TS-DDA framework for malodorous gas classification.

appended after the final block to produce the gas-category prediction. The multi-scale convolutional extractor adopts depth-wise separable convolutions combined with dilated kernels to efficiently extract temporal features across different receptive fields scales, allowing the network to capture the subtle, low-concentration gas dynamics in a compact and effective manner. Building upon the multi-scale convolutional features, the TS-DDA module integrates two core mechanisms: a temporal-shift dynamic aggregation (TSDA) mechanism that establishes cross-time interactions with a very low parameter overhead, and a depth dynamic aggregation (DDA) mechanism that adaptively fuses multi-depth features through learned weights, allowing the network to simultaneously model temporal dependencies and depth-dependent feature coherently. To further improve feature stability, an anti-aliasing pooling unit is incorporated to suppress high-frequency artifacts introduced by down-sampling and to preserve the integrity of temporal patterns. Finally, the aggregated representation is passed through a global average pooling layer and a fully connected classifier to generate the final classification results.

A. Stem Layer and Anti-aliasing Down-sampling Module

Given an E-nose signal $X_{in} \in \mathbb{R}^{C\times T}$ as the input, where C is the sensor channel and T is the time steps, a stem layer consisting of a 1×7 convolution, batch normalization (BN), and GeLU activation function is applied to down-sample the input signals and expand the receptive field rapidly, following common practices in residual network layers [24]. The output of the stem layer F_1 can be expressed as:

$$F_1 = \text{GeLU}\left(\text{BN}(\text{Conv}_{1\times 7}(X_{in}))\right) \tag{4}$$

where $\text{Conv}_{1\times 7}$ denotes a 1-D convolution operation that employs convolution kernels of size 1×7. After the stem layer, an anti-aliasing pooling unit composed of a fixed low-pass depth-wise filter followed by adaptive average pooling (AAP) is applied to suppress high-frequency artifacts caused by down-sampling [25]. This preserves coarse dynamic trends while mitigating aliasing distortions. The anti-aliasing pooling can be expressed as follows:

$$F_2 = \text{AAP}(\text{DWConv}_{LPF}(F_1)) \tag{5}$$

where DWConv_{LPF} is a fixed depth-wise convolution implementing the low-pass filter [0.25, 0.5, 0.25], following the anti-aliasing design BlurPool [25].

B. Multi-Scale Convolutional Extractor

The multi-scale convolutional extractor is designed to extract rich temporal features from the preprocessed E-nose responses while keeping the model lightweight. To this end, we adopt a multi-branch architecture built on depth-wise dilated separable convolutions. Each branch captures gas-response dynamics at a different temporal scale, and their outputs are fused to form a compact yet expressive representation that serves as the input to the subsequent TS-DDA module.

1) 1D Depth-wise Dilated Separable Convolution: To efficiently enlarge the temporal receptive field while keeping the parameters low, depth-wise dilated separable convolution is utilized in each branch of the multi-scale conventional extractor. This technique separates the convolution process into two stages: a depth-wise dilated convolution and a point-wise convolution, as illustrated in Fig. 6. Firstly, a depth-wise dilated convolution is applied with separated and dilated filters to each input channel independently, thereby reducing the computational cost and parameters and enlarging the receptive field compared to traditional convolutions where filters are applied across all channels simultaneously. Then a 1×1

TABLE III
COMPARISON OF PARAMETERS, FLOPS AND RECEPTIVE FIELDS BETWEEN DIFFERENT 1D-CONVOLUTIONS.

Convolution type	Parameter	FLOPs (Computational Cost)	Receptive Field
Traditional	$KC_{in}C_{out}$	$KC_{in}C_{out}T$	K
Depth-wise separable	$KC_{in}+C_{in}C_{out}T$	$KC_{in}T+C_{in}C_{out}T$	K
Depth-wise dilated separable	$KC_{in}+C_{in}C_{out}T$	$KC_{in}T+C_{in}C_{out}T$	$K+(K-1)(D-1)$

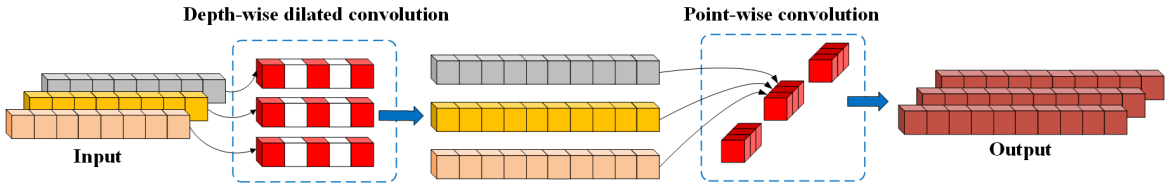


Fig. 6. Depth-wise dilated separable convolution.

point-wise convolution is applied to the output of the depth-wise dilated convolution to mix information across different channels. This approach improves efficiency while maintain a large receptive field, as shown in Table III, which compares the parameter count, computational cost, and receptive field size of traditional convolution, depth-wise separable convolution and depth-wise dilated separable convolution, given the input channel C_{in}, output channel C_{out}, kernel size $1\times K$, dilation rate of D and input signal length of T. It can be observed that depth-wise dilated separable convolution achieves a larger receptive field compared with conventional methods while maintaining fewer parameters.

2) Multi-Scale Local Feature Extraction: The local feature extractor is composed of three parallel branches. Each branch consists of a depth-wise dilated separable convolution with its own kernel size and dilation rate, corresponding to different effective receptive field [20]. In this way, the 1×1 convolution focuses on fine-grained local variations, middle-sized convolution kernel emphasizes medium-range dynamics, and the large kernel captures longer-term temporal dependencies within a single gas exposure. The feature maps are then concatenated along the channel dimension and projected by a 1×1 convolution, yielding a unified multi-scale representation. The outputs of the three branches are denoted by F_{3-1}, F_{3-2} and F_{3-3}, which can be expressed as follows:

$$F_{3-1}=\text{GeLU}\left(\text{BN}(\text{Conv}_{1\times1}(F_2))\right) \tag{6}$$

$$F_{3-2}=\text{GeLU}\left(\text{BN}\left(\text{PWConv}_{1\times1}\left(\text{DDWConv}_{1\times3}^{d=2}(F_2)\right)\right)\right) \tag{7}$$

$$F_{3-3}=\text{GeLU}\left(\text{BN}\left(\text{PWConv}_{1\times1}\left(\text{DDWConv}_{1\times5}^{d=3}(F_2)\right)\right)\right) \tag{8}$$

where $\text{PWConv}_{1\times1}$ denotes the point-wise convolution and $\text{DDWConv}_{1\times K}^{d}$ denotes the depth-wise dilated convolution that applies $1\times K$ kernels with dilation rate of d independently to each channel. The branch outputs are then concatenated and fused through a 1×1 convolution:

$$F_4=\text{BN}\left(\text{Conv}_{1\times1}\left(\text{Concat}\left(F_{3-1},F_{3-2},F_{3-3}\right)\right)\right) \tag{9}$$

C. TS-DDA Module

Although the multi-scale convolutional extractor combined with dilated convolution provides a rich set of multi-scale temporal features and expands the receptive field without increasing parameters, it still suffers two inherent limitations. First, the receptive fields produced remain static; the kernel sizes and dilation rates are fixed and cannot adapt to the varying temporal characteristics of different gas-response patterns. Second, the extractor cannot jointly model temporal dependencies together with the feature relationships that exist across the outputs of successive convolutional blocks. To address these limitations, we propose the TS-DDA module, which introduces adaptive modeling along both dimensions. The TSDA mechanism constructs flexible temporal dependencies by mixing shifted and non-shifted feature paths. The DDA mechanism further performs input-adaptive fusion between the current block output and the features propagated from previous blocks. The TS-DDA module jointly models temporal dependencies and hierarchical feature relationships, thereby complementing the static receptive fields of the multi-scale convolutional extractor and improving overall performance.

1) Temporal-shift Dynamic Aggregation: The TSDA module introduces input-adaptive temporal receptive fields by mixing shifted and original feature paths. In this way, the proposed TSDA enables the network to construct dynamic temporal dependencies at negligible parameter cost. Fig. 7 illustrates the structure of the TSDA module. Given an input feature map $f\in\mathbb{R}^{C\times T}$, we first define a symmetric set of temporal offsets $\Delta T=\{-K,...,-1,0,1,...,K\}$. The collection of shifted features is given by:

$$\{F^{(\Delta k)}\}=\{\text{Shift}(f,\Delta k)\},\quad \Delta k\in\Delta T \tag{10}$$

where $\text{Shift}(\cdot,\Delta k)$ is the shift operator, the empty positions of the shifted features will be padded by all-zero tensors. To adaptively fuse these shifted features, a set of dynamic weights is generated from the global context of input feature f, layer normalization (LN) along the channel dimension followed by multilayer perceptron (MLP) and SoftMax activation function to obtain the distribution weights of the shifted features. The

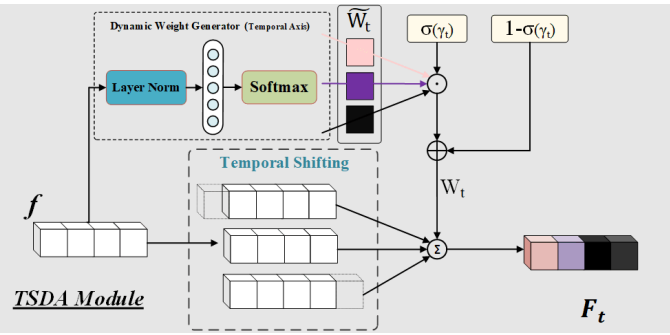


Fig. 7. The proposed TSDA structure.

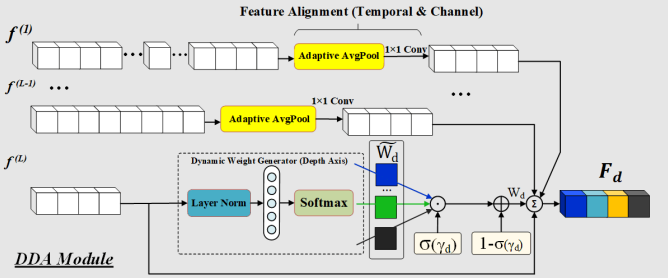


Fig. 8. The proposed DDA structure.

calculation can be expressed as:

$$\tilde{W}_t(\Delta k) = \mathrm{SoftMax}_{\Delta k}\left(\mathrm{MLP}(\mathrm{LN}(f))\right), \quad \Delta k \in \Delta T \tag{11}$$

In addition, a learnable scalar gate $\sigma(\gamma_t) \in [0,1]$ is introduced to control the strength of the dynamic fusion. The effective weight for each offset Δk is defined as:

$$W_t(\Delta k) = \sigma(\gamma_t)\tilde{W}_t + 1 - \sigma(\gamma_t), \quad \Delta k \in \Delta T \tag{12}$$

and the final output of TSDA is computed as weighted aggregation of all shifted branches:

$$F_t = \sum_{\Delta k \in \Delta T} W_t(\Delta k) F^{(\Delta k)}, \quad \Delta k \in \Delta T \tag{13}$$

where $F_t \in \mathbb{R}^{C\times T}$ has the same shape as the input feature f.

2) Depth Dynamic Aggregation: In order to dynamically fuse features at different hierarchical block levels of gas responses, the DDA module learn input-adaptive fusion weights along the depth axis. As illustrated in Fig. 8, let $\{f^{(1)}, ..., f^{(L-1)}, f^{(L)}\}$ denote a set of feature maps extracted from different hierarchical block levels of the network, where each $f^{(l)}$ may have different temporal and channel dimensions. To enable aggregation, these features are first aligned in both the temporal and channel dimensions via adaptive average pooling and 1×1 convolutions. Specifically,

$$\hat{f}^{(l)} = \mathrm{Conv}_{1\times1}\left(\mathrm{AAP}(f^{(l)})\right) \in \mathbb{R}^{C\times T}, \; l = 1, ..., L-1 \tag{14}$$

$$\hat{f}^{(l)} = f^{(l)}, \; l = L \tag{15}$$

where L is the number of depths levels participating in DDA and $f^{(L)}$ denotes the feature from the current input. To adaptively determine the importance of each depth level, dynamic weights $\tilde{W}_d(l)$ are computed form the current feature $f^{(L)}$:

$$\tilde{W}_d(l) = \mathrm{SoftMax}_l\left(\mathrm{MLP}\left(\mathrm{LN}(f^{(L)})\right)\right), \; l = 1, ..., L \tag{16}$$

In addition, the learnable scalar gate $\sigma(\gamma_d) \in [0,1]$ is also introduced to modulate the overall strength of dynamic depth aggregation. The effective fusion weights are defined as:

$$W_d(l) = \sigma(\gamma_d)\tilde{W}_d + 1 - \sigma(\gamma_d), \quad l = 1, ..., L \tag{17}$$

Finally, the aggregated depth representation is computed as:

$$F_d = \sum_{l=1}^{L} W_d(l)\hat{f}^{(l)}, \tag{18}$$

where $F_d \in \mathbb{R}^{C\times T}$ has the same shape as each aligned feature map. In this way, DDA performs input-adaptive selection and weighting over multi-level features, effectively realizing a dynamic receptive field along the hierarchical depth of the network.

3) Unified Aggregation: Based on the above TSDA and DDA designs, the proposed unified TS-DDA module jointly aggregates temporal and hierarchical information from the output of the multi-scale extractor. Let f^L denote the fused multi-scale feature of the current block, and $\{\tilde{f}^{(1)}, ...\tilde{f}^{(L-1)}\}$ be the aligned depth features obtained in the DDA branch. For each depth level l and each temporal offset $\Delta k \in \Delta T$, we denote by $\tilde{f}^{(l,\Delta k)}$ the shifted variant of $\tilde{f}^{(l)}$ generated by the TSDA branch. Given the temporal fusion weights $W_t(\Delta k)$ calculated by equation (12) and the hierarchical fusion weights $W_d(l)$ calculated by equation (18), the unified TS-DDA aggregation is formulated as:

$$F_5 = \sum_{l=1}^{L} \sum_{\Delta k \in \Delta T} W_t(\Delta k) W_d(l) \tilde{f}^{(l,\Delta k)} \tag{19}$$

where $F_5 \in \mathbb{R}^{C\times T}$ has the same shape as F_4. In this way, TS-DDA applies temporally and hierarchically adaptive weighting in a separable manner, yielding a dynamically aggregated representation F_5. After that, an anti-aliasing down-sampling and a local residual connection are employed to produce the block output F_6, Specifically, the block input F_2 is projected by a 1×1 convolution and down-sampled by the same anti-aliasing operator with F_5, followed with GeLU activation function, the calculation process can be expressed as:

$$F_6 = \mathrm{GeLU}\begin{pmatrix} \mathrm{AAP}(DW\mathrm{Conv}_{LPF}(F_5)) \oplus \\ \mathrm{AAP}\left(DW\mathrm{Conv}_{LPF}\left(\mathrm{Conv}_{1\times1}(F_2)\right)\right) \end{pmatrix} \tag{20}$$

This output F_6 is propagated to the following block or, for the last stage, further fed into the classification head. The final classification head consists of a global average pooling layer that collapses the temporal dimension into a compact representation, followed by a fully connected layer that produces the final gas-category prediction.

D. Training Strategy and Data Augmentation

To alleviate the class imbalance and limited sample size in the collected malodorous-gas dataset, a temporal oversampling strategy inspired by the temporal prefix-truncation and near-border sample generation mechanism of T-SMOTE [26] is employed. We further adapt it to the multi-sensor time series

E-nose signals by performing phase-aligned window interpolation and physically constrained synthetic sequence generation as detailed below.

1) Giving a training sequence $X \in \mathbb{R}^{16\times400}$ (16 sensor channels, 400 time steps), it is divided into overlapping windows with a length of 100 and a stride of 50 samples, which produces 7 overlapping sub-windows per sequence, denoted as $W_i \in \mathbb{R}^{16\times100}$.

2) For every pair of consecutive windows within the same sample, multi-channel cross-correlations are applied to estimate their relative temporal alignment. We search an integer time shift Δ within a limited range: $\Delta \in [-\Delta_{max}, \Delta_{max}], \Delta_{max} = 8$, which corresponds to ±0.8 s at 10 Hz. This range is sufficient to cover typical short-term response delays without altering the global response pattern. The multi-channel correlation score is defined as:

$$Score(\Delta) = \sum_{c=1}^{16}\sum_{t=1}^{100} W_i^{(c)}(t) W_{i+1}^{(c)}(t+\Delta) \tag{21}$$

where $W_i^{(c)}(t)$ denotes the value of channel c at time index t in windows W_i. Here, $Score(\Delta)$ is the sum of cross-correlations over all channels. we first compute per-channel temporal correlation and then sum them to obtain a joint multi-channel alignment measure. The optimal shift Δ^* is obtained by maximizing the score:

$$\Delta^* = argmax_{\Delta\in[-\Delta_{max},\Delta_{max}]} Score(\Delta) \tag{22}$$

then define the aligned neighbor window as:

$$\tilde{W}_{i+1} = shift(W_{i+1}, \Delta^*) \tag{23}$$

where $shift(\cdot, \Delta^*)$ denotes a temporal shift by Δ^* along the time axis with boundary padding of edge replication.

3) After alignment, a synthetic window is generated by convex interpolation between W_i and $\tilde{W}_{i+1}$ with a random mixing coefficient:

$$\begin{aligned} W_{sym} &= (1-\lambda)W_i + \lambda\tilde{W}_{i+1} + \varepsilon, \\ \varepsilon^{(c)}(t) &\sim N\left(0, \left(\sigma \cdot \text{std}\left(W_i^{(c)}\right)\right)^2\right) \end{aligned} \tag{24}$$

where $\lambda \in \mathcal{U}(0,1)$, and ε is channel-wise Gaussian noise. In addition, to ensure physical plausibility, the synthesized windows are further filtered using simple shape constraints derived from empirical MOS sensor responses. In particular, synthetic candidates exhibiting numerical abnormalities resulting from interpolation or noise injection, negative response values, excessive peak amplitudes, or unrealistically abrupt temporal variations are discarded, ensuring that the synthetic windows remain consistent with realistic sensor response dynamics.

4) The synthetic window W_{sym} is then reinserted into the original sequence by replacing the corresponding temporal region, yielding a synthetic sequence $\tilde{X} \in \mathbb{R}^{16\times400}$ that preserves the original length and global structure while enriching transient patterns. The augmentation process is repeated on multiple overlapping windows with different interpolation factors. As a result, each augmented sample is a unique composite trajectory assembled from multiple perturbed segments, rather than a simple variant of the original sample. This procedure is also repeated for windows of all training samples within each class and concentration bucket until the number of samples in that bucket reaches a balanced target. The temporal oversampling is applied only to the training set in each 5-fold cross-validation split; validation and test sets remain unchanged to avoid any data leakage.

E. Training Hyperparameters

During the proposed MS-TS-DDA network training, the model parameters were optimized using the Adam optimizer with an initial learning rate of 1×10^{-3}, and the cross-entropy loss function was applied to calculate the training error. A cosine annealing learning rate scheduler was adopted to gradually decay the learning rate to 1×10^{-6} over the course of training. The model is trained for 200 epochs with a batch size of 32. All experiments were conducted on the same laptop equipped with an Intel Core i9-13900HX CPU and an NVDIA RTX 4060 GPU.

IV. Experiments and Discussion

A. Datasets and Gas Information Visualization

The first step of the dataset construction is measuring the eight types of malodorous gases with the proposed instrument. Each sample measurement lasts 225 s at a sampling rate of 10 Hz and consists of four working phases, as summarized in Table II. During the data collection process, only the equilibrium, sampling, and purge phases were recorded, resulting in 105 s of valid data per sample, while the restoration phase was excluded. For dataset construction, only the 40 s time window corresponding to the sampling phase was extracted for subsequent model training and evaluation.

The dataset comprises multi-sensor time-series measurements of 8 distinct gas species across 5 concentration levels per gas. Each sample consists of 400-timestep signals (40 s duration at 10 Hz sampling rate) recorded during the sampling phase of the E-nose captured by 16 sensor channels, resulting in a total of 497 samples. A summary of the dataset is provided in Table IV. After extracting the sampling window, a baseline correction step was applied. Specifically, the response signal is computed as:

$$V_{res} = V_{gas} - V_{air} \tag{25}$$

where V_{res} denotes the baseline-corrected sensor response, and V_{gas} is the raw voltage measured during the sampling phase, and V_{air} is the baseline voltage recorded during the purified-air equilibrium phase. This operation ensures that the extracted sequence reflects only the gas-induced variation.

In the context of MOS-based E-nose systems, volatile compounds are typically detected at ppm and even sub-ppm levels, which are generally regarded as low-concentration regimes [27]. In this work, the target gas concentrations (0.5–60 ppm) fall into this low-concentration range with over 90% of the collected samples distributed below 5 ppm. In this case, the sensor responses become relatively weak, while the noise and drift effects are more observable.

Fig. 9(a) presents the PCA visualization of the original E-nose dataset. The first two principal components explain

only 24.72% and 8.39% of the total variance, respectively, indicating that the intrinsic data structure is high-dimensional and cannot be effectively represented by a low-dimensional linear subspace. As can be observed, samples from different gas categories are largely overlapped in the PCA space, and the inter-class distances are generally small compared with the intra-class variations. This phenomenon is particularly pronounced under low-concentration conditions, where sensor responses are weak and highly correlated across channels. These observations suggest that simple global features or linear projection-based methods are insufficient to discriminate different gas types, thereby motivating the use of multi-scale and dynamically adaptive temporal feature modeling in subsequent analysis.

To further qualitatively illustrate the effect of the proposed data augmentation strategy, Fig. 9(b) shows the t-SNE visualization of the training samples from a representative fold. As observed, the augmented samples are distributed around the original samples of the same gas category, forming locally expanded neighborhoods rather than isolated clusters. This indicates that the proposed augmentation strategy preserves the intrinsic class structure while increasing intra-class diversity.

TABLE IV
SAMPLE GAS CONCENTRATION RANGE AND QUANTITY

Class	Gas composition	Concentration range (ppm)	Number
0	H_2S (Hydrogen sulfide)	0.5, 1, 2, 4, 5	77
1	C_3H_9N (Trimethylamine)	0.5, 1, 2, 4, 5	52
2	CH_4S (Methyl mercaptan)	0.5, 1, 2, 4, 5	70
3	C_2H_6S (Methyl sulfide)	0.5, 1, 2, 4, 5	63
4	$C_2H_6S_2$ (Dimethyl disulfide)	0.5, 1, 2, 4, 5	55
5	CS_2 (Carbon disulfide)	1, 2, 3, 4, 5	53
6	C_8H_8 (Styrene)	0.5, 1, 2, 4, 5	61
7	$C_4H_{10}O$ (N-butanol)	2, 4, 8, 12.5, 60	66

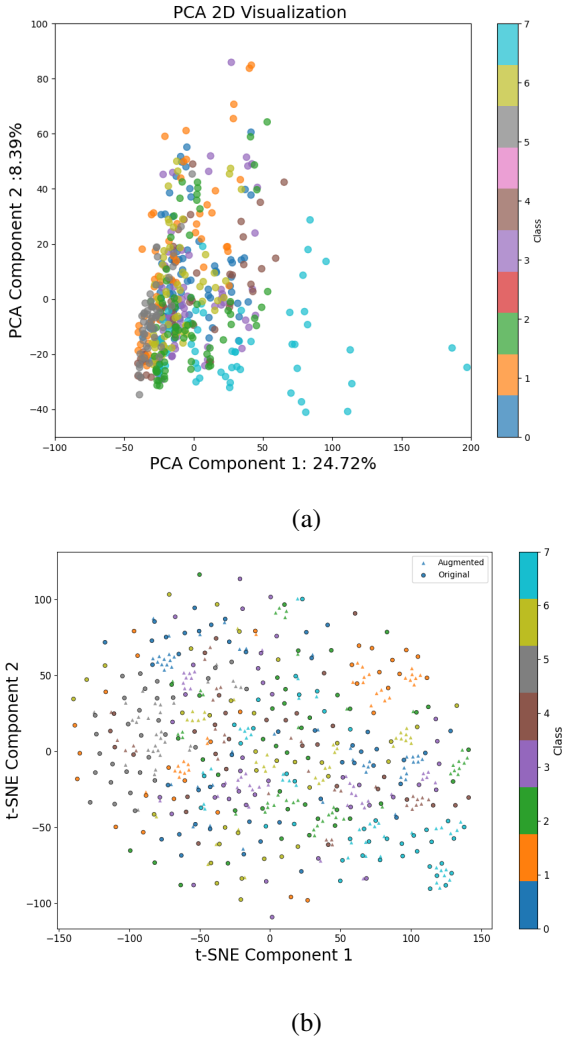


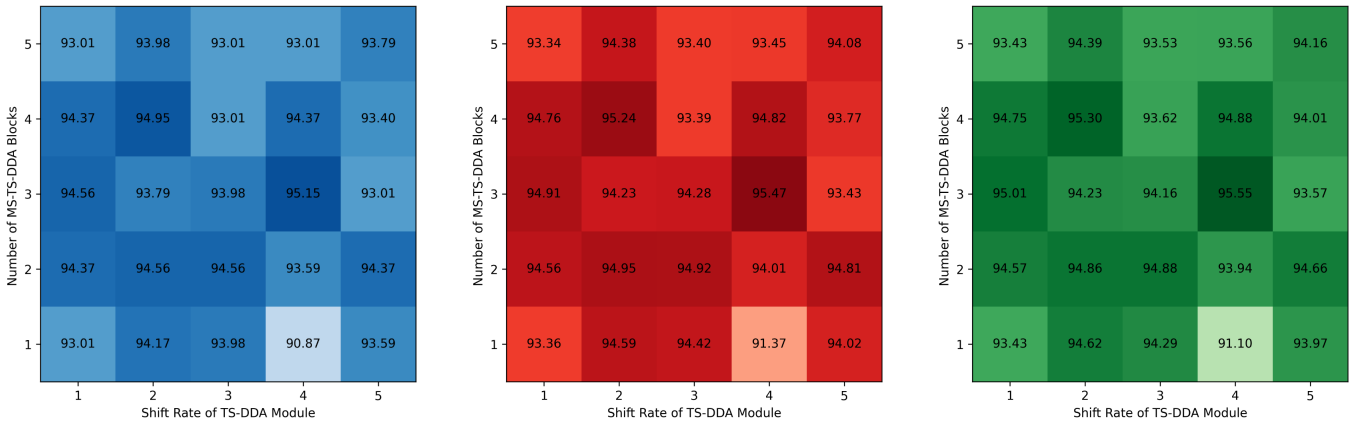
Fig. 9. (a) PCA dimension reduction of eight malodorous gases under low-concentration. (b) t-SNE visualization of original and augmented training samples from the first fold of the cross-validation.

B. Optimal model structure analysis

To balance the trade-off between model complexity and classification performance, a systematic grid search over the number of MS-TS-DDA blocks L and the temporal shift rate K is conducted. The number of blocks directly influence both computational cost and the network's capacity for hierarchical feature extraction. Meanwhile, the shift rate K determines the scope of temporal receptive fields formed by the TS-DDA module. Therefore, the value of MS-TS-DDA blocks and the temporal shift rate are set to 1, 2, 3, 4 and 5. Fig. 10 represents the optimization results of the MS-TS-DDA network structure under different number of MS-TS-DDA blocks and temporal shift rate. When the block number L is set to 3 and the temporal shift rate to 4, the MS-TS-DDA network achieves the best accuracy, F1-score and recall. In terms of the block number, too few blocks fail to fully capture the hierarchical and nonlinear gas features of low-concentration gas responses, while too many blocks introduce redundant parameters and the risk of overfitting. Additionally, a small shift rate K enables the model to focus on short-term temporal variations, whereas a large K allows it to incorporate longer temporal dependencies. However, excessively large shifts may distort local temporal alignment, thus affecting the classification performance. For most shift rates, the overall classification performance improves with increasing depth until reaching a moderate number of blocks, after which deeper configuration tend to yield reduced or unstable gains. Based on these results, we identify the optimal structure as consisting of 3 blocks with a temporal shift rate of 4.

C. Ablation Study

In the MS-TS-DDA network, the anti-aliasing pooling is applied after the stem layer and within the MS-TS-DDA blocks to suppress high-frequency artifacts during down-sampling. The TSDA module is used to expand the temporal receptive field and capture subtle temporal dependencies present in low-ppm gas signals. In addition, the DDA module enhances hierarchical feature fusion. To prove the importance of these three mechanisms for improving the classification performance of the network, ablation experiments are conducted. Table V shows the quantitative results of the ablation experiments.



(a) (b) (c)

Fig. 10. Optimization results of the MS-TS-DDA network structure for different numbers of MS-TS-DDA blocks and shift rate of TS-DDA module. (a)Average accuracy (%). (b) Average macro-F1(%). (c) Average macro-recall (%).

Since the anti-aliasing pooling is a standard component, we mainly focus on evaluating the contributions of the proposed TSDA and DDA modules. Therefore, the ablation experiments are divided into six cases. By comparing the classification results for the baseline model, baseline combined with anti-aliasing pooling, the variants that individually enable TSDA or DDA module, TS-DDA, and finally the full model integrating all three mechanisms, we demonstrate the contribution and effectiveness of each component. Relative to the baseline model, enabling anti-aliasing pooling alone improves 2.91% in average accuracy, with only a marginal increase in parameters and computational cost. Activating the TSDA module alone yields the largest individual performance gain, improving average classification accuracy by 4.47%, macro F1-score by 4.47%, and macro-recall by 4.14% while incurring only a negligible increase in parameters (from 81.93K to 83.94K) and FLOPs (from 9.23M to 9.49M). Similarly, compared to the baseline model, enabling the DDA module alone improves average accuracy by 3.88%, macro F1-score by 3.88% and macro-recall by 3.56% with a moderate increase in parameters and FLOPs. When both the TSDA and DDA modules are activated simultaneously, the network achieves improvements of 4.66% in average accuracy, 4.48% in macro F1-score and 4.26% in macro-recall, demonstrating the complementary nature of temporal and depth-wise dynamic aggregation. Moreover, with the fusion of all three mechanisms, the MS-TS-DDA network achieves the best performance, with an average accuracy of 95.15%, an average macro F1-score of 95.57%, and an average macro-recall of 95.55% while maintaining a relatively low computational cost of 105.32K parameters and 11.12M FLOPs.

In addition to quantitative performance comparisons, we further investigate how different ablation variants affect the feature utilization behavior of the network through gradient-weighted class activation mapping (Grad-CAM) to visualize the gas information processed by three representative variants: the baseline network with DDA, the baseline network with TSDA and the full TS-DDA model, using the same sample as input. Fig. 11 presents the visualization results of gas features. The DDA model focuses on sparse and localized important responses, emphasizing a limited number of sensor channels at specific time instants because the DDA module is designed to enhance cross-layer feature selection rather than long-range dependencies. In contrast the TSDA variant exhibits continuous important patterns along the temporal dimension, demonstrating its ability to enlarge the receptive field and capture long-term response dynamics. When both mechanisms are jointly enabled, the resulting importance maps become significantly more structured and interpretable. The model focuses on a small subset of informative sensor channels while maintaining continuous attention over extended temporal regions, indicating the complementary effect of the DDA and TSDA modules. In summary, the proposed MS-TS-DDA network effectively highlights the long-range key features that impact gas information classification performance, and its effectiveness has been comprehensively validated.

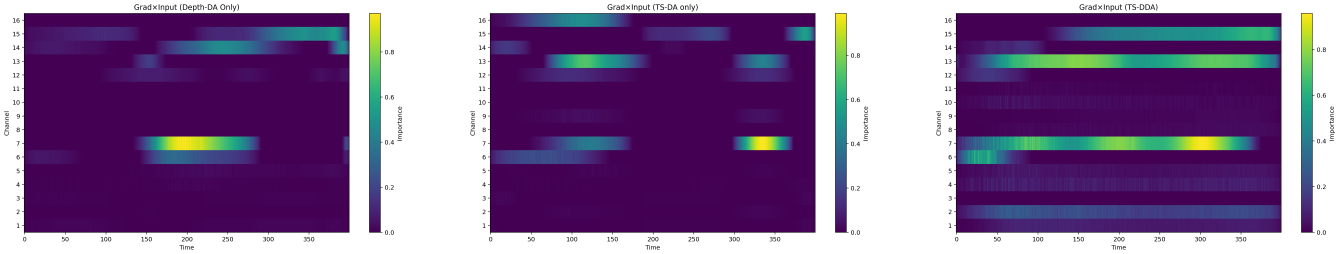
Overall, the ablation results indicate that the TSDA module contributes the largest individual improvement, while the DDA module provides complementary hierarchical information fusion. Their combination yields additional gains, confirming that the proposed TSDA and DDA modules address different aspects of feature modeling. The anti-aliasing pooling further stabilizes temporal representations during down-sampling. In summary, the proposed MS-TS-DDA network effectively enhances temporal feature quality and dynamic aggregation behavior.

D. Performance Comparison

To evaluate the performance advantages of the proposed MS-TS-DDA network, we compare it with four categories of representative classification methods, including traditional machine-learning-based classifiers, classic convolutional neural networks, the classic network of transformer encoder, and state-of-the-art gas information analysis models. The machine-learning classifiers include Support Vector Machine with an RBF kernel (SVM-RBF), Random Forest (RF), and K-Nearest Neighbors (KNN, K=1). The CNN baselines consist of 1D-ResNet18 [24], 1D-DenseNet121 [28], 1D-GoogLeNetV1

TABLE V
COMPARISON RESULTS OF ABLATION EXPERIMENTS. (AVERAGE PERFORMANCE AND CORRESPONDING STANDARD DEVIATIONS)

AA	TSDA	DDA	Accuracy (%)	Macro F1-score (%)	Macro Recall (%)	Params(K)	Flops(M)
			90.29 ± 3.50	90.62 ± 3.35	90.99 ± 3.22	81.93	9.23
√			93.20 ± 2.94	93.44 ± 2.87	93.67 ± 2.70	83.85	9.52
	√		94.76 ± 0.78	95.09 ± 0.90	95.13 ± 0.85	83.94	9.49
		√	94.17 ± 1.62	94.50 ± 1.62	94.55 ± 1.66	102.25	10.71
	√	√	94.95 ± 1.88	95.10 ± 1.93	95.25 ± 2.03	103.40	10.82
√	√	√	95.15 ± 0.87	95.47 ± 0.93	95.55 ± 0.93	105.32	11.12



(a) (b) (c)

Fig. 11. Visualization results of gas features by using Grad-CAM. (a) DDA only. (b) TSDA only. (c)Full TS-DDA model.

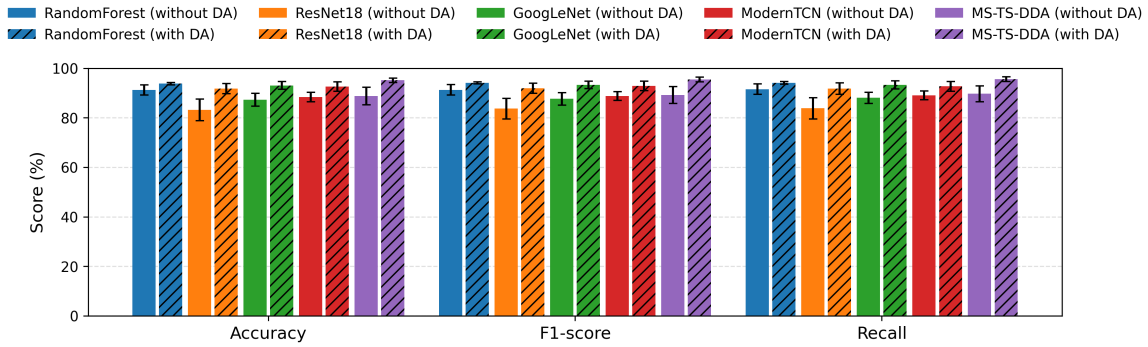


Fig. 12. Comparison of model performance with and without data augmentation (DA) for the proposed MS-TS-DDA model, Random Forest, ResNet18, GoogLeNetV1 and ModernTCN.

[29], TimesNet [30] and ModernTCN [31]. In addition, eight representative state-of-the-art models reported in recent gas-sensing literature are also included. The comparison results are summarized in Table VI.

Due to the lack of a mechanism that jointly and adaptively enlarges the temporal receptive field and dynamically aggregates cross-depth feature representations, the existing state-of-the-art models—despite employing separable convolutions, multi-branch architectures, temporal convolutions, or transformer-style self-attention—struggle to effectively capture the subtle temporal dependencies and hierarchical nonlinear features present in low-ppm gas responses. As a result, they fall short of achieving the same level of classification performance as the proposed MS-TS-DDA network. Moreover, the proposed method not only achieves the highest average accuracy, F1-score, and recall, but also exhibits smaller performance variances across folds, indicating improved robustness and stability in low-concentration gas classification tasks.

To further investigate the impact of data augmentation on different models, we compare the performance of our proposed model with representative deep learning models, 1D-ResNet18, 1D-GoogLeNet-V1, ModernTCN, as well as the best-performing machine learning model, Random Forest. It can be observed from Fig. 12 that data augmentation consistently improves the performance of all learning-based models, demonstrating that the proposed data augmentation strategy effectively enriches the training distribution and alleviates data imbalance. When data augmentation is applied, the MS-TS-DDA model achieves the highest score in all metrics, with Random Forest performing slightly better than other models. In contrast, without data augmentation, 1D-ResNet18 experiences the most observable performance degradation, becoming the worst-performing model without data augmentation. Random Forest, 1D-GoogLeNet-V1, MordernTCN show similar performance without data augmentation, indicating their stronger robustness in handling limited and imbalanced training data compared to 1D-ResNet18. These observations suggest that while data augmentation benefits all models, the proposed MS-TS-DDA architecture is particularly effective in exploiting the additional temporal diversity introduced by

augmented samples.

TABLE VI
COMPARISON RESULTS WITH OTHER CLASSIFICATION METHOD. (AVERAGE PERFORMANCE AND CORRESPONDING STANDARD DEVIATIONS)

Method	Accuracy (%)	F1-score (%)	Recall (%)
SVM-RBF	83.98 ± 2.09	84.68 ± 1.81	85.10 ± 1.69
Random Forest	93.79 ± 0.48	94.02 ± 0.42	94.10 ± 0.51
KNN (K=1)	91.65 ± 2.92	92.11 ± 2.71	92.16 ± 2.76
1D-ResNet18	91.75 ± 2.09	91.90 ± 2.09	91.80 ± 2.31
1D-DenseNet121	91.17 ± 2.55	91.52 ± 2.57	91.68 ± 2.51
1D-GoogLeNet-V1	93.01 ± 1.55	93.26 ± 1.52	93.32 ± 1.63
Transformer-Encoder	92.43 ± 3.27	92.74 ± 3.30	93.12 ± 3.32
ModernTCN [31]	92.62 ± 1.90	92.80 ± 1.91	92.70 ± 1.98
TimesNet [30]	92.04 ± 1.13	92.38 ± 1.16	92.47 ± 1.11
CNN-AE [14]	90.58 ± 1.23	91.45 ± 2.23	91.58 ± 2.03
MCNN [17]	92.04 ± 3.10	92.07 ± 3.12	92.12 ± 3.06
1D-DNR [15]	91.84 ± 3.28	92.21 ± 3.21	92.23 ± 3.40
BM-Net [16]	90.87 ± 1.90	91.24 ± 1.86	91.15 ± 1.93
TEA-CNN [32]	88.93 ± 2.54	89.58 ± 2.49	89.77 ± 2.32
AKCA-Net [33]	87.96 ± 2.09	88.38 ± 2.10	88.43 ± 2.05
GFAN-Net [34]	90.10 ± 5.37	90.54 ± 5.04	90.50 ± 5.11
TETCN [35]	92.82 ± 1.90	93.25 ± 1.86	93.22 ± 1.86
Ours	95.15 ± 0.87	95.47 ± 0.93	95.55 ± 0.93

V. CONCLUSION

In this work, we designed an all-in-one E-nose framework that integrates instrument design, data augmentation, and deep learning-based pattern recognition. Aiming to jointly address the challenges of fast odor absorption and desorption, low-concentration sensing, limited sample size and weakly discriminative temporal responses, the proposed approach provides an effective solution for malodorous gas classification under practical conditions. The main conclusions of this study can be summarized as follows:

1) An all-in-one instrument was developed, integrating a multi-channel MOS sensor array, an annular gas chamber and a data acquisition and control board. The annular chamber design facilitates rapid and uniform gas exchange, while the coordinated hardware-software control ensures stable signal acquisition under low-ppm conditions. This integrated design provides a reliable and repeatable sensing platform for subsequent data collection and gas recognition.

2) A low-concentration gas dataset was established based on the developed instrument. The dataset comprised response signals of eight categories of malodorous gases, forming a challenging dataset characterized by weak responses, high inter-class similarity, and class imbalance. To mitigate these issues, a temporal oversampling strategy tailored to multi-sensor time-series signals was proposed. By combining phase-aligned window interpolation with physically constrained synthetic sequence generation, the augmentation method effectively enriches the training distribution while preserving realistic sensor response dynamics.

3) The MS-TS-DDA network was proposed for low-concentration malodorous gas classification, which introduces TSDA and DDA mechanisms to jointly enhance temporal receptive fields and hierarchical feature fusion. Extensive experiments, including data visualization, structure optimization, ablation studies, performance comparisons with classical machine-learning methods, deep convolutional networks, transformer-based models and recent SOTA gas sensing approaches, demonstrate that the proposed network achieves superior accuracy, achieving an average accuracy of 95.15%, F1-score of 95.47% and recall of 95.55% under 5-fold cross-validation.

In summary, this work establishes a tightly coupled framework that links instrument design, data augmentation, and dynamic feature modeling for low-ppm gas recognition. The proposed approach not only improves classification performance but also provides improved interpretability of learned features. Future work will focus on larger-scale datasets with broader concentration range and long-term sensor drift analysis.

REFERENCES

[1] H.-T. Tran, Q. A. Binh, T. Van Tung, D. T. Pham, H.-G. Hoang, N. S. H. Nguyen, S. Xie, T. Zhang, S. Mukherjee, and N. S. Bolan, "A critical review on characterization, human health risk assessment and mitigation of malodorous gaseous emission during the composting process," *Environmental Pollution*, vol. 351, p. 124115, 2024.

[2] G. Wang, Z. Zhai, J. Geng, M. Han, and F. Lu, "Testing and determination of the olfactory thresholds of the 40 kinds of typical malodorous substances," *Journal of Safety and Environment*, vol. 15, no. 6, pp. 348–351, 2015.

[3] T. W. Lambert, V. M. Goodwin, D. Stefani, and L. Strosher, "Hydrogen sulfide (h2s) and sour gas effects on the eye. a historical perspective," *Science of the total environment*, vol. 367, no. 1, pp. 1–22, 2006.

[4] A. P. v. Harreveld, "Odor concentration decay and stability in gas sampling bags," *Journal of the Air & Waste Management Association*, vol. 53, no. 1, pp. 51–60, 2003.

[5] Z. Yuan, F. Lu, M. Peng *et al.*, "Selective colorimetric detection of hydrogen sulfide based on primary amine-active ester cross-linking of gold nanoparticles," *Analytical Chemistry*, vol. 87, no. 14, pp. 7267–7273, 2015.

[6] D. Yu, Y. Xu, J. M. Regenstein, W. Xia, F. Yang, Q. Jiang, and B. Wang, "The effects of edible chitosan-based coatings on flavor quality of raw grass carp (ctenopharyngodon idellus) fillets during refrigerated storage," *Food Chemistry*, vol. 242, pp. 412–420, 2018.

[7] J. A. Covington, S. Marco, K. C. Persaud, S. S. Schiffman, and H. T. Nagle, "Artificial olfaction in the 21 st century," *IEEE Sensors Journal*, vol. 21, no. 11, pp. 12 969–12 990, 2021.

[8] Y. Wang, X. Yan, S. Wang, S. Gao, K. Yang, R. Zhang, M. Zhang, M. Wang, L. Ren, and J. Yu, "Electronic nose application for detecting different odorants in source water: Possibility and scenario," *Environmental Research*, vol. 227, p. 115677, 2023.

[9] J. Qian, A. Zhang, Y. Lu, J. Zhang, and P. Xu, "A novel multisensor detection system design for odor classification," *IEEE Sensors Journal*, vol. 23, no. 16, pp. 18 624–18 633, 2023.

[10] J.-Y. Wang, Q.-H. Meng, X.-W. Jin, and Z.-H. Sun, "Design of hand-held electronic nose bionic chambers for chinese liquors recognition," *Measurement*, vol. 172, p. 108856, 2021.

[11] M. He, L. Xiong, S. Han, X. Luo, Y. Hou, X. Tang, and B. Zhang, "Rapid detection of mixed odor intensity level in composting plants based on e-nose," *IEEE Sensors Journal*, vol. 23, no. 22, pp. 27 795–27 803, 2023.

[12] W. M. Sanjaya, A. Roziqin, A. W. Temiesela, M. F. B. Zaman, A. Taqwim, I. Opialisti, P. Sintia, A. Mulyawan, D. Anggraeni, and T. Sa'adah, "Developing an electronic nose for formalin detection in meatballs using support vector machine (svm) method and raspberry pi 4," *Physica scripta*, vol. 99, no. 9, p. 096009, 2024.

[13] C. Qu, Z. Zhang, J. Liu, P. Zhao, B. Jing, W. Li, C. Wu, and J. Liu, "Multi-scenario adaptive electronic nose for the detection of environmental odor pollutants," *Journal of Hazardous Materials*, vol. 489, p. 137660, 2025.

[14] Z. Ye, Y. Li, R. Jin, and Q. Li, "Toward accurate odor identification and effective feature learning with an ai-empowered electronic nose," *IEEE Internet of Things Journal*, vol. 11, no. 3, pp. 4735–4746, 2023.

[15] F. Li, Y. Li, B. Sun, H. Cui, J. Yan, P. Feng, and X. Peng, "A novel densenet with warm restarts for gas recognition in complex airflow environments," *Microchemical Journal*, vol. 197, p. 109864, 2024.

[16] Y. Wang, H. Wang, X. Wen, J. Liu, Y. Shi, and H. Men, "Origin identification of angelica dahurica using a bidirectional mixing network combined with an electronic nose system," *Sensors and Actuators B: Chemical*, vol. 429, p. 137356, 2025.
[17] J. Guo, X. Li, X. Li, Z. Liang, J. Cao, and X. Wei, "Anti-drift gas detection algorithm based on neural network," *IEEE Transactions on Instrumentation and Measurement*, 2024.
[18] Z. Zhu, Q. Jiang, M. Wang, M. Xu, Y. Zhang, F. Shuang, and P. Jia, "A co concentration prediction method for electronic nose based on trellisnet with gated recurrent unit and dilated convolution," *Microchemical Journal*, vol. 199, p. 110014, 2024.
[19] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.
[20] X. Chen, X. Xia, J. Zhuo, P. Wu, X. Peng, and J. Chu, "Saddcnhff: A lightweight hybrid network electronic nose system for mixed gas concentration prediction," *Sensors and Actuators B: Chemical*, p. 138733, 2025.
[21] L. Yang, Z. Zheng, Y. Han, H. Cheng, S. Song, G. Huang, and F. Li, "Dyfadet: Dynamic feature aggregation for temporal action detection," in *European Conference on Computer Vision*. Springer, 2024, pp. 305–322.
[22] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 510–519.
[23] A. Rabehi, H. Helal, D. Zappa, and E. Comini, "Advancements and prospects of electronic nose in various applications: a comprehensive review," *Applied Sciences*, vol. 14, no. 11, p. 4506, 2024.
[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
[25] R. Zhang, "Making convolutional networks shift-invariant again," in *International conference on machine learning*. PMLR, 2019, pp. 7324–7334.
[26] P. Zhao, C. Luo, B. Qiao, L. Wang, S. Rajmohan, Q. Lin, and D. Zhang, "T-smote: Temporal-oriented synthetic minority oversampling technique for imbalanced time series classification." in *IJCAI*, 2022, pp. 2406–2412.
[27] A. Khorramifar, H. Karami, L. Lvova, A. Kolouri, E. Łazuka, M. Piłat-Rożek, G. Łagód, J. Ramos, J. Lozano, M. Kaveh *et al.*, "Environmental engineering applications of electronic nose systems based on mox gas sensors," *sensors*, vol. 23, no. 12, p. 5716, 2023.
[28] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.
[29] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.
[30] H. Wu *et al.*, "Timesnet: Temporal 2d-variation modeling for general time series analysis," *The Eleventh International Conference on Learning Representations*, 2023.
[31] D. Luo and X. Wang, "Moderntcn: A modern pure convolution structure for general time series analysis," in *The Twelfth International Conference on Learning Representations*, 2024, pp. 1–43.
[32] G. Ren, R. Wu, L. Yin, Z. Zhang, and J. Ning, "Description of tea quality using deep learning and multi-sensor feature fusion," *Journal of Food Composition and Analysis*, vol. 126, p. 105924, 2024.
[33] H. Sun, Z. Hua, C. Yin, F. Li, and Y. Shi, "Geographical traceability of soybean: An electronic nose coupled with an effective deep learning method," *Food chemistry*, vol. 440, p. 138207, 2024.
[34] Y. Shi, B. Wang, C. Yin, Z. Li, and Y. Yu, "Performance improvement: A lightweight gas information classification method combined with an electronic nose system," *Sensors and Actuators B: Chemical*, vol. 396, p. 134551, 2023.
[35] F. Wu, R. Ma, Y. Li, F. Li, S. Duan, and X. Peng, "A novel electronic nose classification prediction method based on tetcn," *Sensors and Actuators B: Chemical*, vol. 405, p. 135272, 2024.

Chenlong Gu Received the B.E. in information engineering from East China University of Science and Technology, Shanghai, China, in 2023, where he is currently pursuing the master's degree in electronic information technologies and the Ph.D. degree in control science and engineering. His main research interests include hardware design of multi-sensor system, machine learning algorithms and signal processing.

Qianshen Wu Received the B.E. in Energy and Power Engineering from Shanghai Jiao Tong University, Shanghai, China, in 2023, where he is currently pursuing the master's degree in Energy and Power Engineering. His main research interests include the Lattice Boltzmann Method, multiphase flow simulation, and PEM hydrogen production.

Nan Wang (Member, IEEE) received the B.E. degree in computer science from Nanjing University, Nanjing, China, in 2009, and the MS and PhD degrees from the Graduate School of Information, Production and Systems, Waseda University, Shinjuku, Japan, in 2011, and 2014, respectively. He is currently an associate professor with the School of Information Science and Engineering, East China University of Science and Technology, Shanghai, China. His research interests include VLSI design automation, hardware security, network-on-chip and reconfigurable architectures. He is also a member of IEICE.

Yuxuan Zhang (Member, IEEE) received the Ph.D. degree in Electronics from Mid Sweden University, Sundsvall, Sweden, in 2025, the M.Sc. degree in Embedded Systems Engineering from University of Leeds, Leeds, UK, in 2019 and the B.Eng. in New Energy Science and Engineering from Beijing Information Science and Technology University, Beijing, China, in 2018.

He is currently an Assistant Professor in Embedded Systems and IoT with the College of Intelligent Science and Engineering, Beijing University of Agriculture, China. He is also an Affiliated Researcher with the Department of Computer and Electrical Engineering, Mid Sweden University, Sweden. He serves as a reviewer for several leading journals, such as *ACM Computing Survey, IEEE TIM, Meas, IEEE TII, IEEE IoTJ, MSSP, RESS, EAAI, ESWA, NTE, INPA and COMPAG*. He was recognized as an Outstanding Reviewer of *IEEE TIM* in 2024 and 2025, and *INPA* in 2025.

His research interests include Edge AI and IoT for Structural Health Monitoring, Smart Agriculture, and Industrial Maintenance.

Sebastian Bader (Senior Member, IEEE) received the Ph.D. degree in electronics from Mid Sweden University, Sundsvall, Sweden, in 2013, and the Dipl.-Ing. degree from the University of Applied Sciences, Wilhelmshaven, Germany in 2008. He is currently an Associate Professor of embedded systems with the Department of Computer and Electrical Engineering, Mid Sweden University. His research interests focus on energy aspects of embedded systems, including energy harvesting, low-power sensing systems, and machine learning on resource-constrained devices.

Xiaofeng Ling received the B.S. and Ph.D. degrees from Shanghai Jiao Tong University, Shanghai, China, in 2006 and 2012, respectively.,From 2013 to 2015, he served as the Director of research and development in a start-up company. He is currently an Associate Professor with the School of Information Science and Engineering, East China University of Science and Technology, Shanghai. His research interests include wideband array signal processing, wireless communications, and MIMO technique.

Yongjing Wan received the Ph.D. degree in control science and engineering from East China University of Science and Technology, Shanghai, China, in 2008. She is currently a Professor with the Department of Electronic and Communication Engineering, East China University of Science and Technology, Shanghai, China. Her research interest include computer vision and pattern recognition, audio signal processing, speech signal processing and digital image processing.

Daqi Gao received the Ph.D. degree in industrial automation from Zhejiang University, Hangzhou, China, in 1996.He is currently a Professor with the Department of Computer Science, East China University of Science and Technology, Shanghai, China. He has authored or co-authored more than 100 papers. His current research interests include pattern recognition, machine learning, neural networks, and artificial olfactory.