

A Communication Aware Masked Denoising Autoencoder Framework for Reliability Aware and Fault Tolerant Gas Classification in Electronic Nose Systems

Saifur Rahman^{1,2,*}

¹ Electrical Engineering Department, College of Engineering, Najran University, Najran, Saudi Arabia

² Scientific and Engineering Research Centre, Najran University, Najran, Saudi Arabia

INFORMATION

Keywords:

Electronic nose
fault tolerant communication
systems
masked denoising autoencoder
gas classification
sensor reliability

DOI: 10.23967/j.rimni.2026.10.84609

Revista Internacional
Métodos numéricos
para cálculo y diseño en ingeniería

RIMNI



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

In cooperation with

CIMNE[®]

A Communication Aware Masked Denoising Autoencoder Framework for Reliability Aware and Fault Tolerant Gas Classification in Electronic Nose Systems

Saifur Rahman^{1,2,*}

¹Electrical Engineering Department, College of Engineering, Najran University, Najran, Saudi Arabia

²Scientific and Engineering Research Centre, Najran University, Najran, Saudi Arabia

ABSTRACT

Electronic noses (e-noses) have become a promising technology for gas classification in environmental and industrial safety and smart sensing. However, sensor noise, redundancy, drift, and partial sensor failures usually impair performance and reduce reliability in practice. Communication-wise, the multi-sensor array could be viewed as a distributed sensing and information transmission system, noise, sensor failures, and drift are considered channel impairments. To overcome these issues, this paper presents a reliability-conscious, fault-tolerant e-nose model built around a masked denoising autoencoder (MDAE) with probabilistic classification. The approach is based on a masking process that causes the model to learn strong latent representations that can be used to recreate lost or corrupt sensor signals. These acquired representations are then used to perform correct multi-class classification. The framework was evaluated using an MQ gas sensor dataset containing 40,706 samples. After exact duplicate removal, 28,441 unique samples from five valid gas classes were retained for leakage-controlled evaluation. The proposed MDAE + calibrated Fusion SVM achieved 100.00% accuracy, macro-F1 of 1.000, macro-AUC of 1.000, Brier score of 0.000015, NLL of 0.002172, and ECE of 0.002167. These results confirm excellent classification and probability calibration. While some calibrated baselines performed slightly better on Brier, the proposed framework successfully integrates denoising reconstruction, latent feature learning, sensor fault restoration, and sensor dropout and noise robustness in a novel manner. The efficacy of the masking strategy and denoising architecture is shown through ablation studies, and the stability and generalizability of the proposed framework are verified through 5-fold cross-validation. Furthermore, the robustness analysis shows that the proposed framework performs very well even in the presence of sensor drops and maintains around 72%–74% accuracy even in severe multi-sensor drops. It exhibits graceful degradation under increasing Gaussian noise, corresponding to a resilient communication system over increasingly degraded channels. The importance of features based on permutation provides a key insight into the importance of particular sensors and justifies the model's learning ability in terms of redundancy awareness. Overall, the proposed framework achieves high classification accuracy and demonstrates robustness, reliability, interpretability, and fault tolerance, and can be applied to intelligent sensing and communication-aware systems in practical applications.

OPEN ACCESS

Received: 26/04/2026

Accepted: 20/07/2026

Published: 21/09/2026

DOI

10.23967/j.rimni.2026.10.84609

Keywords:

Electronic nose
fault tolerant communication
systems
masked denoising autoencoder
gas classification
sensor reliability

1 Introduction

In the field of gas detection and environmental studies, electronic noses (e-noses) have become a valuable tool because they can mimic the human sense of smell by employing sensor arrays and pattern recognition algorithms. The use of an Electronic nose device for the classification of bacteria was found to be feasible in initial research by Dutta et al. [1], which serves as a basis for intelligent sensing applications. Over the years, an e-nose has evolved from its original concept into an advanced technology, with improvements in sensor design, data acquisition, and pattern recognition. The use of electronic noses in various applications, such as food quality, environmental monitoring, and industrial safety, was discussed in an in-depth overview by Wilson and Baietto [2]. Machine learning algorithms, including support vector machines, have been extensively applied to solve gas identification problems; e.g., Zhang et al. [3] designed a hybrid SVM framework to detect indoor air contaminants.

As more and more sensing environments are integrated with the IoT, current gas sensing platforms have evolved into multivariable, large-scale sensor arrays. Potyrailo [4] highlighted the importance of multivariable sensors in enabling ubiquitous gas monitoring in industrial and smart settings. On the same note, sophisticated sensing methods, such as optoelectronic sensor arrays, have been proposed to improve sensitivity and selectivity, as described by Li et al. [5].

Recent research has also increased the uses of e-nose systems. For example, Sanislav et al. [6] recently conducted an extensive overview of e-nose applications in food quality and safety and noted their practical relevance. Moreover, Yu et al. [7] discussed the shift toward sensor arrays for multivariate chemical analysis. More sophisticated hardware deployments, including biomimetic olfactory chips, have also been suggested to enhance sensing abilities, as shown by Wang et al. [8]. Additionally, Sun et al. [9] examined applications for detecting adulterants in meat, whereas Tasmurzayev et al. [10] examined applications of IoT-based air quality sensors.

Electronic nose systems have also been successfully implemented in industrial and environmental monitoring situations. For example, Han et al. [11] developed a gas-sensing system to monitor volatile compounds in tea processing. Likewise, Wang et al. [12] reported that e-nose technology was used to monitor wastewater treatment plants. Nevertheless, even with these developments, it is difficult to apply e-nose systems in practice due to sensor limitations.

Sensor drift is one of the greatest problems in e-nose systems, leading to deterioration in model performance over time. Hu et al. [13] explored gas recognition in the presence of sensor drift using deep learning, demonstrating its greater adaptability than conventional approaches. Smoothing and feature transformation are also suggested as signal preprocessing techniques that can enhance data quality, as demonstrated by Wang et al. [14]. Furthermore, sensor-based detection systems have also been considered in agriculture, which are discussed by Fundurulic et al. [15].

Multi-sensor data fusion techniques have been proposed to increase a system's robustness. A proposal for the adaptive weighting of a multi-sensor fusion strategy for condition monitoring applications was presented by Zhao et al. [16]. Nonetheless, they generally assume a consistent sensor input and do not explicitly consider sensor failures or missing data.

With the advent of deep learning, the ability of e-nose systems to learn complex patterns directly from raw sensor data has greatly improved. Goodfellow et al. [17] developed the basic principles of deep learning architectures that enable automatic feature extraction and improved generalization. Besides prediction accuracy, recent studies have aimed to enhance model reliability through calibration and uncertainty estimation.

The neural networks have been investigated by Guo et al. [18], who showed that contemporary deep models tend to be overconfident and need after-processing methods. Deep learning strategies, such as Bayesian deep learning, have been proposed to estimate predictive uncertainty, as described by Kendall and Gal [19]. The concept of dropout-based uncertainty estimation was proposed by Gal and Ghahramani [20] as an approximate method to Bayesian inference. On the same note, deep ensemble techniques have been shown to enhance predictive uncertainty estimation, as observed by Lakshminarayanan et al. [21]. Abdar et al. [22] conducted a thorough review of methods for quantifying uncertainty.

The selective prediction, also called classification with a reject option, was first proposed by Chow [23] and has been studied extensively in the pattern recognition literature. It's designed to improve decision-making reliability by enabling a model to abstain from making predictions when it's unsure. More recently, Geifman and El-Yaniv [24] applied the same concept to deep neural networks by developing a method for selective classification and later proposed SelectiveNet [25] for end-to-end learning of prediction and selection functions. These methods have significantly advanced the deployment of trustworthy AI to critical safety applications by reducing the risk of incorrect predictions.

Although electronic nose systems and deep learning-based classification have made remarkable progress, there are still several limitations:

- Most current researchers focus on classification accuracy but have little fault tolerance or robustness. Sensor drift and noise are considered to some extent, but sensor dropout and failure conditions are still mainly ignored.
- Current deep learning-based approaches do not explicitly include reconstruction processes for distorted sensor data.
- Reliability methods focus on estimating uncertainty, and robustness is not considered alongside fault restoration.

Although significant progress has been made in electronic nose systems, most prior work has focused on increasing classification accuracy under normal operating conditions. Few efforts have been made to simultaneously recover sensor corruption, fault tolerance, and prediction reliability in a joint learning framework. Furthermore, from a communication point of view, multi-sensor arrays are still underexploited, where sensor measurements can be considered as information streams corrupted by channel-like impairments such as sensor outages, noise, and drift. Inspired by these limitations, this work proposes a communication-aware masked denoising autoencoder framework capable of restoring sensor signals, robustly learning representations, and reliably characterizing the gas.

From a communication systems perspective, the electronic nose can be viewed as a distributed sensing and information transmission system. Every sensor is an information source that produces a stream of measurements which characterize the gas environment. Like communication channels, sensors can suffer from problems such as noise, drift, missing data, and hardware failures, which can make the information sent inaccurate and affect decision-making. Thus, its key challenge is to improve the reliability of the sensor information transmitted and recovered. Inspired by this view, the proposed framework introduces principles of communication-aware learning and treats sensor degradations as channel impairments, thereby facilitating the learning of apprising fault-resilient latent representations that can be reconstructed from degraded sensor information via masked denoising autoencoding.

In this paper, a new communication-aware masked denoising autoencoder (MDAE) framework is proposed to address these challenges for reliability-aware, fault-tolerant gas classification in electronic nose (ENoS) systems. The proposed method combines sensor restoration and latent representation learning within a unified MDAE framework, followed by a calibrated linear SVM for probabilistic classification. The major technical achievements of this work are outlined as follows:

- A communication-aware masked denoising autoencoder (MDAE) framework is developed in which sensor noise, drift, missing measurements, and partial sensor failures are modeled as communication-channel impairments. Through stochastic masking and denoising reconstruction, the framework learns fault-resilient latent representations that can recover corrupted or missing sensor information.
- A fault-restorative learning mechanism is introduced that combines reconstruction-based signal recovery with latent-feature-driven gas classification. The proposed framework explicitly improves fault tolerance and reliability under sensor degradation conditions, whereas traditional autoencoder-based methods aim only at feature compression.
- The robustness and reliability of the algorithm are assessed comprehensively through ablation studies, 5-fold cross-validation, sensor dropout analysis, noise injection analysis, computational complexity analysis, and probabilistic calibration analysis. The results show that the proposed method can provide highly reliable, well-calibrated classification and is robust even in poor sensing environments.

2 Materials and Methods

2.1 Dataset Description

The proposed approach is tested on a publicly available MQ sensor dataset comprising 40,706 samples from six gas sensors (MQ-2, MQ-5, MQ-6, MQ-8, MQ-9, and MQ-135). Each sample is a multivariate sensor response vector that is associated with varying gas exposure conditions. Before model training, exact duplicate feature-label rows were removed to avoid possible data leakage. This reduced the dataset from 40,706 samples to 28,442 unique samples. After de-duplication, Class 3 retained only one unique sample and was therefore excluded because it could not support stratified train/validation/test splitting or cross-validation. The final leakage-controlled dataset contained 28,441 samples from five gas classes.

The dataset can be mathematically represented as in [Eq. \(1\)](#)

$$\mathbf{X} \in \mathbb{R}^{N \times d} \quad (1)$$

where:

- N is the number of samples
- d is the number of sensors (features)

The respective class labels are defined as given in [Eq. \(2\)](#):

$$\mathbf{y} \in \{1, 2, \dots, C\}^N \quad (2)$$

where C represents the number of gas classes.

The resulting leakage-controlled dataset exhibits a five-class distribution that is imbalanced but sufficiently represented, suitable for supervised learning and statistical validation. [Table 1](#) presents the summary of the dataset.

Table 1: Dataset summary

Property	Value
Original samples	40,706
Unique samples after de-duplication	28,442
Final samples used	28,441
Removed duplicate rows	12,264
Final number of classes	5
Sensors used	MQ-2, MQ-5, MQ-6, MQ-8, MQ-9, MQ-135

To ensure an unbiased evaluation, all subsequent train/validation/test partitions and cross-validation experiments were performed exclusively on the de-duplicated dataset, thereby eliminating exact feature-label overlap between data partitions.

2.2 Data Preprocessing

To improve numerical stability, accelerate network convergence, and ensure consistent feature scaling, z-score normalization is applied to each sensor channel before model training, as defined in (3). To prevent information leakage, the normalization parameters are estimated solely from the training set and subsequently applied to the validation and test sets without modification.

$$\tilde{x}_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad (3)$$

where:

- μ_j and σ_j denote the mean and standard deviation of the j^{th} sensor, respectively.

This normalization eliminates scale differences among sensor channels, improves optimization stability, and prevents features with larger numerical ranges from dominating the learning process.

2.3 Masked Denoising Strategy

To simulate realistic sensor failures and measurement corruption, a stochastic masking mechanism is introduced before feature learning. A binary mask matrix is determined as given in Eq. (4):

$$\mathbf{M} \sim \text{Bernoulli}(1 - p) \quad (4)$$

where p is the masking probability; in all experiments, the masking probability was fixed at $p = 0.20$, indicating that 20% of sensor measurements were randomly masked during training. The masking operation was applied together with Gaussian noise injection and random scaling perturbations to emulate realistic sensor degradation conditions. Validation and testing were performed without intentionally masking sensor channels, thereby ensuring an unbiased assessment of reconstruction and classification performance.

The corrupted input is then generated as given in Eq. (5):

$$\mathbf{X}_{\text{mask}} = \mathbf{M} \odot \mathbf{X} \quad (5)$$

where $\odot$ denotes element-wise multiplication.

The formulation also enables the model to learn even when there is sensor dropout and partial data loss, as is common in actual e-nose systems.

2.4 Masked Denoising Autoencoder (MDAE)

The main element of the suggested framework is a denoising autoencoder, trained to restore the original input from corrupted observations.

The encoder projects the corrupted input to a latent representation as shown in Eq. (6):

$$z = f_{\theta}(X_{\text{mask}}) = \sigma(W_e X_{\text{mask}} + b_e) \quad (6)$$

The decoder recreates the original signal as given in Eq. (7):

$$\hat{X} = g_{\phi}(z) = \sigma(W_d z + b_d) \quad (7)$$

where:

- z is the latent representation
- $\hat{X}$ is the reconstructed output

The corrupted inputs are masked sensor measurements, additive Gaussian noise, and multiplicative random scaling perturbations, so that the network learns robust representations under various conditions of sensor degradation during training. The objective for the reconstruction is formulated in terms of Mean Squared Error (MSE) as presented in Eq. (8):

$$\mathcal{L}_{\text{rec}} = \frac{1}{N} \sum_{i=1}^N \|x_i - \hat{x}_i\|^2 \quad (8)$$

This loss guarantees that the model learns to extract the original sensor values from corrupted inputs.

The masking probability directly controls the trade-off between reconstruction accuracy and robustness. If a low masking probability is used, reconstruction accuracy would be high, but the ability to withstand faults would be limited, as only a small proportion of the sensor information was corrupted during training. However, a very high masking probability can result in the loss of significant information, making reconstruction more challenging and impacting the quality of the latent features and classification accuracy. Based on the experimental evaluation, a masking probability of 20% was selected as an effective compromise, providing reliable signal restoration while maintaining robustness against sensor dropouts, noise injection, and partial sensor failures

2.5 Latent and Reconstruction-Assisted Fusion Classification

The normalized sensor measurements, reconstructed sensor signals, and latent representations learned by the MDAE are concatenated to form a unified fusion feature vector. This feature vector is subsequently classified using a calibrated linear Support Vector Machine (SVM). The SVM decision function is expressed as in Eq. (9):

$$s = W_f F + b_f \quad (9)$$

where

- F is fusion feature vector
- s is SVM decision scores

The calibrated posterior probabilities are subsequently evaluated using standard probabilistic performance measures including Negative Log-Likelihood (NLL), Brier Score, and Expected Calibration Error (ECE). The corresponding cross-entropy formulation is expressed as in Eq. (10):

$$\mathcal{L}_{\text{cls}} = - \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(\hat{y}_{i,c}) \quad (10)$$

The raw SVM decision scores are converted into posterior probabilities using sigmoid (Platt) calibration before probabilistic evaluation.

The proposed framework Overview is shown in Fig. 1. The model consists of preprocessing, masking-based corruption, denoising autoencoder reconstruction, latent feature extraction, and classification. The detailed architecture of the proposed MDAE-based framework is shown in Fig. 2. It demonstrates the masked input, encoder-decoder architecture, latent representation, and classification layer. The architecture combines reconstruction and classification into a single framework. The encoder is used to squeeze the input into a latent space, and the decoder is used to rebuild the signal. The classifier is applied to the latent representation, making it robust.

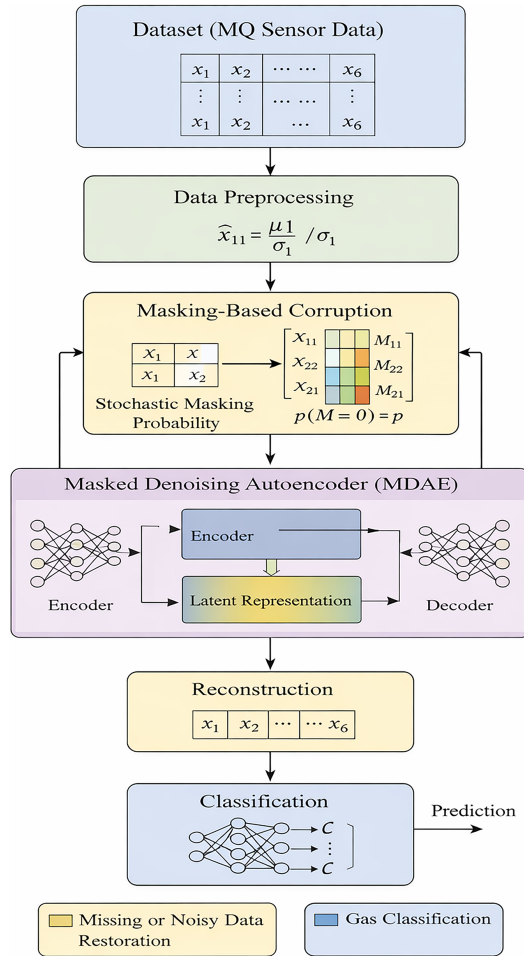


Figure 1: Overall flowchart of the proposed method

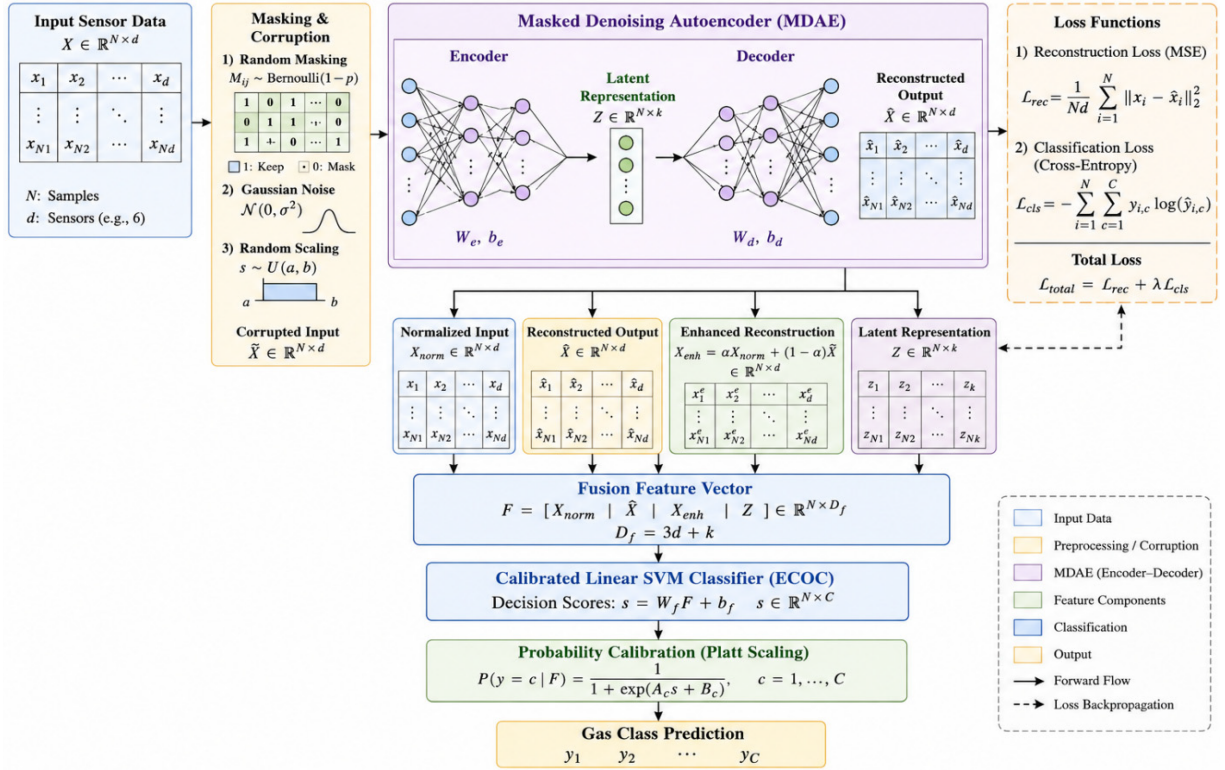


Figure 2: Detailed architecture of the proposed MDAE-based framework

2.6 Experimental Setup

The dataset was partitioned into 69.9975% training, 14.9995% validation, and 15.0030% testing subsets using stratified sampling, as summarized in Table 2. Stratified partitioning preserves the class distribution across all subsets, ensuring that each gas category is proportionally represented in the training, validation, and testing sets. This approach minimizes sampling bias and improves the reliability of performance evaluation. Consequently, the relative proportions of all five gas classes remain approximately unchanged across the three data partitions. Exact feature-label overlap was verified after dataset partitioning. No duplicate feature-label pairs were observed between the training, validation, and testing subsets, confirming complete leakage-free evaluation.

Table 2: Data split configuration

Set	Samples	Ratio
Training	19,908	69.9975%
Validation	4266	14.9995%
Testing	4267	15.0030%

The initial MDAE framework was developed and validated in MATLAB. The final implementation, including probabilistic calibration, robustness analysis, leakage-controlled evaluation, cross-validation, and repeated hold-out validation, was performed in Python (Anaconda) using PyTorch and scikit-learn. A fixed random seed (42) was used throughout all experiments to ensure reproducible data partitioning and model training. The proposed masked denoising autoencoder (MDAE) employed an encoder architecture with 32 hidden neurons followed by a 16-neuron bottleneck layer and symmetric decoder reconstruction layers. ReLU activation functions were used throughout the hidden layers. The network was trained using the Adam optimizer with a learning rate of 0.001, a mini-batch size of 256, L2 regularization of 1×10^{-4} , and a maximum of 120 training epochs with early stopping (patience = 20) based on validation loss. During training, a masking probability of 0.20 was applied, along with Gaussian corruption noise ($\sigma = 0.05$) and random signal scaling between 0.95 and 1.05, to simulate realistic sensor degradation conditions. For classification, the normalized sensor measurements, reconstructed outputs, and latent representations were concatenated to form a fusion feature vector, which was subsequently classified using a calibrated linear Fusion SVM with sigmoid (Platt) probability calibration. All experiments were executed on a workstation equipped with an NVIDIA RTX 5090 GPU and 64 GB RAM.

2.7 Ablation Study Design

To assess the impact of the individual component, the combinations as given in [Table 3](#) are explored:

Table 3: Ablation configurations

Model	Description
A0	Raw features + SVM
A1	Raw features + MLP
A2	Autoencoder + SVM
A3	Proposed MDAE + classifier

This experimental design separates the impact of reconstruction from that of masking.

2.8 Cross-Validation Strategy

To evaluate the stability and generalization capability of the proposed framework, 5-fold stratified cross-validation was performed, as shown in [Eq. \(11\)](#):

$$\mathcal{D} = \bigcup_{k=1}^5 \mathcal{D}_k \quad (11)$$

Each fold was used once as the validation fold, while the remaining four were used for training.

The last performance is calculated as given in [Eq. \(12\)](#):

$$\text{Metric}_{avg} = \frac{1}{5} \sum_{k=1}^5 \text{Metric}_k \quad (12)$$

3 Results and Discussion

3.1 Dataset Characteristics and Sensor Relationships

The final class distribution of the leakage-controlled MQ sensor dataset is shown in Fig. 3 after removing duplicates and the under-represented class. While the five gas classes are not perfectly balanced, each class has enough samples to provide reliable supervised learning and stratified evaluation. Removing duplicate observations reduces the risk of overly optimistic performance while preserving the distribution of the remaining gas categories.

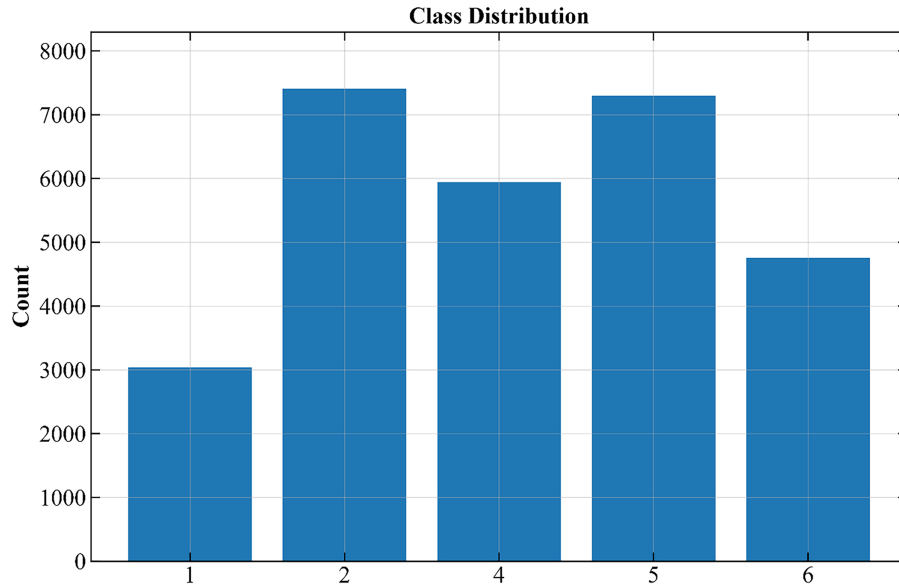


Figure 3: Class distribution of the MQ sensor dataset

The class imbalance lies between 0.02 and 0.30, which gives an idea of real-world conditions to assess the performance of intelligent gas classification systems. The proposed masked denoising autoencoder (MDAE) learns robust representations in the latent space across all sensor channels, thereby mitigating the impact of redundancy, measurement noise, and partial sensor failures. As a result, even with the different class frequencies, the learned feature representations are discriminative for all gas categories.

Fig. 4 illustrates the Pearson correlation matrix for the six MQ gas sensors after preprocessing. The correlation analysis provides insight into the redundancy and complementarity of the sensor responses. Strong positive correlations are observed between certain sensor pairs, such as MQ-5 and MQ-6, indicating partially overlapping sensing characteristics. In contrast, weaker or negative correlations are observed for pairs such as MQ-5 and MQ-9, suggesting that these sensors capture complementary information about the gas environment. The coexistence of both correlated and weakly correlated sensor channels justifies the proposed fusion strategy, which learns complementary latent representations from multiple sensors to improve classification robustness and reduce sensitivity to individual sensor failures or measurement noise.

The observed sensor diversity also supports the masking-based denoising strategy adopted in the proposed MDAE, enabling the framework to recover useful latent information even when individual sensor measurements are corrupted or unavailable.

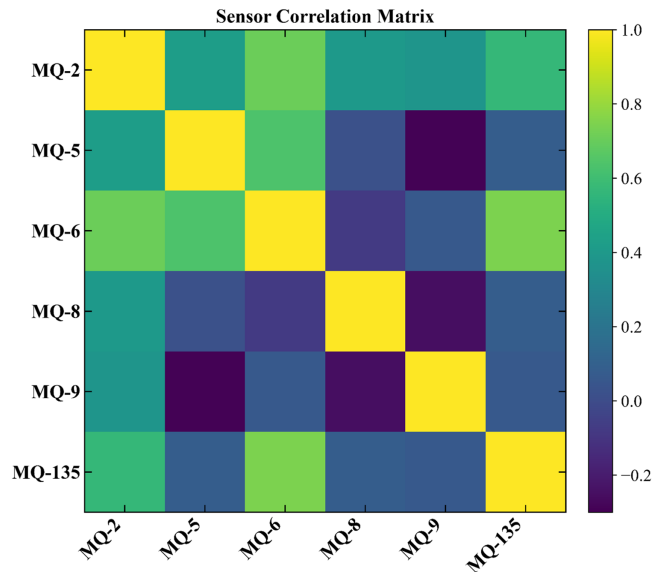


Figure 4: Sensor correlation matrix

3.2 Classification Performance

The classification performance of the proposed framework was first evaluated using the confusion matrix shown in Fig. 5, which provides a class-wise assessment of prediction accuracy on the independent leakage-controlled test set.

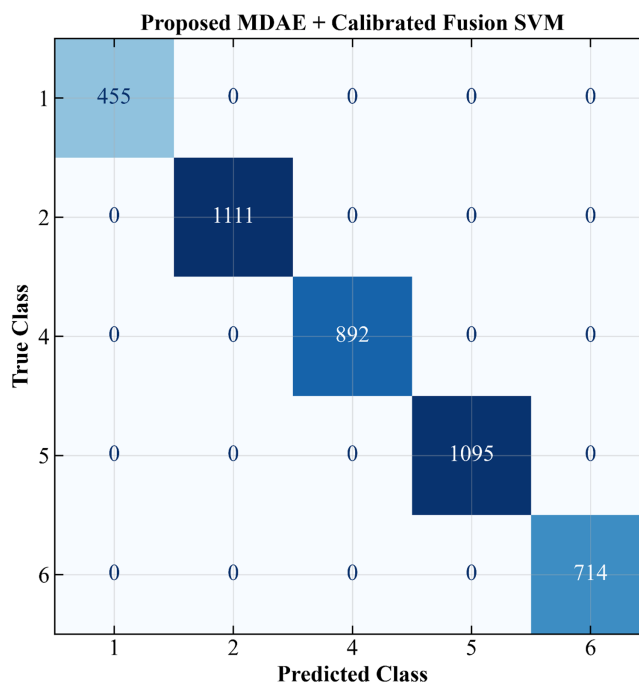


Figure 5: Confusion matrix of the proposed model demonstrating perfect classification performance across all classes

The independent de-duplicated test set consisted of 4267 samples. The proposed MDAE-based fusion framework correctly classified every test sample, resulting in a confusion matrix with no off-diagonal entries. Consequently, all five gas classes achieved 100% class-wise recognition accuracy, indicating that no false positives or false negatives were observed under the leakage-controlled evaluation protocol.

Fig. 6 presents the receiver operating characteristic (ROC) curves of the proposed framework. All five gas classes achieved an area under the ROC curve (AUC) of 1.000, indicating perfect class separability on the independent leakage-controlled test set. The ROC curves closely follow the upper-left boundary, demonstrating excellent discriminative capability with negligible overlap among the class distributions.

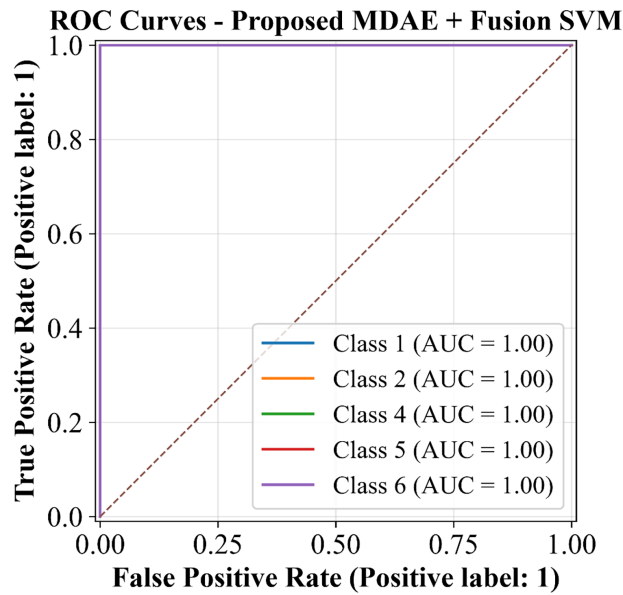


Figure 6: ROC curves of the proposed model

The proposed fusion strategy is based on the assumption that multiple sensors share correlated and weakly correlated sensor channels, suggesting that complementary latent representations can be learned across sensors to make classification robust to the failure of any single sensor or measurement noise. The perfect classification performance is largely due to the natural separability of the MQ sensor dataset, enabled by leakage-controlled preprocessing and duplicate removal. Only five (1, 2, 4, 5, and 6) of the six gas classes in the final dataset had sufficient numbers of unique samples; Gas Class 3 had only one sample left after de-duplication. The other classes show sufficiently different characteristics of their sensor response, allowing for near-perfect discrimination in nominal operating conditions with both the conventional classifiers and the proposed framework.

Thus, classification accuracy is not a good measure for distinguishing the proposed framework from traditional machine learning approaches on this dataset. Rather, the key benefit of the proposed MDAE framework is its ability to reconstruct corrupted sensor measurements, sustain reliable classification even in the presence of sensor degradation, provide well-calibrated probabilistic predictions, and increase robustness to sensor dropouts and measurement noise. The benefits of these are further illustrated in the ablation study, cross-validation, probability calibration analysis, and robustness experiments discussed in the next sections.

3.3 Ablation Study

The ablation study results are summarized in Figs. 7 and 8, where the classification accuracy of the configurations (A0–A3) tested is compared, and the probabilistic calibration of the configurations tested is compared using the Brier Score in Fig. 8.

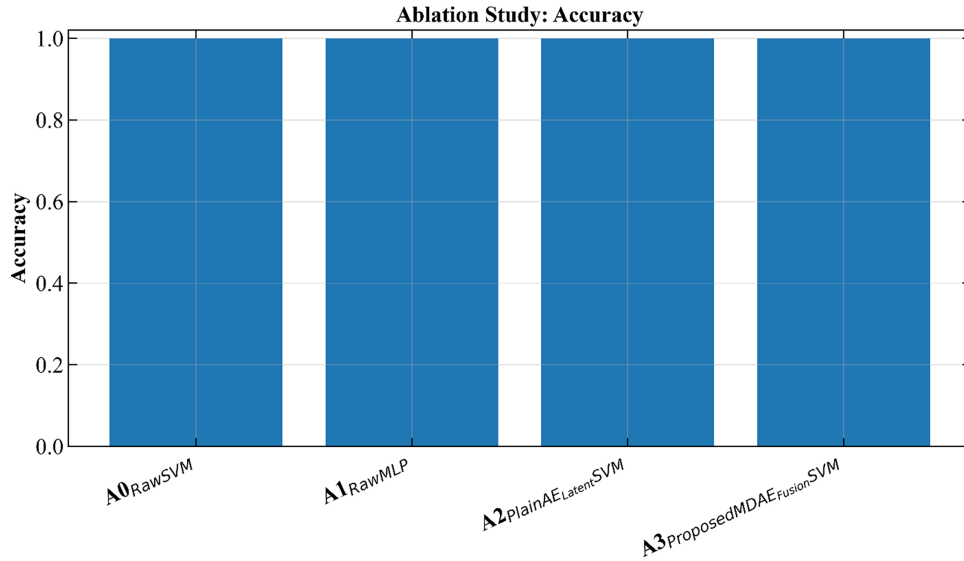


Figure 7: Ablation study: accuracy comparison

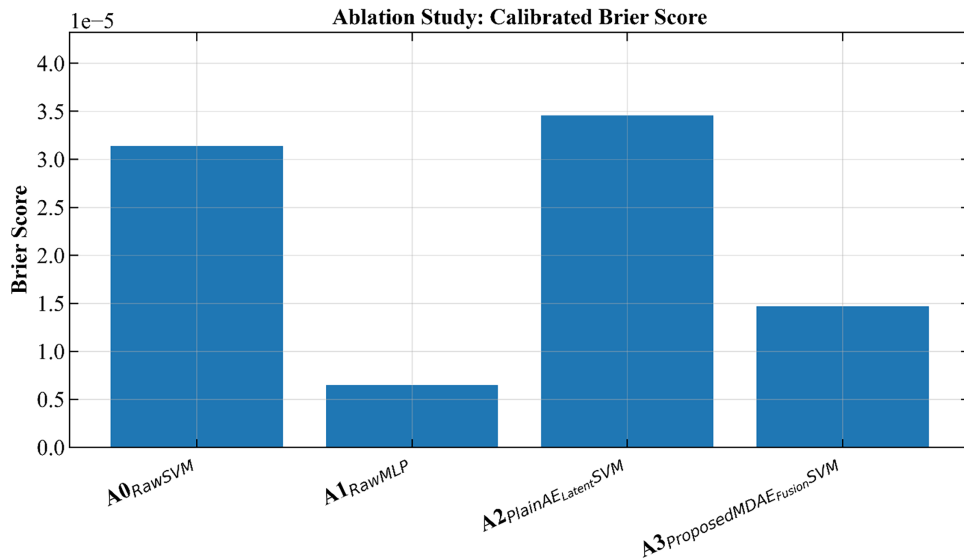


Figure 8: Ablation study: brier score comparison

As observed in Fig. 7, all the analyzed configurations (A0–A3) yielded almost 100% classification accuracy on the leakage-controlled MQ sensor data set. This observation verifies that the data set is intrinsically well separated and that even common machine learning models can achieve very high classification accuracy under nominal operating conditions. It is therefore not sufficient to rely solely on classification accuracy to differentiate the proposed approach from simpler baseline approaches.

It will be more informative to compare the quality of the predicted probabilities, as done in Fig. 8. The proposed MDAE-based Fusion SVM has been calibrated using the sigmoid (Platt) calibration method, and the resulting Brier Score, NLL, and ECE values are 0.000015, 0.002172, and 0.002167, respectively, demonstrating high reliability of probabilistic predictions. The calibrated Random Forest and MLP gave slightly lower Brier Scores; however, these models lack the ability to reconstruct, learn latent features, restore sensors, or be robust to sensor loss and measurement noise.

The ablation study on the proposed framework therefore shows that the major benefit lies not in classification accuracy, but in integrating all these steps—masking-based denoising, reconstruction-assisted feature learning, probabilistic calibration, and fault-tolerant sensing—into a single framework. Such capabilities are especially useful when considering realistic operating conditions in which sensors degrade, as shown in the following robustness experiments.

The comparison also demonstrates that each of the suggested framework elements can provide the necessary functionality, masking makes it more robust, reconstruction makes it more feature-rich, and calibrated Fusion SVM provides more reliable probabilistic decisions.

3.4 Performance Validation

3.4.1 5-Fold Cross-Validation

The generalization ability of the proposed framework was tested using 5-fold stratified cross-validation, and the resulting classification accuracy is shown in Fig. 9.

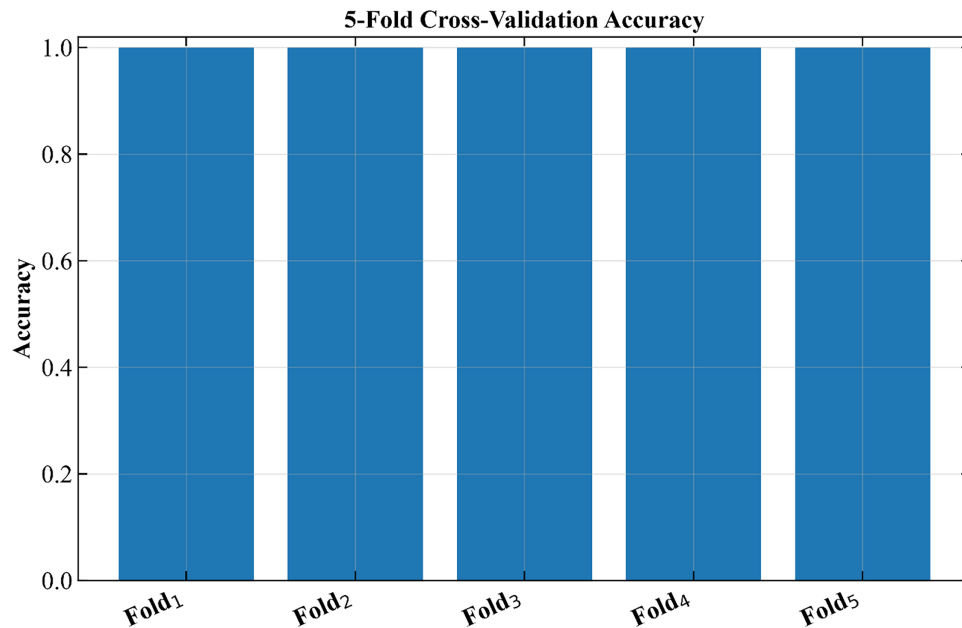


Figure 9: 5-fold cross-validation accuracy

The proposed framework showed almost similar classification accuracy for all five folds with an average accuracy of 1.000 ± 0.000 , as shown in Fig. 9. The small difference between the folds suggests that the feature representations learned are robust with respect to the split of the training and validation data sets and are not sensitive to the specific split.

The results indicate that the proposed framework has achieved good generalization across all folds and does not amplify the negative impact when the validation partition changes. This is because the leakage-controlled MQ sensor set has a natural separability that is unlikely to be advantageous for a train-test split. Additionally, there is no significant variation between the folds that would indicate that overfitting to a particular fold is the cause of the results reported.

The proposed framework was observed to have consistently high Macro-F1 and Macro-AUC, and excellent probabilistic calibration across all folds, as evidenced by the Brier Score, Negative Log Likelihood (NLL), and Expected Calibration Error (ECE) reported in [Table 4](#). All these results show the robustness, stability, and repeatability of the proposed framework.

Table 4: Statistical validation of the proposed framework using 5-fold cross-validation

Metric	Mean	Standard deviation	95% Confidence interval
Accuracy	1.000000	0.000000	[1.000000, 1.000000]
Macro-F1	1.000000	0.000000	[1.000000, 1.000000]
Macro-AUC	1.000000	0.000000	[1.000000, 1.000000]
Brier score	0.000029	0.000013	[0.000013–0.000045]
NLL	0.002547	0.000230	[0.002261–0.002833]
ECE	0.002539	0.000227	[0.002257–0.002820]

3.4.2 Statistical Validation

The results from the 5-fold cross-validation experiments were used to assess the reliability, stability, and generalization potential of the proposed framework through statistical validation. The mean, standard deviation, and 95% confidence interval were calculated for each evaluation metric. A confidence interval is the interval computed as shown in [Eq. \(13\)](#):

$$CI_{95\%} = \bar{x} \pm t_{0.975, n-1} \frac{s}{\sqrt{n}} \quad (13)$$

where $\bar{x}$ is the mean value, s is the standard deviation, $n = 5$ is the number of folds, and $t_{0.975, n-1}$ is the critical value of Student's t -distribution. The statistical validation is given in [Table 4](#).

The statistical findings show good consistency across all folds of cross-validation. The standard deviation is zero for Accuracy, Macro-F1, and Macro-AUC, signifying that the classification performance is the same for all folds. However, the MQ sensor dataset is highly separable post-leakage-controlled preprocessing, which causes these measures to become saturated and thus not distinguishable between competing methods. The Brier Score, Negative Log-Likelihood (NLL), and Expected Calibration Error (ECE) are more informative, as they provide a measure of the quality of the predicted probabilities beyond just the predicted class labels. The small standard deviations and consistently low mean values of these calibration metrics for the proposed framework give further evidence of the stability, reliability, and well-calibration of probabilistic predictions across the different data partitions.

3.5 Reconstruction Capability

The restoration presentation is shown in [Fig. 10](#).

Input vs. Reconstructed Sensor Patterns

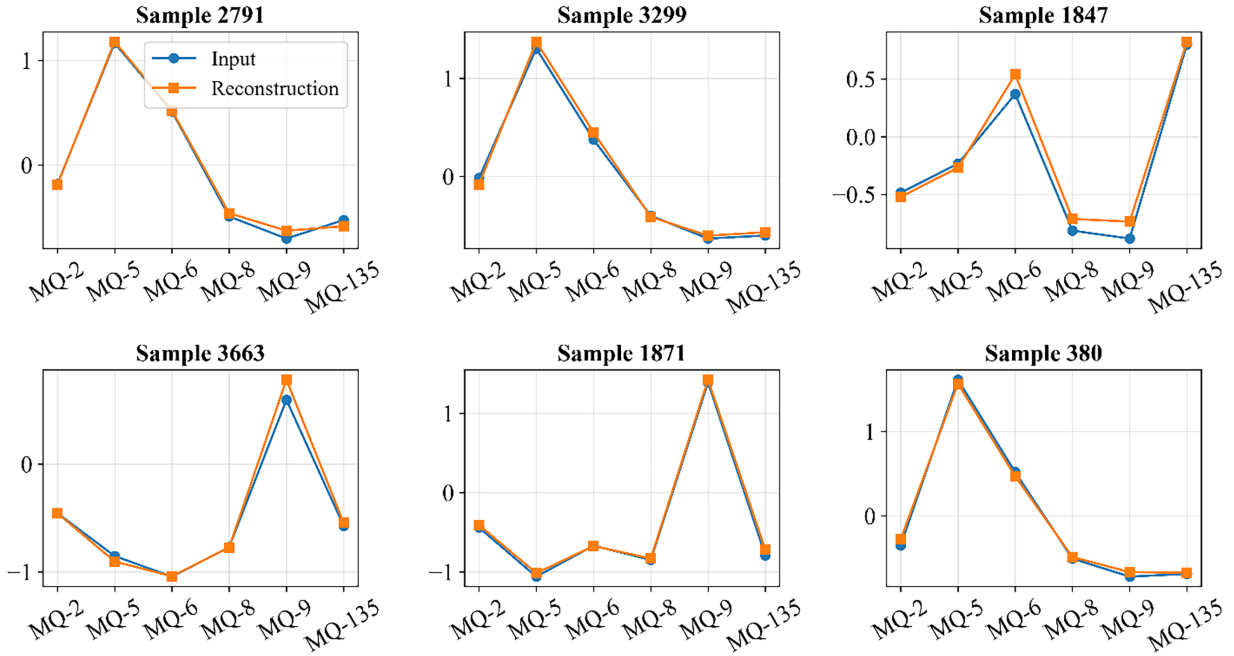


Figure 10: Input vs. reconstructed sensor patterns

The reconstructed signals have close resemblance to the original inputs in all the samples and maintain the following:

- Amplitude patterns
- Relative sensor relationships
- Signal trends

This confirms that the masked denoising autoencoder can effectively learn the underlying data distribution and reconstruct corrupted or missing signals.

3.6 Feature Importance Analysis

The permutation importance results are shown in Fig. 11.

From the permutation analysis, it is seen that MQ-135 and MQ-6 are most strongly influencing the classification, and MQ-2 is least strongly influencing it. This establishes that the model uses redundant sensor channels in addition to dominant ones for robust decision-making.

3.7 Robustness under Sensor Dropout

The robustness results under sensor failure are shown in Fig. 12.

When the individual sensors (MQ-2, MQ-5, MQ-6, MQ-8, or MQ-9) were dropped, the associated performance drop was negligible. In moderate cases of MQ-135 dropping, degradation was moderate, and the most severe cases of multi-sensor dropout lowered accuracy to about 72%–74%. The results show that these systems can provide fault tolerance in the presence of single-sensor and severe multi-sensor failures.

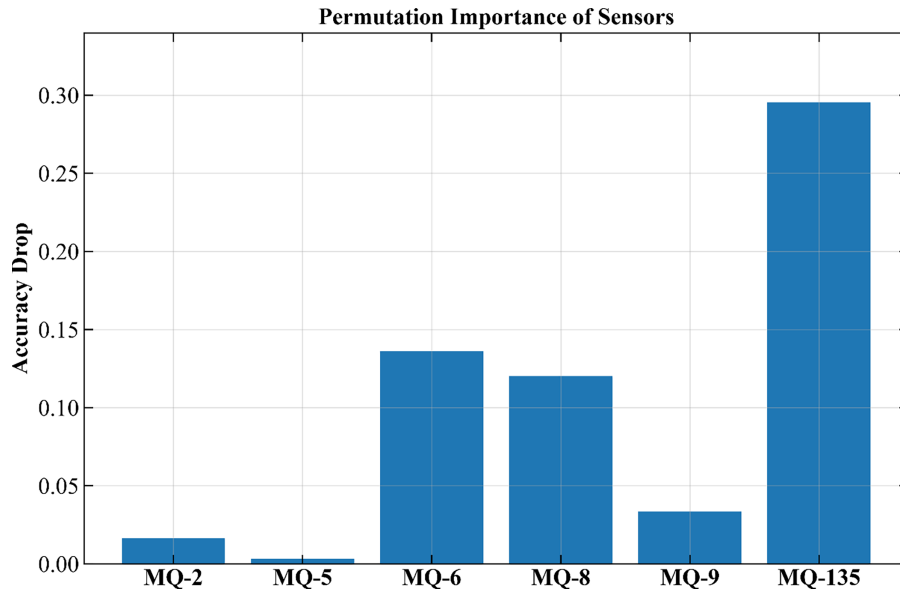


Figure 11: Permutation importance of sensors

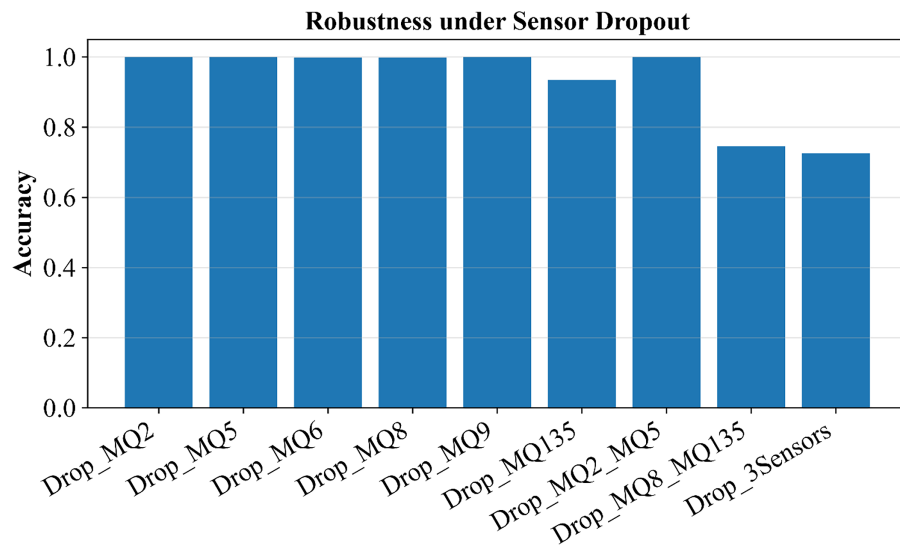


Figure 12: Performance under sensor dropout scenarios

3.8 Robustness under Noise Injection

The performance under Gaussian noise is shown in Fig. 13.

At low and moderate noise levels up to $\sigma = 0.20$, the proposed framework maintained nearly perfect accuracy and macro-F1. Performance decreased gradually at higher perturbation levels, reaching approximately 96% accuracy at $\sigma = 0.50$. This confirms graceful degradation under severe sensor noise.

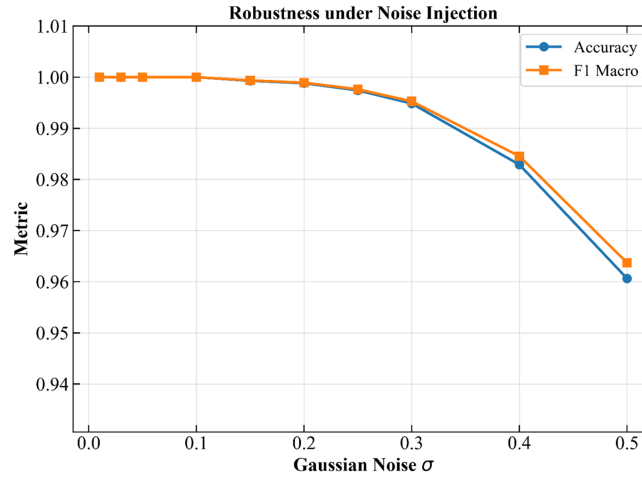


Figure 13: Robustness under noise injection

The outcomes show that the proposed model not only improves classification accuracy but also enhances reliability and fault tolerance, which are essential for the application of an e-nose system in the real world.

The alternative interpretation of the robustness results is from a communication perspective. The scenarios of Gaussian noise injection and sensor dropouts simulate channel impairments commonly found in real-life distributed sensing networks. The proposed MDAE framework can correctly classify data with high accuracy even in the presence of these disturbances, demonstrating its ability to extract useful information from corrupted sensor data. Consequently, the proposed method improves information reliability in a manner analogous to fault-resilient communication systems that employ redundancy and error-recovery mechanisms to preserve data integrity under adverse channel conditions.

3.9 Computational Complexity and Deployment Considerations

To evaluate practical deployability, the computational cost of the proposed MDAE framework was compared with the baseline models used in the ablation study. The results are summarized in Table 5, where both offline training cost and online inference latency are reported for all ablation-study models. The raw ECOC-SVM model (A0) and raw MLP model (A1) have the lowest computational cost because they operate directly on the normalized sensor features. The plain autoencoder-based model (A2) introduces an additional representation learning stage and therefore requires longer training time. The proposed MDAE model (A3) further incorporates stochastic masking, denoising reconstruction, latent feature extraction, and reconstruction-assisted fusion. Consequently, a higher offline training time is required compared with the baseline models.

However, the inference cost remains very low for all models. The proposed A3 model required only 0.00258 ms per test sample, only slightly higher than A2 and still suitable for real-time gas classification. The masking operation is used only during training and does not increase inference-time complexity. Therefore, although the proposed method incurs higher offline training costs, its deployment-time latency remains negligible for practical electronic noses and edge-based intelligent sensing systems.

Table 5: Computational complexity and deployment-time performance comparison of the evaluated models

Model	Main components	Training time (s)	Inference time (s)	Inference time per sample (ms)
A0	Raw linear SVM	3.247	0.0165	0.00387
A1	Raw MLP	0.550	0.0200	0.00469
A2	Plain AE + Latent SVM	17.846	0.00652	0.00153
A3	Proposed MDAE + Fusion SVM	19.074	0.0110	0.00258

These results indicate that the proposed MDAE framework primarily increases offline training cost, while maintaining a very low inference latency of only 0.00258 ms per sample. Therefore, the method remains suitable for real-time electronic nose and intelligent sensing applications.

3.10 Comparison with Recent State-of-the-Art Methods

To further evaluate the effectiveness of the proposed framework, its performance was qualitatively compared with recent state-of-the-art methods reported in the electronic nose and intelligent sensing literature. Table 6 summarizes this comparison. Traditional machine learning approaches, such as the hybrid SVM (VSVM) proposed by Zhang et al. [3], achieve excellent classification performance but do not explicitly address sensor corruption, missing data, or fault recovery. Hu et al. [13] proposed a deep learning method for sensor drift compensation that is robust to sensor drift but does not reconstruct corrupted sensor measurements. Adaptive multi-sensor fusion methods, such as those proposed by Zhao et al. [16], improve information integration but assume that valid sensor measurements remain available. Reliability-oriented methods proposed by Guo et al. [18], Kendall and Gal [19], and Lakshminarayanan et al. [21] mainly focus on probability calibration and uncertainty estimation rather than sensor reconstruction or fault recovery. The proposed communication-aware MDAE framework addresses the key limitations of existing electronic nose methods by jointly integrating sensor reconstruction, robust latent representation learning, reliability-aware probabilistic classification, and fault-tolerant gas classification within a unified architecture. Unlike previous approaches that focus primarily on classification accuracy, sensor drift compensation, multi-sensor fusion, or uncertainty estimation individually, the proposed framework combines these capabilities into a single end-to-end solution. Experimental results demonstrate excellent classification performance (100% accuracy and macro-AUC of 1.000), high-quality probability calibration (Brier score = 0.000015, NLL = 0.002172, and ECE = 0.002167), and strong robustness under sensor dropout and Gaussian noise. These characteristics make the proposed framework well-suited for reliable, intelligent sensing and communication-aware electronic nose applications.

3.11 Comparison with Classical Machine Learning Baselines

To further evaluate the effectiveness of the proposed communication-aware masked denoising autoencoder framework, additional experiments were conducted using several widely adopted machine learning classifiers. The baseline methods selected are Random Forest (RF), Kernel Support Vector Machine (Kernel SVM), Boosted Decision Trees (BDT), Multilayer Perceptron (MLP) and ECOC-SVM. These are some of the most popular algorithms that are used for gas classification and intelligent sensing. The comparative performance results on the test set are shown in Table 7.

Table 6: Comparison with recent state-of-the-art electronic nose gas classification methods

Ref.	Method	Dataset/ Scenario	Accuracy (%)	Fault tolerance	Reliability analysis	Reconstruction capability
Zhang et al. [3]	Hybrid SVM	Indoor contaminants	95–98	No	No	No
Hu et al. [13]	Deep learning under sensor drift	Drift dataset	~97	Partial	No	No
Zhao et al. [16]	Multi-sensor fusion	Pump monitoring	High	Partial	No	No
Wang et al. [14]	SG + Image conversion	Gas recognition	~98	No	No	No
Guo et al. [18]	Calibration framework	Generic DL	N/A	No	Yes	No
Kendall and Gal [19]	Bayesian DL	Generic DL	N/A	No	Yes	No
Lakshminaray anan et al. [21]	Deep ensemble	Generic DL	N/A	No	Yes	No
Proposed	MDAE + Fusion SVM	MQ dataset	100	Yes	Yes	Yes

Table 7: Comparison with classical machine learning baselines

Model	Accuracy (%)	Macro-F1	Macro- AUC	Sensor reconstruction	Fault restoration	Robustness analysis
ECOC-SVM	100.00	1.000	1.000	No	No	No
MLP	100.00	1.000	1.000	No	No	No
Plain AE + SVM	100.00	1.000	1.000	Limited	Partial	No
Kernel SVM (RBF)	100.00	1.000	1.000	No	No	No
Random forest	100.00	1.000	1.000	No	No	No
Boosted trees	100.00	1.000	1.000	No	No	No
Proposed MDAE + Fusion SVM	100.00	1.000	1.000	Yes	Yes	Yes

The achieved results show that the separation between classes in the measurements collected from the MQ sensor data set is very good, and all classical machine learning methods tested in the work perform excellent classification. Although classification accuracy is important, it is not sufficient to fully describe a model's usefulness for practical deployment of an electronic nose. The proposed communication-aware MDAE framework integrates masked denoising reconstruction, sensor fault restoration, latent feature learning, probabilistic reliability assessment, sensor-dropout resilience, and robustness evaluation under noise perturbations, features not included in traditional classifiers. As a result, the proposed framework will offer greater robustness to sensor degradation and impairments that are characterized by conditions similar to those of communication channels, while still achieving high classification accuracy for gases. The proposed method has capabilities that make it more appropriate for real-world intelligent sensing and communication-related electronic nose applications.

3.12 External Validation on an Independent Dynamic Gas Mixture Dataset

To evaluate the generalizability of the proposed method, another validation study was conducted on the publicly available Gas Sensor Array under Dynamic Gas Mixtures dataset. This dataset contains dynamic gas-mixture measurements that are collected under different sensing conditions. The dataset includes sensor measurements from various gas-sensor channels under ethylene–CO and ethylene–methane mixture scenarios. In order to evaluate transferability, a balanced subset containing 100,000 samples (50,000 samples per mixture class) was constructed. The proposed methodology has been followed in the same manner as the preprocessing pipeline. The results obtained are summarized in Table 8. The proposed approach has yielded classification accuracy very close to the optimum. It achieves an accuracy of 99.99%, precision of 99.99%, recall of 99.99%, macro-F1 score of 99.99%, and macro-AUC of 0.9999. The successful performance on this independent dataset provides additional evidence that the proposed method is transferable beyond the original MQ sensor dataset. The presented results demonstrate strong robustness and generalization across different gas-sensing scenarios.

Table 8: External validation results on the dynamic gas mixture dataset

Metric	Proposed MDAE framework
Accuracy (%)	99.99
Precision (%)	99.99
Recall (%)	99.99
Macro-F1 (%)	99.99
Macro-AUC	0.9999

Although the proposed method is performing well on both the MQ sensor dataset and the independent Dynamic Gas Mixture dataset. However, the practical deployment will be an important future research direction. Future work will focus on hardware-level implementation using real-time electronic nose systems. Furthermore, extensive field testing under dynamic environmental conditions will be conducted to further evaluate the robustness and practical applicability of the proposed framework.

4 Conclusions

This paper presents a reliability-aware, fault-tolerant electronic nose (e-nose) framework based on a communication-aware masked denoising autoencoder (MDAE) for robust gas classification. The proposed framework addresses key limitations of conventional e-nose systems, including sensor noise, redundancy, missing measurements, and partial sensor failures, by learning robust latent representations through structured masking and denoising reconstruction. Experimental evaluation on a leakage-controlled MQ sensor dataset demonstrated excellent classification performance, achieving 100.00% accuracy, a macro-AUC of 1.000, and outstanding probability calibration with a Brier score of 0.000015, NLL of 0.002172, and ECE of 0.002167. The ablation study confirmed the effectiveness of the masking strategy, denoising reconstruction, and reconstruction-assisted feature learning, while the 5-fold cross-validation results demonstrated the stability and generalization capability of the proposed framework across different data partitions. Robustness experiments further showed that the framework maintained reliable classification performance under challenging conditions, including sensor dropout and Gaussian noise injection, exhibiting graceful performance degradation under severe

sensing impairments. In addition, the reconstruction capability enabled effective recovery of corrupted sensor signals, thereby improving the overall system's reliability and fault tolerance. Permutation-based feature importance analysis also provided valuable insights into the contribution of individual sensors, enhancing the interpretability of the proposed framework. Furthermore, external validation on the independent Dynamic Gas Mixture dataset demonstrated that the proposed framework maintained excellent classification performance across substantially different sensing environments and gas-mixture conditions, providing additional evidence of its generalization capability. Overall, the proposed communication-aware MDAE framework integrates sensor reconstruction, robust latent representation learning, probability-calibrated classification, interpretability, and fault tolerance within a unified architecture, making it well suited for reliable intelligent sensing and communication-aware electronic nose applications. Future work will focus on real-time hardware implementation, adaptive sensor calibration, and integration with edge-based intelligent sensing systems to further enhance practical deployment in industrial and environmental monitoring applications.

Acknowledgement: The author is thankful to the Deanship of Graduate Studies and Scientific Research at Najran University for funding this work under the Growth Funding Program grant code NU/GP/SERC/14/3682-1. Furthermore, The Grammarly professional version has been used for language assistance, paraphrasing and grammar improvement.

Funding Statement: The author is thankful to the Deanship of Graduate Studies and Scientific Research at Najran University for funding this work under the Growth Funding Program grant code NU/GP/SERC/14/3682-1.

Availability of Data and Materials: The data could be accessed through <https://www.kaggle.com/datasets/thehanyss/mq-sensors-for-electronic-nose-system> (accessed on 20 January 2026). External Validation Dataset Link: <https://archive.ics.uci.edu/dataset/322/gas+sensor+array+under+dynamic+gas+mixtures> (accessed on 29 May January 2026).

Ethics Approval: Not applicable. This study did not involve human participants, animal subjects, or any biological materials requiring ethical approval or Institutional Review Board (IRB) approval.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Dutta R, Hines EL, Gardner JW, Boilot P. Bacteria classification using Cyranose 320 electronic nose. *Biomed Eng Online*. 2002;1:4. doi:10.1186/1475-925x-1-4.
2. Wilson AD, Baietto M. Applications and advances in electronic-nose technologies. *Sensors*. 2009;9(7):5099–148. doi:10.3390/s90705099.
3. Zhang L, Tian F, Nie H, Dang L, Li G, Ye Q, et al. Classification of multiple indoor air contaminants by an electronic nose and a hybrid support vector machine. *Sens Actuators B Chem*. 2012;174:114–25. doi:10.1016/j.snb.2012.07.021.
4. Potyrailo RA. Multivariable sensors for ubiquitous monitoring of gases in the era of Internet of Things and industrial Internet. *Chem Rev*. 2016;116(19):11877–923. doi:10.1021/acs.chemrev.6b00187.
5. Li Z, Askim JR, Suslick KS. The optoelectronic nose: colorimetric and fluorometric sensor arrays. *Chem Rev*. 2019;119(1):231–92. doi:10.1021/acs.chemrev.8b00226.
6. Sanislav T, Mois GD, Zeadally S, Folea S, Radoni TC, Al-Suhaimi EA. A comprehensive review on sensor-based electronic nose for food quality and safety. *Sensors*. 2025;25(14):4437. doi:10.3390/s25144437.

7. Yu X, Fu L, Wang T, Liu Z, Niu N, Chen L. Multivariate chemical analysis: from sensors to sensor arrays. *Chin Chem Lett.* 2024;35(7):109167. doi:10.1016/j.ccllet.2023.109167.
8. Wang C, Chen Z, Chan CLJ, Wan ZA, Ye W, Tang W, et al. Biomimetic olfactory chips based on large-scale monolithically integrated nanotube sensor arrays. *Nat Electron.* 2024;7(2):157–67. doi:10.1038/s41928-023-01107-7.
9. Sun X, Wang S, Jia W. Research progress of electronic nose and near-infrared spectroscopy in meat adulteration detection. *Chemosensors.* 2024;12(3):35. doi:10.3390/chemosensors12030035.
10. Tasmurzayev N, Amangeldy B, Smagulova G, Baigarayeva Z, Imash A. A low-cost IoT sensor and preliminary machine-learning feasibility study for monitoring in-cabin air quality: a pilot case from Almaty. *Sensors.* 2025;25(14):4521. doi:10.3390/s25144521.
11. Han Z, Ahmad W, Rong Y, Chen X, Zhao S, Yu J, et al. A gas sensors detection system for real-time monitoring of changes in volatile organic compounds during oolong tea processing. *Foods.* 2024;13(11):1721. doi:10.3390/foods13111721.
12. Wang B, Yan Y, An R, Weng X, Chang Z. Application of electronic nose technology in wastewater treatment plants monitoring: progress and future developments. *Microchem J.* 2026;224:117768. doi:10.1016/j.microc.2026.117768.
13. Hu X, Liu Q, Cai H, Li F. Gas recognition under sensor drift by using deep learning. In: *Practical applications of intelligent systems.* Berlin, Heidelberg: Springer; 2014. p. 23–33. doi:10.1007/978-3-642-54927-4_3.
14. Wang X, Qian C, Zhao Z, Li J, Jiao M. A novel gas recognition algorithm for gas sensor array combining Savitzky–Golay smooth and image conversion route. *Chemosensors.* 2023;11(2):96. doi:10.3390/chemosensors11020096.
15. Fundurulic A, Faria JM, Inácio ML. Advances in electronic nose sensors for plant disease and pest detection. *Eng Proc.* 2023;48(1):14. doi:10.3390/CSAC2023-14890.
16. Zhao Y, Yang X, Wu X, Huang J, Peng K. Adaptive weighting strategy based multi-sensor data fusion method for condition monitoring of reciprocating pump. In: *Proceedings of the 2021 CAA Symposium on Fault Detection, Supervision, and Safety for Technical Processes; 2021 Dec 17–18; Chengdu, China.* p. 1–6. doi:10.1109/SAFEPROCESS52771.2021.9693735.
17. Goodfellow I, Bengio Y, Courville A. *Deep learning.* Cambridge, MA, USA: MIT Press; 2016.
18. Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. *arXiv:1706.04599.* 2017. doi:10.48550/arXiv.1706.04599.
19. Kendall A, Gal Y. What uncertainties do we need in Bayesian deep learning for computer vision? *arXiv:1703.04977.* 2017. doi:10.48550/arXiv.1703.04977.
20. Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. *arXiv:1506.02142.* 2015. doi:10.48550/arXiv.1506.02142.
21. Lakshminarayanan B, Pritzel A, Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. *arXiv:1612.01474.* 2016. doi:10.48550/arXiv.1612.01474.
22. Abdar M, Pourpanah F, Hussain S, Rezazadegan D, Liu L, Ghavamzadeh M, et al. A review of uncertainty quantification in deep learning: techniques, applications and challenges. *Inf Fusion.* 2021;76:243–97. doi:10.1016/j.inffus.2021.05.008.
23. Chow C. On optimum recognition error and reject tradeoff. *IEEE Trans Inf Theory.* 1970;16(1):41–6. doi:10.1109/TIT.1970.1054406.
24. Geifman Y, El-Yaniv R. Selective classification for deep neural networks. *arXiv:1705.08500.* 2017. doi:10.48550/arXiv.1705.08500.
25. Geifman Y, El-Yaniv R. SelectiveNet: a deep neural network with an integrated reject option. *arXiv:1901.09192.* 2019. doi:10.48550/arXiv.1901.09192.