SIS-NET: SENSOR IDENTITY-AWARE SET MODELING FOR SENSOR-CONFIGURATION-AGNOSTIC ELECTRONIC NOSE RECOGNITION

Anonymous authors
Paper under double-blind review

ABSTRACT

Electronic-nose recognition models are commonly trained with a fixed sensor array and therefore implicitly depend on a fixed number, identity, and ordering of input sensors. This assumption is fragile in deployment, where sensors may be unavailable because of hardware variants, maintenance, or externally identified unreliable measurements. We study whether a single model can recognize samples under different subsets of a known sensor universe without configuration-specific retraining. We propose SIS-NET, a sensor-set model that applies a shared temporal encoder to each available sensor, incorporates an explicit sensor identity embedding, and aggregates the resulting variable-size token set with masked mean pooling. During training, sensor subsets are sampled from multiple availability levels so that the model experiences configuration variation before inference. We evaluate the same checkpoint under full, fixed-subset, random-subset, and held-out-combination conditions. The experiments are designed to distinguish configuration-agnostic recognition from fixed-input zero filling, electronic-nose fault-tolerant ensembles, and multivariate time-series baselines.

1 INTRODUCTION

Electronic noses (e-noses) are multisensor systems that characterize volatile chemical environments from the collective responses of gas sensors. They have been applied to a wide range of recognition tasks, including gas identification, food quality assessment, malodour classification, and other chemical sensing applications (Fang et al., 2022; Li et al., 2024; Ozer et al., 2023). Recent systems have also explored data-centric sensor representations and adaptive learning strategies to improve recognition under practical sensing conditions (Sung et al., 2024; Shtepliuk et al., 2025). These developments have made learning-based e-nose recognition increasingly effective. However, they also expose a fundamental property of chemical sensor arrays: individual sensors are often not highly selective to a single compound. Their responses can overlap across different chemical species. As a result, an array with more sensors does not necessarily provide proportionally more useful information.

This property creates an important practical trade-off. Additional sensors may provide complementary measurements, but they can also introduce redundant information, higher hardware cost, calibration effort, maintenance requirements, and greater system complexity. These issues have motivated extensive research on sensor-array optimization and sensor selection. For example, recent methods explicitly search for informative sensor subsets while reducing redundancy (Zhang et al., 2024; Mei et al., 2025). Other studies have investigated how e-nose systems can remain usable when individual sensors fail or become unavailable (Fonollosa et al., 2013; Kaya et al., 2020; Wang et al., 2023). Such methods address important practical constraints. Yet they mainly focus on identifying a desirable sensor array or maintaining recognition under a predefined failure scenario.

Our problem is different. We consider a known sensor universe in which the sensors available at inference time may vary across samples. The goal is not to find one optimal subset and then keep it fixed. Instead, we ask whether a single trained recognition model can operate directly on different numbers and combinations of available sensors. This distinction separates *sensor-array optimization* from *sensor-configuration-agnostic recognition*. The former asks which sensors should be selected.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

The latter asks whether one model can recognize a sample when the currently available sensor subset changes.

This setting is difficult for conventional multivariate time-series models. Most such models are designed around a fixed set of channels. Channel-dependent architectures explicitly model interactions among predefined channels, while widely used temporal models also assume a stable input dimensionality (Wu et al., 2023; Liu et al., 2024; Wang et al., 2024). When a sensor becomes unavailable, a common solution is to remove the sensor, replace its signal, or fill its input with a predefined value. Such operations preserve the original input shape but do not explicitly represent which physical sensor remains available. More importantly, they make the model dependent on the original channel layout. A sensor should therefore be treated not only as a channel position, but also as an identifiable sensing element.

This observation suggests a set-based formulation. Instead of representing an input as a fixed ordered vector of sensor channels, we represent the currently available sensors as a variable-size set of sensor tokens. Set-based representation learning provides a natural foundation for this formulation because it can process inputs with different cardinalities without requiring a fixed ordering (Zaheer et al., 2017; Lee et al., 2019; Qi et al., 2017). Perceiver and Perceiver IO further demonstrate how variable-size inputs can be mapped into fixed-dimensional latent representations through input-agnostic architectures (Jaegle et al., 2021; 2022). Related ideas have also emerged in neural sensing systems with heterogeneous sensor configurations. Recent EEG foundation models, for example, explicitly consider variation in sensing setups and sensor topology (Döner et al., 2025; El Ouahidi et al., 2025; Xiao et al., 2025). These studies motivate treating sensor configuration as part of the representation problem rather than assuming that it is permanently fixed.

Based on this view, we propose SIS-NET, a compact sensor-set model for sensor-configuration-agnostic electronic-nose recognition. Each available sensor is processed by a shared temporal encoder and represented together with an explicit sensor identity. The resulting sensor representations form a variable-size set. A masked aggregation operation then maps the available sensor set to a fixed-dimensional representation for classification. During training, we expose the model to different sensor subsets so that it does not learn to rely exclusively on the complete sensor array. Importantly, the model is trained only once. The resulting checkpoint is subsequently evaluated under multiple sensor configurations without retraining.

We evaluate this property through a unified sensor-configuration protocol. The same frozen checkpoint is tested with different sensor availability levels, ordinary fixed subsets, random per-sample subsets, and combinations withheld from training. This protocol separates two effects that are often conflated: performance degradation caused simply by observing fewer sensors, and the ability of a trained model to generalize to previously unseen sensor combinations. We further compare SIS-Net with fixed-input and simple set-based baselines, and conduct controlled ablations to examine the roles of sensor identity and training-time sensor-subset sampling.

Our contributions are summarized as follows:

- We formulate **sensor-configuration-agnostic electronic-nose recognition**, where a single trained model operates under different numbers and combinations of available sensors.
- We introduce **SIS-Net**, which represents available sensors as identifiable sensor tokens and uses shared temporal encoding with masked set aggregation to obtain a fixed-dimensional representation from variable sensor subsets.
- We introduce a **training-time sensor-subset sampling strategy** that exposes the model to changing sensor configurations without requiring configuration-specific models.
- We establish a **unified sensor-configuration evaluation protocol** based on a single frozen checkpoint, covering different sensor availability levels, fixed subsets, random subsets, and held-out sensor combinations.
- We conduct controlled comparisons and ablations to quantify configuration robustness and the contributions of sensor identity and sensor-subset training.

2 Related Work

Electronic-nose recognition, sensor optimization, and fault tolerance. Learning-based electronic-nose systems have been developed for gas identification, chemical recognition, and odor-related classification (Fang et al., 2022; Li et al., 2024; Ozer et al., 2023). Recent work has also explored data-centric representations of sensor arrays and adaptive learning for practical chemical sensing (Sung et al., 2024; Shtepliuk et al., 2025). At the same time, the redundancy and cost of multisensor arrays have motivated sensor-array optimization. Deep learning methods have been used to identify effective sensor subsets, while interpretable selection strategies explicitly consider informative and non-redundant sensors (Zhang et al., 2024; Mei et al., 2025). Another line of work studies sensor failure tolerance. Algorithmic compensation, fault-tolerant learning, and elastic sensor architectures have been investigated to maintain recognition performance when individual sensors become unavailable (Fonollosa et al., 2013; Kaya et al., 2020; Wang et al., 2023). These studies address closely related practical constraints, but their primary objective is to select or maintain a suitable sensor array. Our setting is different: the sensor universe is fixed, while the available subset may change from sample to sample, and one trained model is required to operate across these configurations.

Variable-channel modeling for multivariate time series. Changing channel availability is also relevant to multivariate time-series learning. Conventional multivariate models generally assume a fixed channel dimension and often model interactions according to a predefined channel layout (Wu et al., 2023; Liu et al., 2024; Wang et al., 2024). Channel-selection methods instead aim to identify relevant and non-redundant variables, as in TSelect and related feature-selection approaches (Nuyts et al., 2025; Peng et al., 2005). More recent work investigates robustness to channel permutation and changing channel relationships. CPiRi, for example, explicitly studies channel-permutation-invariant relational interaction, while ChannelTokenFormer considers broader real-world issues involving dependency, asynchrony, and missingness (Xu et al., 2026; Jang et al., 2026). RAINDROP addresses missing and irregularly observed multivariate time series by modeling dependencies among sensors with a graph-based architecture (Zhang et al., 2022). These approaches demonstrate the importance of relaxing fixed-channel assumptions. However, they primarily target forecasting, irregular observation, or missing-data settings. Our focus is recognition under changing sensor subsets, where the identity and availability of physical sensors are part of the input configuration.

Set-based representation learning and variable sensor configurations. Set-based models provide a natural way to represent collections whose cardinality and ordering may vary. Deep Sets establishes permutation-invariant set representation through shared element-wise transformations and symmetric aggregation, while Set Transformer extends this framework with attention-based interactions among set elements (Zaheer et al., 2017; Lee et al., 2019). PointNet similarly demonstrates that unordered collections of elements can be mapped to useful global representations without relying on a fixed ordering (Qi et al., 2017). Perceiver and Perceiver IO provide another relevant perspective by separating input representation from a fixed-dimensional latent representation, allowing the architecture to process flexible input structures (Jaegle et al., 2021; 2022). Related ideas have recently appeared in neural sensing systems. LUNA, REVE, and BrainOmni explicitly consider variation in EEG sensing configurations and sensor topology (Döner et al., 2025; El Ouahidi et al., 2025; Xiao et al., 2025). These works motivate modeling sensors as identifiable elements rather than permanently fixed channel positions. However, the sensor-set formulation has not been systematically investigated for electronic-nose recognition under changing sensor availability. SIS-Net therefore combines shared sensor-wise temporal encoding, explicit sensor identity, and set aggregation for this setting.

3 Methodology

3.1 Problem Formulation

We consider electronic-nose recognition with a known sensor universe

$$\mathcal{S} = \{1, 2, \ldots, M\}, \tag{1}$$

where M denotes the total number of sensor identities. In our experiments, $M = 16$.

For an input sample i, only a subset of sensors may be available at inference time. We denote the available sensor set by

$$\mathcal{S}_i \subseteq \mathcal{S}, \tag{2}$$

and its cardinality by

$$K_i = |\mathcal{S}_i|. \tag{3}$$

Thus, K_i may vary across samples.

For each available sensor $j \in \mathcal{S}_i$, the observed response is a one-dimensional temporal sequence

$$x_{i,j} \in \mathbb{R}^T, \tag{4}$$

where T denotes the number of temporal observations. The input to the model is therefore a variable-size collection of sensor responses,

$$\mathcal{X}_i = \{x_{i,j} \mid j \in \mathcal{S}_i\}, \tag{5}$$

rather than a fixed $M \times T$ sensor matrix.

Our objective is to learn a single recognition model

$$f_\Theta : \mathcal{X}_i \to \hat{y}_i, \tag{6}$$

where $\hat{y}_i$ denotes the predicted class distribution. The same trained model should operate on different numbers and combinations of available sensors without configuration-specific retraining.

This formulation differs from conventional sensor selection, where a fixed sensor subset is first selected and subsequently used for all samples. Here, the sensor universe remains fixed, while the available subset may change across samples. The model must therefore represent both the temporal response of each sensor and the identity of the physical sensor that produced it.

3.2 SIS-Net Architecture

Figure 1 illustrates the overall architecture of SIS-Net. The model transforms a variable collection of sensor responses into a fixed-dimensional representation for classification. The processing pipeline consists of shared sensor-wise temporal encoding, sensor identity-aware tokenization, variable-size sensor-set construction, masked mean aggregation, and classification. For a sample with K_i available sensors, the overall transformation can be summarized as

$$\mathcal{X}_i \to \{h_{i,j}\}_{j\in\mathcal{S}_i} \to \{z_{i,j}\}_{j\in\mathcal{S}_i} \to \mathcal{Z}_i \to r_i \to \hat{y}_i, \tag{7}$$

where $h_{i,j}$ denotes the temporal representation of sensor j, $z_{i,j}$ denotes its sensor-identity-aware token, $\mathcal{Z}_i$ denotes the variable-size sensor token set, and r_i denotes the resulting fixed-dimensional sample representation.

The complete mapping can be written as

$$\hat{y}_i = C_\phi \left(\frac{1}{K_i} \sum_{j\in\mathcal{S}_i} [E_\theta(x_{i,j}) + e_j] \right), \tag{8}$$

where E_θ is the shared temporal encoder, e_j is the learnable embedding associated with sensor identity j, and C_ϕ is the classification head.

3.3 Sensor-wise Temporal Encoding

The temporal response of each available sensor is first encoded independently. Importantly, the same temporal encoder is shared across all sensor identities:

$$h_{i,j} = E_\theta(x_{i,j}), \qquad j \in \mathcal{S}_i, \tag{9}$$

where

$$h_{i,j} \in \mathbb{R}^d \tag{10}$$

is the sensor-level temporal representation and d is the feature dimension.

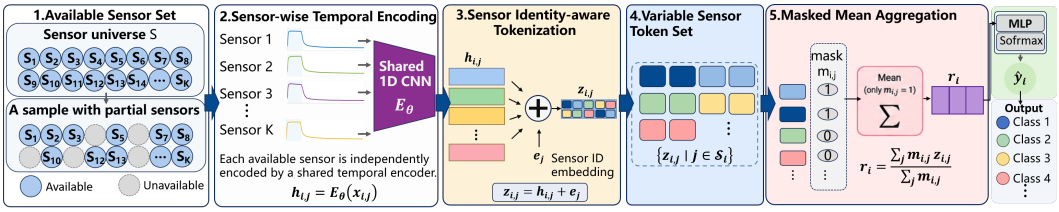


Figure 1: Overview of SIS-Net. Given a fixed sensor universe, each sample may contain a different subset of available sensors. SIS-Net independently encodes the temporal response of each available sensor, incorporates sensor identity into sensor tokens, represents the available sensors as a variable-size token set, and aggregates them into a fixed-dimensional representation for classification.

We implement E_θ as a three-layer one-dimensional convolutional neural network. The encoder parameters are shared across all available sensors. Therefore, each sensor response is processed by the same temporal feature extractor rather than by a sensor-specific encoder.

This shared design separates temporal response modeling from sensor identity. The encoder captures the temporal characteristics observed in a sensor response, while the identity information is introduced explicitly in the subsequent tokenization stage.

3.4 Sensor Identity-aware Tokenization

A temporal representation alone does not explicitly specify which physical sensor generated the response. This distinction is important because different sensors may exhibit different response characteristics even when their temporal representations are similar.

We therefore associate each sensor identity $j \in \mathcal{S}$ with a learnable embedding

$$e_j \in \mathbb{R}^d. \tag{11}$$

For an available sensor, its temporal representation and identity embedding are combined by

$$z_{i,j} = h_{i,j} + e_j, \qquad j \in \mathcal{S}_i. \tag{12}$$

The resulting $z_{i,j} \in \mathbb{R}^d$ is an identity-aware sensor token. Thus, each token contains two complementary types of information: the temporal response pattern extracted from the observed signal and the identity of the physical sensor that produced that signal.

3.5 Variable Sensor Token Set

The identity-aware sensor tokens are collected into a set

$$\mathcal{Z}_i = \{z_{i,j} \mid j \in \mathcal{S}_i\}. \tag{13}$$

The cardinality of this set is determined by the currently available sensors:

$$|\mathcal{Z}_i| = K_i. \tag{14}$$

Because K_i may vary across samples, $\mathcal{Z}_i$ is a variable-size set. In contrast, the feature dimension of every token remains fixed:

$$z_{i,j} \in \mathbb{R}^d. \tag{15}$$

This representation removes the requirement that the model receive a fixed number of sensor channels at its input. A sample with 16 available sensors therefore produces 16 sensor tokens, while a sample with 8 available sensors produces only 8 tokens. The model does not need to introduce artificial sensor responses for unavailable sensors.

The set representation also avoids making the number of available sensors part of the dimensionality of the downstream representation. Instead, the variable-size sensor set is converted into a fixed-dimensional representation by the aggregation operation described next.

3.6 MASKED MEAN AGGREGATION

Since the number of available sensors varies across samples, SIS-Net requires an aggregation operation whose output dimension is independent of K_i. We use masked mean aggregation over the available sensor tokens:

$$r_i = \frac{1}{K_i} \sum_{j \in \mathcal{S}_i} z_{i,j}, \tag{16}$$

where

$$r_i \in \mathbb{R}^d \tag{17}$$

is the resulting sample-level representation.

Equivalently, let

$$m_{i,j} \in \{0, 1\} \tag{18}$$

denote the availability mask, where $m_{i,j} = 1$ if sensor j is available for sample i and $m_{i,j} = 0$ otherwise. The aggregation can then be written as

$$r_i = \frac{\sum_{j=1}^{M} m_{i,j} z_{i,j}}{\sum_{j=1}^{M} m_{i,j}}. \tag{19}$$

Only available sensor tokens contribute to the aggregation. An unavailable sensor therefore contributes neither a temporal representation nor an artificial response value. At the same time, the dimensionality of r_i remains fixed regardless of the number of available sensors.

The resulting representation is passed to a classification head:

$$\hat{y}_i = C_\phi(r_i), \tag{20}$$

where C_ϕ denotes the classifier and $\hat{y}_i$ denotes the predicted class distribution.

Combining the preceding operations gives the complete SIS-Net mapping:

$$\hat{y}_i = C_\phi \left(\frac{1}{K_i} \sum_{j \in \mathcal{S}_i} \left[E_\theta(x_{i,j}) + e_j \right] \right). \tag{21}$$

The architecture therefore remains unchanged when the number or combination of available sensors varies.

3.7 TRAINING WITH SENSOR-SUBSET SAMPLING

A variable-size input architecture alone does not guarantee robustness to changing sensor configurations. If the model were trained only with the complete sensor array, it could still learn representations that implicitly depend on the full configuration.

We therefore introduce sensor-subset sampling during training. For each training sample, an availability level is first sampled from

$$K_i \in \{16, 12, 8, 4\}. \tag{22}$$

A corresponding sensor subset $\mathcal{S}_i$ is then sampled from the known sensor universe according to the predefined configuration protocol. The model receives only the selected sensor responses and their corresponding sensor identities.

Training can therefore be formulated as minimizing the expected classification loss over both the data distribution and the sensor-subset sampling process:

$$\min_{\theta, \phi, \{e_j\}} \mathbb{E}_{(\mathcal{X}_i, y_i)} \mathbb{E}_{\mathcal{S}_i} \left[\mathcal{L} \left(C_\phi(r_i), y_i \right) \right], \tag{23}$$

where $\mathcal{L}$ denotes the classification loss.

The purpose of subset sampling is to expose the model to configuration variation during training rather than only at inference time. After training, the resulting checkpoint is frozen and evaluated under multiple sensor configurations without configuration-specific retraining or test-time fine-tuning.

Thus, SIS-Net combines two complementary mechanisms. The sensor-set architecture makes variable sensor configurations structurally compatible with a single model, while sensor-subset sampling exposes the model to such configuration variation during training.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Dataset and task. We evaluate sensor-configuration-agnostic electronic-nose recognition on a dynamic gas-mixture dataset collected with a 16-sensor chemical sensor array under continuously varying gas concentrations (Fonollosa et al., 2015). We use the $\mathrm{ethylene_CO}$ measurement from the original dataset. The processed data contain 975 model samples, each represented as a multivariate time series with shape $[16, 300]$. The 16 sensor channels form a fixed sensor universe, while the processed task contains four concentration-based classification classes. Each sample therefore provides a complete reference configuration of 16 sensors before sensor-availability variation is introduced.

Rather than assuming that all sensors are always available, we evaluate the same recognition model under four sensor-availability levels: 16, 12, 8, and 4 sensors. The experimental scope is restricted to this known sensor universe. We do not consider unseen sensor identities or zero-shot recognition of new sensor types. Sensor-configuration variation is introduced by masking subsets of the available sensor channels during training and evaluation; the underlying sensor responses are not synthetically generated.

Training and evaluation protocol. All formal results are obtained using five group-based folds, where model samples originating from the same stable concentration segment are kept within the same fold. Within each fold, the checkpoint is selected according to validation Macro-F1 under the Full16 condition. The selected checkpoint is then frozen and evaluated under all sensor configurations, without configuration-specific retraining or test-time fine-tuning. Thus, comparisons across sensor configurations measure the ability of a single trained model to operate under changing sensor availability rather than the ability of independently optimized models to solve each configuration.

During training, the number of available sensors is uniformly sampled from $\{16, 12, 8, 4\}$, followed by sampling a legal sensor combination at the selected sensor count. The proposed SIS-Net uses a shared three-layer 1D CNN to encode each sensor time series, augments the resulting sensor-wise feature with a learnable sensor-identity embedding, and aggregates the resulting variable-size sensor token set before classification. The baselines and the proposed model are evaluated using the same sensor-configuration protocol.

Sensor configuration protocol. We consider ten evaluation conditions: Full16, Fixed12, Random12, HeldOut12, Fixed8, Random8, HeldOut8, Fixed4, Random4, and HeldOut4. In a Fixed-k condition, all samples are evaluated using the same k-sensor subset. In a Random-k condition, each sample is assigned a deterministic sensor subset generated from a configuration seed. In a HeldOut-k condition, the complete k-sensor combination is excluded from the training subset sampler, although all individual sensor identities in that combination remain known during training. This protocol separates ordinary variation in sensor combinations from generalization to combinations that were not jointly observed during training.

Baselines and metrics. We compare SIS-Net with a Full-Sensor Reference, Zero-Fill Fixed Input, Sensor Swarm Ensemble, Multi-Kernel Fault-Tolerant E-nose, and RAINDROP. The Full-Sensor Reference is evaluated only under the Full16 condition and represents the conventional fixed-array setting. Zero-Fill Fixed Input represents a fixed-input strategy in which unavailable standardized sensor channels are filled with zeros. Sensor Swarm Ensemble and Multi-Kernel Fault-Tolerant E-nose represent sensor-level and predefined sensor-subset fault-tolerant strategies, respectively. RAINDROP provides a multivariate time-series baseline with observation-aware sensor-relation modeling.

Macro-F1 is the primary evaluation metric because it gives equal weight to the four classes. Accuracy is used as a secondary metric but is omitted from the main tables for compactness. Unless otherwise stated, results in the main comparison and component ablation are reported as mean $\pm$ standard deviation across the five folds.

Table 1: Macro-F1 under the unified sensor-configuration protocol. Results are mean ± standard deviation over five folds. "Fixed/Full" denotes Full16 for the 16-sensor setting and Fixed-k for reduced sensor settings. R and H denote Random and HeldOut configurations, respectively. Bold indicates the best result among the variable-configuration methods within each sensor-availability level.

#Sensors	Method	Fixed / Full	Random	HeldOut
16	Full-Sensor Reference	$\mathbf{0.8482 \pm 0.0149}$	–	–
	Zero-Fill Fixed Input	0.8061 ± 0.0418	–	–
	Sensor Swarm Ensemble	0.8360 ± 0.0159	–	–
	Multi-Kernel Fault-Tolerant E-nose	0.8418 ± 0.0125	–	–
	RAINDROP	0.8459 ± 0.0378	–	–
	SIS-Net (Ours)	0.8397 ± 0.0245	–	–
12	Zero-Fill Fixed Input	0.7118 ± 0.0253	0.6713 ± 0.0143	0.6882 ± 0.0148
	Sensor Swarm Ensemble	0.7917 ± 0.0337	0.7737 ± 0.0378	0.7546 ± 0.0459
	Multi-Kernel Fault-Tolerant E-nose	0.8067 ± 0.0196	0.7833 ± 0.0113	0.7701 ± 0.0216
	RAINDROP	0.8142 ± 0.0145	0.7924 ± 0.0547	$\mathbf{0.7809 \pm 0.0417}$
	SIS-Net (Ours)	$\mathbf{0.8698 \pm 0.0348}$	$\mathbf{0.8263 \pm 0.0531}$	0.7793 ± 0.0429
8	Zero-Fill Fixed Input	0.6657 ± 0.0166	0.5976 ± 0.0053	0.6716 ± 0.0563
	Sensor Swarm Ensemble	0.7264 ± 0.0209	0.6926 ± 0.0251	0.6830 ± 0.0123
	Multi-Kernel Fault-Tolerant E-nose	0.7441 ± 0.0435	0.7295 ± 0.0118	0.7023 ± 0.0312
	RAINDROP	0.7679 ± 0.0574	0.7412 ± 0.0452	0.7312 ± 0.0022
	SIS-Net (Ours)	$\mathbf{0.7734 \pm 0.0143}$	$\mathbf{0.8048 \pm 0.0688}$	$\mathbf{0.7775 \pm 0.0127}$
4	Zero-Fill Fixed Input	0.6213 ± 0.0468	0.5556 ± 0.0357	0.3026 ± 0.0235
	Sensor Swarm Ensemble	0.6467 ± 0.0466	0.6133 ± 0.0254	0.5916 ± 0.0245
	Multi-Kernel Fault-Tolerant E-nose	0.6648 ± 0.0128	0.6433 ± 0.0253	0.6212 ± 0.0114
	RAINDROP	0.6905 ± 0.0346	0.6777 ± 0.0471	0.6509 ± 0.0112
	SIS-Net (Ours)	$\mathbf{0.7892 \pm 0.0348}$	$\mathbf{0.7565 \pm 0.0407}$	$\mathbf{0.7108 \pm 0.0292}$

4.2 Results and Analysis

Main configuration generalization. Table 1 compares all methods under the unified sensor-configuration protocol. The Full-Sensor Reference achieves the highest Full16 Macro-F1 of 0.8482 ± 0.0149, while SIS-Net obtains a competitive 0.8397 ± 0.0245. More importantly, SIS-Net is designed to use the same trained model across all variable-sensor conditions, whereas the Full-Sensor Reference is evaluated only with the complete sensor array.

As sensor availability decreases, SIS-Net maintains relatively stable recognition performance. Its Macro-F1 changes from 0.8397 ± 0.0245 under Full16 to 0.8698 ± 0.0348, 0.7734 ± 0.0143, and 0.7892 ± 0.0348 under Fixed12, Fixed8, and Fixed4, respectively. The corresponding random-per-sample configurations obtain 0.8263 ± 0.0531, 0.8048 ± 0.0688, and 0.7565 ± 0.0407. SIS-Net also remains effective on held-out known combinations, achieving 0.7793 ± 0.0429, 0.7775 ± 0.0127, and 0.7108 ± 0.0292 for HeldOut12, HeldOut8, and HeldOut4, respectively.

In contrast, Zero-Fill Fixed Input shows a substantially stronger degradation as sensors become unavailable. Its Macro-F1 decreases from 0.8061 ± 0.0418 under Full16 to 0.6213 ± 0.0468 under Fixed4, 0.5556 ± 0.0357 under Random4, and 0.3026 ± 0.0235 under HeldOut4. The results indicate that explicitly modeling the available sensors as a variable set is more suitable for configuration changes than retaining a fixed input structure and filling unavailable channels.

Among the variable-configuration baselines, RAINDROP remains competitive, particularly under several fixed and held-out settings. SIS-Net does not dominate every individual condition; for example, RAINDROP obtains a slightly higher Macro-F1 on HeldOut12. Nevertheless, SIS-Net provides consistently strong performance across fixed, random, and held-out configurations while using one checkpoint throughout the evaluation.

Configuration robustness. To determine whether the observed performance depends on one particular sensor combination, we perform an inference-only configuration robustness analysis using frozen checkpoints. For each sensor count, Fixed-k results are averaged over 10 ordinary fixed sensor combinations, while Random-k results are averaged over five configuration seeds. These combinations are ordinary evaluation configurations and are not treated as additional held-out combinations.

Table 2: Configuration robustness evaluated with frozen checkpoints. Fixed results are mean ± standard deviation over 10 ordinary fixed sensor combinations, while Random results are mean ± standard deviation over five configuration seeds. These configurations are not additional held-out combinations.

#Sensors	Method	Fixed	Random
12	Zero-Fill Fixed Input	0.7551 ± 0.0317	0.6624 ± 0.0178
	SIS-Net (Ours)	$\mathbf{0.8172 \pm 0.0139}$	$\mathbf{0.8075 \pm 0.0209}$
8	Zero-Fill Fixed Input	0.6334 ± 0.0122	0.5841 ± 0.0309
	SIS-Net (Ours)	$\mathbf{0.8003 \pm 0.0132}$	$\mathbf{0.7933 \pm 0.0197}$
4	Zero-Fill Fixed Input	0.5503 ± 0.0375	0.4801 ± 0.0459
	SIS-Net (Ours)	$\mathbf{0.7482 \pm 0.0130}$	$\mathbf{0.7414 \pm 0.0147}$

Table 3: Component ablation results. Values are mean ± standard deviation over five folds. Avg-k is computed within each fold as the mean of Fixed-k, Random-k, and HeldOut-k, followed by aggregation across folds.

Variant	Full16	Avg-12	Avg-8	Avg-4
SIS-Net	$\mathbf{0.8397 \pm 0.0245}$	$\mathbf{0.8251 \pm 0.0284}$	$\mathbf{0.7852 \pm 0.0262}$	$\mathbf{0.7522 \pm 0.0168}$
SIS-Net w/o Sensor ID	0.8333 ± 0.0346	0.7973 ± 0.0254	0.7216 ± 0.0142	0.6546 ± 0.0146
SIS-Net w/o Subset Training	0.8626 ± 0.0137	0.7076 ± 0.0575	0.5542 ± 0.0501	0.4001 ± 0.0367

Table 2 shows that SIS-Net maintains higher average Macro-F1 than Zero-Fill Fixed Input across all evaluated sensor counts and both configuration types. The advantage becomes particularly clear when only four sensors are available: SIS-Net achieves 0.7482 ± 0.0130 and 0.7414 ± 0.0147 on Fixed4 and Random4, compared with 0.5503 ± 0.0375 and 0.4801 ± 0.0459 for Zero-Fill Fixed Input. The smaller configuration-level standard deviations of SIS-Net in these settings also indicate that its performance is less sensitive to the particular sensor subset used for evaluation.

Component ablation. We next examine the contributions of sensor identity encoding and sensor-subset training. Table 3 reports five-fold results for the complete SIS-Net and two variants. For each fold, Avg-k is first computed as the mean of Fixed-k, Random-k, and HeldOut-k, after which the resulting fold-level values are aggregated across the five folds.

Removing the Sensor ID embedding has a limited effect under Full16, with Macro-F1 changing from 0.8397 ± 0.0245 to 0.8333 ± 0.0346. However, its effect becomes more pronounced as sensor availability decreases: Avg-8 decreases from 0.7852 ± 0.0262 to 0.7216 ± 0.0142, while Avg-4 decreases from 0.7522 ± 0.0168 to 0.6546 ± 0.0146. This pattern suggests that explicitly encoding the identity of the available sensors becomes more useful when the model must operate with fewer sensor observations.

The effect of subset training is substantially larger. Without subset training, SIS-Net achieves 0.8626 ± 0.0137 under Full16, but its performance drops to 0.7076 ± 0.0575, 0.5542 ± 0.0501, and 0.4001 ± 0.0367 for Avg-12, Avg-8, and Avg-4, respectively. Thus, exposure to variable sensor subsets during training is important for transferring a model trained on the complete sensor universe to reduced and changing sensor configurations.

5 Conclusion

We presented SIS-Net, a sensor-set model for electronic-nose recognition under changing sensor availability. SIS-Net combines shared sensor-wise temporal encoding, explicit sensor identity, variable-size sensor token sets, masked mean aggregation, and training-time sensor-subset sampling. This design enables a single trained model to operate under different numbers and combinations of available sensors without configuration-specific retraining. Our experiments demonstrate its behavior under varying sensor availability and sensor configurations, and verify the contributions of the main components. Future work will extend the study to additional electronic-nose datasets and more diverse sensing conditions.

AI USE STATEMENT

The authors used generative AI tools (specifically ChatGPT by OpenAI) to assist with manuscript writing, editing, and figure design. The authors reviewed and verified all AI-assisted content and take full responsibility for the final manuscript.

REFERENCES

Berkay Döner, Thorir Mar Ingolfsson, Luca Benini, and Yawei Li. Luna: Efficient and topology-agnostic foundation model for eeg signal analysis. In *Advances in Neural Information Processing Systems*, volume 38, pp. 78772–78798, 2025.

Yassine El Ouahidi, Jonathan Lys, Philipp Thölke, Nicolas Farrugia, Bastien Pasdeloup, Vincent Gripon, Karim Jerbi, and Giulia Lioi. Reve: A foundation model for eeg – adapting to any setup with large-scale pretraining on 25,000 subjects. 38:25583–25619, 2025.

Cong Fang, Hua-Yao Li, Long Li, Hu-Yin Su, Jiang Tang, Xiang Bai, and Huan Liu. Smart electronic nose enabled by an all-feature olfactory algorithm. *Advanced Intelligent Systems*, 4(7): 2200074, 2022.

Jordi Fonollosa, Alexander Vergara, and Ramón Huerta. Algorithmic mitigation of sensor failure: Is sensor replacement really necessary? *Sensors and Actuators B: Chemical*, 183:211–221, 2013.

Jordi Fonollosa, Sadique Sheik, Ramón Huerta, and Santiago Marco. Reservoir computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring. *Sensors and Actuators B: Chemical*, 215:618–629, 2015.

Andrew Jaegle, Felix Gimeno, Andy Brock, Oriol Vinyals, Andrew Zisserman, and Joao Carreira. Perceiver: General perception with iterative attention. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 4651–4664. PMLR, 2021.

Andrew Jaegle, Sebastian Borgeaud, Jean-Baptiste Alayrac, Carl Doersch, Catalin Ionescu, David Ding, Skanda Koppula, Daniel Zoran, Andrew Brock, Evan Shelhamer, Olivier Henaff, Matthew M. Botvinick, Andrew Zisserman, Oriol Vinyals, and Joao Carreira. Perceiver io: A general architecture for structured inputs & outputs. In *International Conference on Learning Representations*, 2022.

Jinkwan Jang, Hyungjin Park, Jinmyeong Choi, and Taesup Kim. Towards robust real-world multivariate time series forecasting: A unified framework for dependency, asynchrony, and missingness. In *International Conference on Learning Representations*, 2026.

Aydin Kaya, Ali Seydi Keçeli, Cagatay Catal, and Bedir Tekinerdogan. Sensor failure tolerable machine learning-based food quality prediction model. *Sensors*, 20(11):3173, 2020.

Juho Lee, Yoonho Lee, Jungtaek Kim, Adam R. Kosiorek, Seungjin Choi, and Yee Whye Teh. Set transformer: A framework for attention-based permutation-invariant neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pp. 3744–3753, 2019.

Ziyang Li, Siyuan Kang, Ninghui Feng, Chongbo Yin, and Yan Shi. Pscformer: A lightweight hybrid network for gas identification in electronic nose system. *Pattern Recognition*, 145:109912, 2024.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *International Conference on Learning Representations*, 2024.

Haixia Mei, Jingyi Peng, Tao Wang, Bowei Zhang, and Fuzhen Xuan. An interpretable sensor selection strategy for array optimization and performance enhancement. *ACS Sensors*, 2025.

Loren Nuyts, Lorenzo Perini, and Jesse Davis. Tselect: Selecting relevant and non-redundant channels for multivariate time series classification. *Data Mining and Knowledge Discovery*, 2025.

Emre Ozer, Jedrzej Kufel, John Biggs, Anjit Rana, Francisco J. Rodriguez, Thomas Lee-Clark, Antony Sou, Catherine Ramsdale, Scott White, Suresh Kumar Garlapati, et al. Malodour classification with low-cost flexible electronics. *Nature Communications*, 14(1):777, 2023.

Hanchuan Peng, Fuhui Long, and Chris Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005.

Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 652–660, 2017.

Ivan Shtepliuk, Kerstin Montelius, Jens Eriksson, and Donatella Puglisi. Adaptive machine learning for electronic nose-based forensic voc classification. *Advanced Science*, 12(36):e04657, 2025.

Seung-Hyun Sung, Jun Min Suh, Yun Ji Hwang, Ho Won Jang, Jeon Gue Park, and Seong Chan Jun. Data-centric artificial olfactory system based on the eigengraph. *Nature Communications*, 15(1):1211, 2024.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y. Zhang, and Jun Zhou. Timemixer: Decomposable multiscale mixing for time series forecasting. In *International Conference on Learning Representations*, 2024.

Tao Wang, Yu Wu, Yongwei Zhang, Wen Lv, Xiyu Chen, Min Zeng, Jianhua Yang, Yanjie Su, Nantao Hu, and Zhi Yang. Portable electronic nose system with elastic architecture and fault tolerance based on edge computing, ensemble learning, and sensor swarm. *Sensors and Actuators B: Chemical*, 375:132925, 2023.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *International Conference on Learning Representations*, 2023.

Qinfan Xiao, Ziyun Cui, Chi Zhang, SiQi Chen, Wen Wu, Andrew Thwaites, Alexandra Woolgar, Bowen Zhou, and Chao Zhang. Brainomni: A brain foundation model for unified eeg and meg signals. In *Advances in Neural Information Processing Systems*, volume 38, pp. 46081–46114, 2025.

Jiyuan Xu, Wenyu Zhang, Xin Jing, Jiahao Nie, Shuai Chen, and Shuai Zhang. Cpiri: Channel permutation-invariant relational interaction for multivariate time series forecasting. In *International Conference on Learning Representations*, 2026.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan Salakhutdinov, and Alexander J. Smola. Deep sets. In *Advances in Neural Information Processing Systems*, volume 30, pp. 3391–3401, 2017.

Xiang Zhang, Marko Zeman, Theodoros Tsiligkaridis, and Marinka Zitnik. Graph-guided network for irregularly sampled multivariate time series. In *International Conference on Learning Representations*, 2022.

Xijia Zhang, Tao Wang, Wangze Ni, Yongwei Zhang, Wen Lv, Min Zeng, Jianhua Yang, Nantao Hu, Rui Zhan, Guang Li, et al. Sensor array optimization for the electronic nose via different deep learning methods. *Sensors and Actuators B: Chemical*, 410:135579, 2024.