Celal Bayar University Journal of Science
Volume 22, Issue 2, 2026, p 401-409
Doi: 10.18466/cbayarfbe.1926951
Research Article

Celal Bayar University Journal of Science

A Compact and Explainable Machine Learning Pipeline for Low-Concentration Gas Sensor Array Classification

Bora Canbula[1*]

[1] Manisa Celal Bayar University, Department of Computer Engineering, Yunusemre, 45140 Manisa, Türkiye.
* bora.canbula@cbu.edu.tr
* Orcid No: 0000-0003-1088-2804

Received: April 9, 2026
Accepted: June 10, 2026
DOI: 10.18466/cbayarfbe.1926951

Abstract

This study investigates whether compact and physically interpretable time-domain descriptors can support accurate low-concentration gas classification without relying on full multichannel waveforms. Using the Gas Sensor Array Low-Concentration dataset, each sample was transformed from raw sensor signals into 120 descriptors that summarize baseline behavior, variability, transient dynamics, and response magnitude. Four classical learning pipelines were evaluated through repeated stratified 5-fold cross-validation, supported by principal component analysis, descriptor ranking, sensor correlation analysis, and sensor subset experiments. The best-performing model, a radial support vector machine trained on the compact descriptor set, achieved a mean accuracy of 0.9456 and a macro-averaged F1 score of 0.9437 across 100 test folds while using 75 times fewer inputs than the raw waveform representation. An important finding is that classification performance improved systematically with concentration: accuracy was 0.9067 at 50 ppb, increased to 0.9333 at 100 ppb, and reached 0.9967 at 200 ppb. Response mean and baseline mean emerged as the most informative descriptor families, while VOCS-P, 2M012, and MQ-137 were the most discriminative sensors. Overall, the results show that compact and interpretable descriptors provide an efficient, reproducible, and practically useful benchmark for low-concentration gas classification in resource-constrained electronic nose systems.

Keywords: classification, electronic nose, feature engineering, interpretability, low-concentration gas, sensor array, support vector machine

1. Introduction

Gas sensor arrays transform complex chemical environments into multichannel electrical response patterns and play a central role in electronic nose systems used for environmental monitoring, industrial process control, food quality evaluation, and breath-based diagnostics [1–7]. In such applications, achieving high classification accuracy is essential; however, it is not sufficient on its own. Practical systems also require representations that are reproducible, computationally efficient, and interpretable enough to support sensor selection, system deployment, and operational reliability.

A fundamental challenge in electronic nose data analysis arises from the structure of the data itself. Each observation is typically a multivariate time series composed of long response curves obtained from multiple sensors with partially overlapping sensitivities. When these high-dimensional raw waveforms are directly used for classification, the combination of dimensionality, inter-sensor redundancy, and limited sample size increases the likelihood of overfitting [4,5,8–13]. This issue becomes more critical at low-concentration levels, where sensor responses are weak and class-discriminative patterns can be masked by baseline variations and noise [5,8,14–16].

Recent advances in machine learning and deep learning have improved performance in various sensor-based applications, including gas classification, human activity recognition, and odor descriptor prediction [6,7,14,17–22]. In electronic nose research, feature extraction, bioinspired preprocessing, classifier ensembles, and edge-oriented architectures have also been investigated to improve separability, robustness, and deployment feasibility [23–26]. Nevertheless, complex models are not always the most appropriate solution for compact datasets. When the number of samples is limited and data acquisition is costly, feature

engineering approaches remain valuable because they provide transparent representations that facilitate interpretation and enable direct analysis of the relationship between sensor response characteristics and classification performance [8,12,27–32].

The Gas Sensor Array Low-Concentration dataset introduced by Zhao et al. [16] provides a relevant benchmark for investigating this problem. The dataset contains high-dimensional waveform measurements from six volatile gases at concentrations as low as 50 ppb, with a total of 90 samples. In the original study, strong performance was achieved using feature-ensemble extreme learning machines constructed from downsampled signal representations and classifier aggregation [16]. A recent reproducible benchmark further revisited the same dataset using time-series classifiers and time-series-to-image convolutional neural networks, showing that high performance can also be obtained with more complex temporal representations [33]. While these results demonstrate that the dataset is discriminative, they do not directly address how much of this information can be preserved in a significantly reduced and physically interpretable feature space.

1.1. Problem Definition and Contributions

The problem addressed in this study is small-sample, high-dimensional low-concentration gas classification: can a 9000-point multichannel waveform be compressed into a small set of descriptors while retaining classification performance and exposing which sensors and response phases are informative? The novelty of the proposed approach is therefore not a new classifier architecture, but a deliberately compact and explainable time-domain feature benchmark that complements downsampled waveform ensembles, bioinspired preprocessing, multi-classifier systems, and recent time-series benchmark models [16,23–26,33].

The main contributions are fourfold. First, each sample is represented by 120 fixed, physically meaningful descriptors derived from baseline and response segments, reducing the input dimensionality by a factor of 75. Second, four classical supervised learning pipelines are evaluated under repeated stratified cross-validation to provide a reproducible benchmark. Third, descriptor-family ranking, sensor-level analysis, correlation analysis, and sensor subset experiments are used to interpret model behavior. Fourth, performance is analyzed separately across concentration levels to show how the compact representation behaves under the most challenging 50 ppb regime and at higher signal strengths.

1.2. Paper Organization

The remainder of the paper is organized as follows. Section 2 presents the dataset, descriptor extraction procedure, learning pipelines, validation protocol, and interpretability analyses. Section 3 reports gas–sensor response patterns, descriptor rankings, classification results, concentration-specific behavior, and sensor subset experiments. Section 4 concludes the study and outlines future directions related to drift robustness, concentration estimation, and confidence-aware real-time decisions.

2. Materials and Methods

The analysis was performed using the Gas Sensor Array Low-Concentration dataset, which is provided in comma-separated format and was experimentally characterized by Zhao et al. [16]. In this study, measurements were obtained from a 10-channel metal oxide semiconductor sensor array operating at a sampling rate of 1 Hz, with 5 min of baseline acquisition, 10 min of gas exposure, and 15 min of desorption. The dataset includes 6 target gases—ethanol, acetone, toluene, ethyl acetate, isopropanol, and n-hexane—measured at 3 concentration levels (50, 100, and 200 ppb), with 5 repeated measurements for each gas–concentration combination [16].

Table 1. Structure of the Gas Sensor Array Low-Concentration dataset.

Item	Value
Gas classes	6
Concentration levels	3 (50, 100, 200 ppb)
Replicates per gas-concentration pair	5
Total samples	90
Sensors	10
Time points per sensor	900
Raw features per sample	9000
Compact descriptors per sample	120

In the released dataset, each sample consists of a gas label, a concentration label, and 9000 sensor readings. After reshaping, each observation is represented as 10 sensor channels with 900 time points per channel. Following the acquisition protocol, the first 300 points were treated as the baseline segment, while the remaining 600 points were considered as the response segment [16]. This formulation results in a balanced yet small-sample classification problem characterized by high dimensionality and substantial temporal and cross-sensor redundancy. The overall structure of the dataset is summarized in Table 1.

To address these challenges, the raw waveform representation was replaced by a compact time-domain descriptor set designed to preserve key signal characteristics while significantly reducing

dimensionality. For each sensor trace, the baseline segment, full response segment, and early, middle, and late portions of the response were summarized independently. A total of 12 descriptors were extracted per sensor, yielding 120 descriptors per sample across all channels. The definitions and interpretive roles of these descriptors are summarized in Table 2. This representation reduces the dimensionality of the input space by a factor of 75 relative to the original waveform, while maintaining physically interpretable features related to signal level, variability, transient behavior, and overall response magnitude.

Table 2. Compact descriptor families extracted from each sensor trace.

Descriptor	Definition	Interpretive role
baseline mean	Mean value of points 1-300	Nominal sensor level before gas injection
baseline standard deviation	Standard deviation of points 1-300	Baseline stability and noise
response mean	Mean value of points 301-900	Average response magnitude
response standard deviation	Standard deviation of points 301-900	Response variability
peak delta	Maximum response minus baseline mean	Largest positive deviation
trough delta	Minimum response minus baseline mean	Largest negative deviation
early delta	Mean of points 301-450 minus baseline mean	Initial transient change
mid delta	Mean of points 451-750 minus baseline mean	Intermediate regime
late delta	Mean of points 751-900 minus baseline mean	Late or plateau behavior
response area	Integrated response relative to baseline	Overall response burden
early slope	Early delta divided by early-window length	Approximate onset rate
recovery	Late-window mean minus early-window mean	Drift from early to late response

The selected descriptor families were designed to retain both signed and unsigned information. Signed descriptors, such as peak delta, trough delta, and recovery, capture directional deviations from baseline, including cases where sensor responses decrease under gas exposure. In contrast, unsigned descriptors, such as standard deviation measures, characterize signal stability and variability. This combined representation was intentionally chosen because low-concentration gas recognition is often influenced by subtle baseline shifts and by differences in the temporal evolution of sensor responses [8,16].

Four supervised learning pipelines were evaluated to assess the effectiveness of the compact representation. The first pipeline used uniformly downsampled waveform data with a radial support vector machine and served as a baseline model. The remaining three pipelines employed the compact descriptor set in combination with logistic regression, a linear support vector machine, and a radial support vector machine, respectively. In all cases, feature standardization was applied exclusively within the training portion of each fold to prevent information leakage from the test data.

Model performance was evaluated using 20 repetitions of stratified 5-fold cross-validation, resulting in a total of 100 test folds. For each fold and each model, classification accuracy and macro-averaged F1 score were computed. For the best-performing model, a 95% confidence interval was also estimated for the mean accuracy. Given that the dataset is class-balanced, accuracy and macro-averaged F1 score provide complementary and directly comparable measures of classification performance.

In addition to predictive performance, several analyses were conducted to better understand model behavior and the contribution of individual features. Principal component analysis was used to visualize class separability in the standardized descriptor space. One-way analysis of variance F-statistics were computed for each descriptor with respect to gas class, and the resulting scores were aggregated by descriptor family and by sensor to identify the most informative signal components. Correlation analysis of sensor-wise response shifts was performed to detect redundancy among channels and to identify sensors with complementary or opposing response patterns. Furthermore, sensor subset experiments were carried out by ranking sensors within each training fold and re-evaluating the best-performing model using only the highest-ranked channels. This procedure was treated as a nested exploratory analysis rather than a separate benchmark model.

All analyses were implemented in Python within a fully scripted pipeline to ensure reproducibility. Descriptor extraction, cross-validation, metric aggregation, and result generation were performed without manual intervention. This design choice is particularly important for small electronic nose datasets, which are susceptible to preprocessing inconsistencies, unintended

Celal Bayar University Journal of Science
Volume 22, Issue 2, 2026, p 401-409
Doi: 10.18466/cbayarfbe.1926951

data leakage, and undocumented train–test splits. Accordingly, the study adopts a deliberately transparent and reproducible workflow in which all results are derived from fixed descriptor definitions and repeated cross-validation procedures. While this does not guarantee global optimality of the selected descriptors, it ensures that the benchmark can be reliably audited, extended, and compared with future approaches that may incorporate calibration transfer, drift compensation, or confidence-aware decision mechanisms.

3. Results and Discussion

The results are first examined from a descriptive perspective before evaluating classification performance, since the effectiveness of the compact descriptor representation depends on whether it preserves physically meaningful signal characteristics. Accordingly, the analysis begins with gas–sensor response patterns and their relationship to descriptor discriminability.

As shown in Figure 1, the 6 gases do not activate the 10 sensors uniformly. Ethanol and acetone produce the strongest positive responses on MQ-137, whereas toluene induces negative shifts on several channels, including TGS813, TGS822, and MQ-135. Ethyl acetate and isopropanol exhibit partially overlapping response patterns, particularly on TGS822 and MQ-135, suggesting a chemically consistent source of confusion. In addition, several channels display weak responses under the current operating conditions, indicating that not all sensors contribute equally to class discrimination.

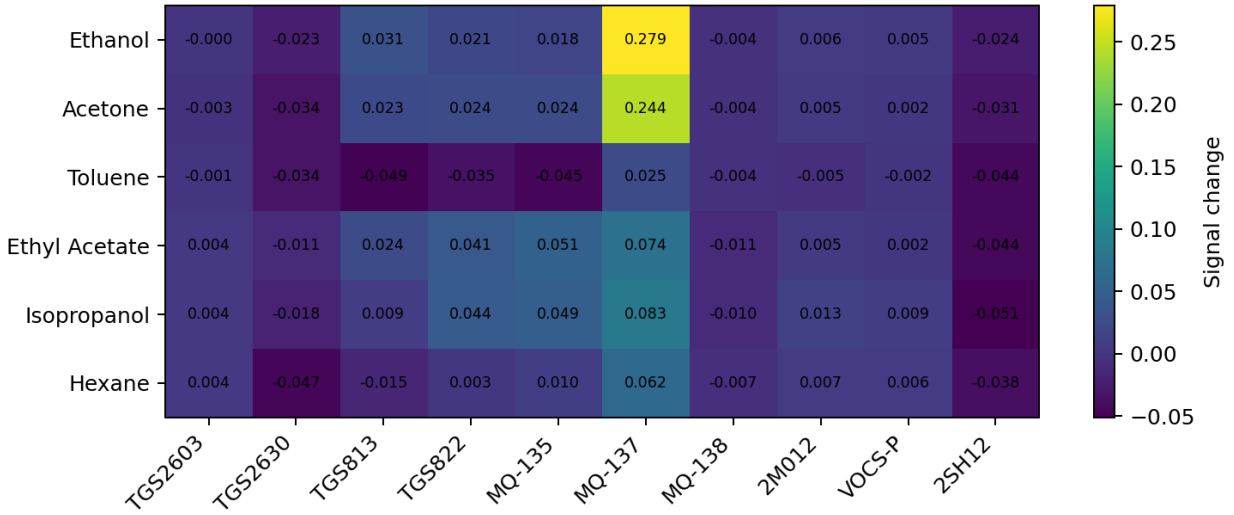

Figure 1. Mean response-window change relative to baseline for each gas-sensor pair.

To further examine discriminative behavior, representative response curves for the most informative channels are presented in Figure 2, which focuses on VOCS-P, 2M012, and MQ-137. VOCS-P and 2M012 primarily separate gases through stable level differences, whereas MQ-137 shows a pronounced transient increase for ethanol and acetone. This contrast explains why descriptors based on response magnitude and early-stage dynamics remain highly informative despite substantial dimensionality reduction.

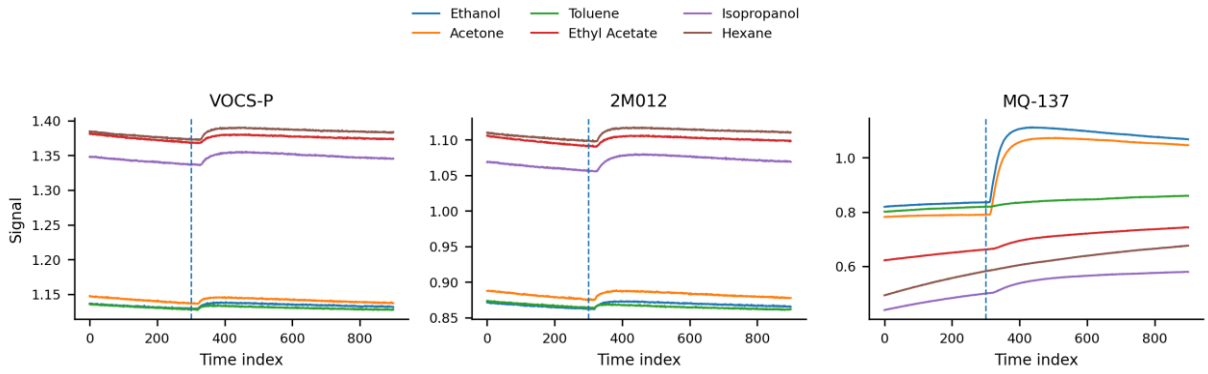

Figure 2. Mean response curves at 100 ppb for the three most discriminative sensors. The dashed line marks the start of the response window.

The relationship between sensors was further analyzed through correlation structure, as shown in Figure 3. Several channels exhibit strong redundancy, with correlations close to 0.99 between TGS822 and MQ-135, and near-perfect correlation between 2M012 and VOCS-P. These results indicate that a significant portion of the discriminative information is duplicated across sensors. In contrast, 2SH12 shows negative correlations with multiple informative channels, suggesting a complementary but limited role. These findings explain why carefully selected sensor subsets can recover most of the full-array performance.

The contribution of individual descriptor families is summarized in Figure 4, which shows that response mean and baseline mean dominate the discriminative ranking, with mean F-scores of 96.14 and 67.27, respectively. All remaining descriptor families cluster near or below approximately 11, indicating that the steady-state signal level is more informative than fine-grained waveform details for this dataset. Nevertheless, transient descriptors such as early slope and early delta remain relevant for distinguishing rapidly responding gases from slower or negatively responding classes.

The most discriminative descriptors are listed in Table 3, which shows that top-ranked features are concentrated on VOCS-P, 2M012, and MQ-137. Response mean and baseline mean repeatedly appear among the highest-ranked features, confirming that classification is primarily driven by exposure-induced shifts combined with stable pre-exposure offsets. This result is particularly important from an application perspective, as it favors computationally efficient descriptors suitable for embedded implementations.

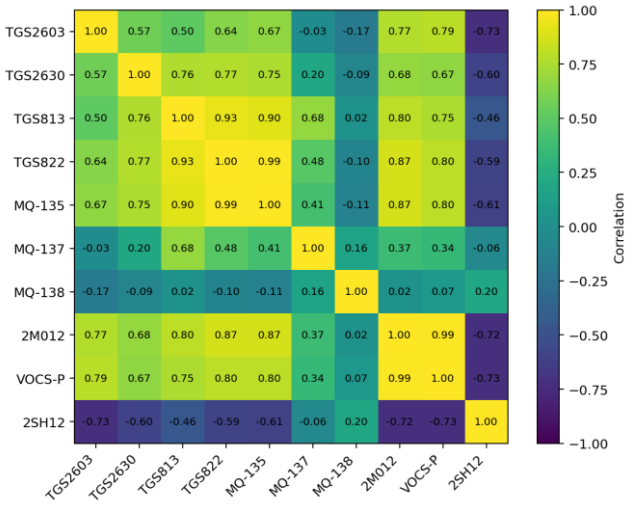

Figure 3. Correlation heatmap of sensor-wise mean response shifts across all 90 samples.

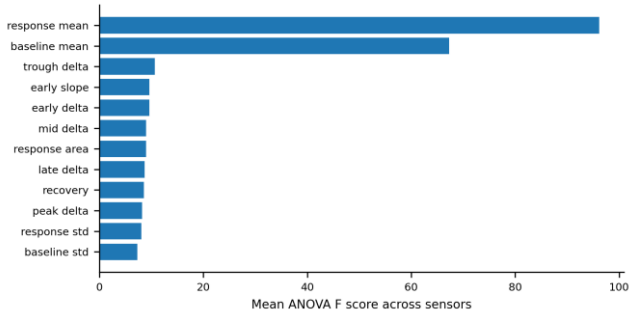

Figure 4. Mean analysis-of-variance F score of each descriptor family, averaged across sensors

Table 3. 10 most discriminative compact descriptors according to one-way analysis of variance.

Rank	Sensor	Descriptor	F score
1	VOCS-P	response mean	353.27
2	VOCS-P	baseline mean	293.28
3	2M012	response mean	285.97
4	2M012	baseline mean	242.46
5	MQ-137	response mean	131.59
6	MQ-135	response mean	54.16
7	TGS2603	response mean	51.03
8	MQ-137	early slope	36.91
9	MQ-137	early delta	36.91
10	MQ-137	baseline mean	35.82

The geometric structure of the descriptor space is visualized in Figure 5, where the first two principal components explain 60.7% of the variance. Several gases form well-separated clusters, while partial overlap is observed among oxygenated solvents. This structure indicates that the compact descriptor representation preserves most of the discriminative information, although perfect separation is not expected under low-signal conditions.

Quantitative performance results are summarized in Table 4 and further illustrated in Figure 6. The compact descriptor representation combined with a radial support vector machine achieves the best performance, with an accuracy of 0.9456 ± 0.0547 and a 95% confidence interval of [0.9347, 0.9564]. All descriptor-based models outperform the downsampled waveform baseline, demonstrating that compact, physically meaningful features are more effective than simple temporal decimation. Notably, the obtained performance is comparable to the 0.9444 accuracy reported by Zhao et al.

[16]. A recent reproducible benchmark on the same GSA-LC dataset reported 0.9578 ± 0.0433 accuracy with ROCKET [33], which is slightly higher, but that approach is based on transformed time-series kernels rather than a small set of physically interpretable descriptors. These comparisons suggest that most of the discriminative information can be preserved within a significantly reduced and interpretable feature space.

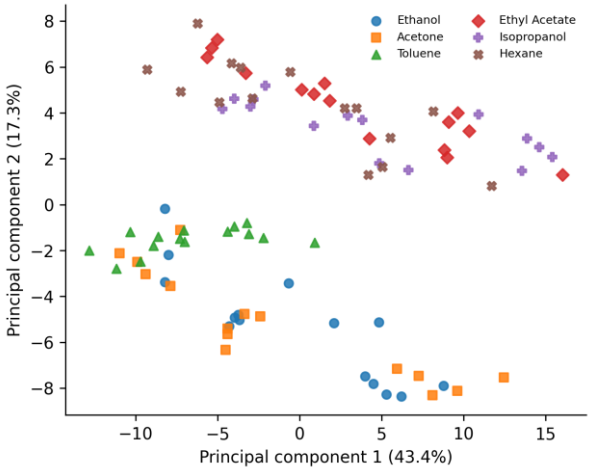

Figure 5. Principal component projection; the first two components explain 60.7% of the variance.

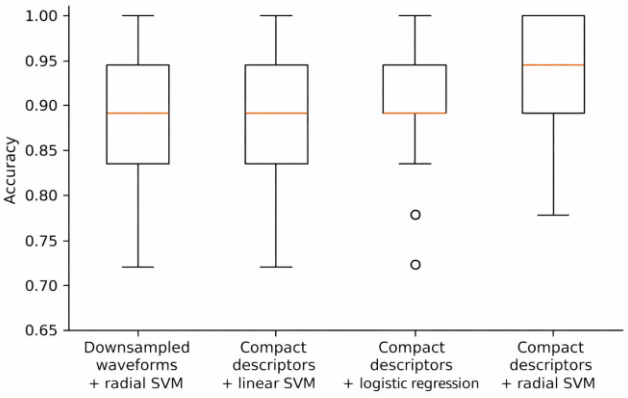

Figure 6. Accuracy distribution across 100 test folds for the compared pipelines.

A more detailed understanding of model behavior is obtained from the confusion matrix in Figure 7. Ethanol and acetone are occasionally confused with each other, while ethyl acetate and isopropanol form a second confusion pair. Toluene exhibits the lowest recall, particularly due to misclassification at 100 ppb, whereas hexane is rarely misclassified except at the lowest signal levels. These patterns are consistent with the response characteristics observed in Figures 1 and 2.

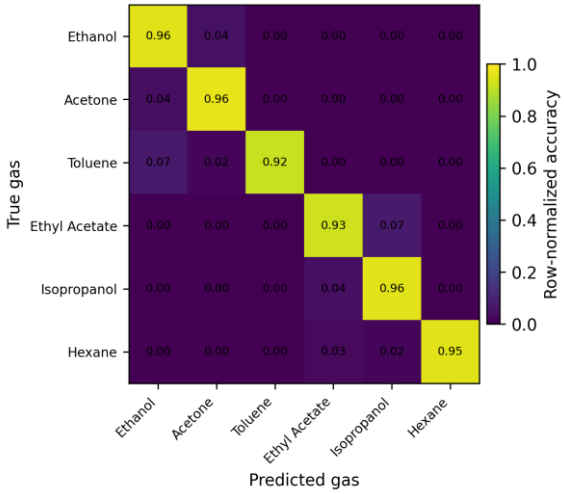

Figure 7. Row-normalized confusion matrix of the best-performing compact descriptor model.

Class-wise performance metrics are reported in Table 5, where toluene achieves perfect precision but lower recall (0.9167), indicating conservative classification behavior. Hexane and acetone yield the highest F1 scores, while ethanol, ethyl acetate, and isopropanol remain more prone to confusion.

An important empirical finding of this study is the systematic improvement in classification performance with increasing concentration, as summarized in Table 6 and visualized in Figure 8. Accuracy increases from 0.9067 at 50 ppb to 0.9333 at 100 ppb and reaches 0.9967 at 200 ppb.

Table 4. Repeated stratified five-fold cross-validation results for the compared pipelines.

Model	Dim.	Accuracy	Macro-F1
Downsampled waveforms + radial support vector machine	150	0.8744 ± 0.0705	0.8695 ± 0.0761
Compact descriptors + linear support vector machine	120	0.9017 ± 0.0646	0.8984 ± 0.0668
Compact descriptors + logistic regression	120	0.9067 ± 0.0688	0.9044 ± 0.0705
Compact descriptors + radial support vector machine	120	0.9456 ± 0.0547	0.9437 ± 0.0563

Table 5. Class-wise precision, recall, and F1 score aggregated from repeated cross-validation predictions of the best model.

Gas	Precision	Recall	F1	Support
ethanol	0.9025	0.9567	0.9288	300
acetone	0.9383	0.9633	0.9507	300
toluene	1.0000	0.9167	0.9565	300
ethyl acetate	0.9329	0.9267	0.9298	300
isopropanol	0.9140	0.9567	0.9349	300
hexane	0.9965	0.9533	0.9744	300

Table 6. Concentration-specific performance of the best compact descriptor model.

Concentration	N	Accuracy	Macro-F1
50 ppb	600	0.9067	0.9074
100 ppb	600	0.9333	0.9336
200 ppb	600	0.9967	0.9967

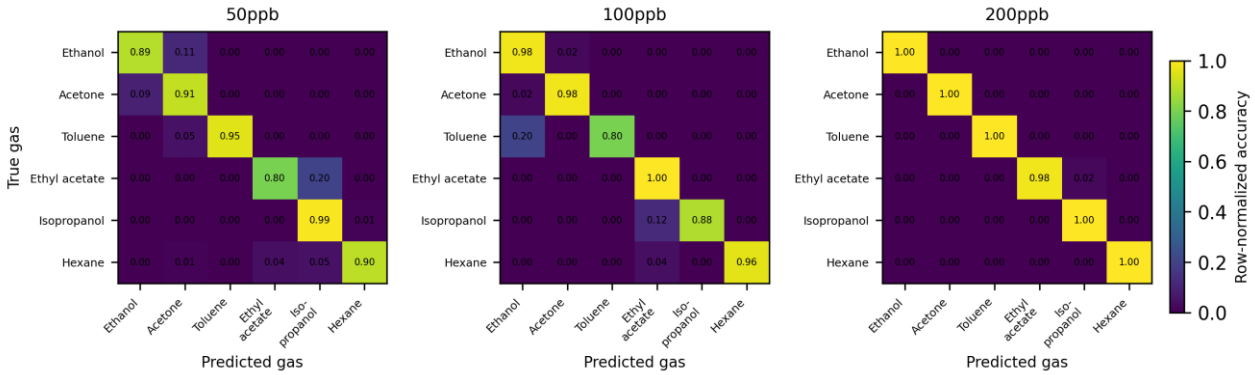

Figure 8. Concentration-specific row-normalized confusion matrices for the best model.

This trend demonstrates that classification performance is more limited under low-concentration conditions, whereas near-perfect discrimination is achieved at higher concentrations. Importantly, the errors observed at 50 ppb exhibit a structured pattern rather than random behavior, with ethanol and acetone being mutually confused, ethyl acetate occasionally misclassified as isopropanol, and hexane sometimes shifting toward alcohol-like classes. At 200 ppb, however, the confusion matrix becomes nearly perfectly diagonal, indicating that the descriptor set effectively captures the relevant signal structure when sufficient signal strength is available.

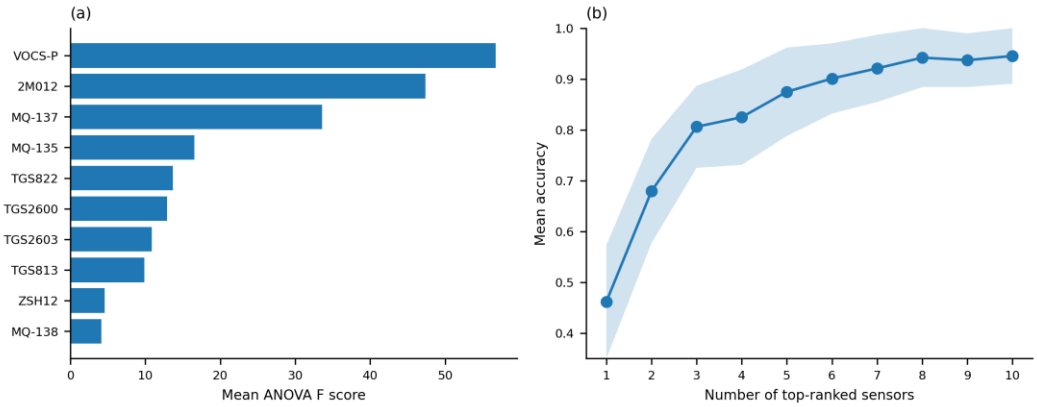

Figure 9. (a) Sensor ranking based on grouped analysis-of-variance scores. (b) Mean accuracy of the best classifier when only the top-ranked sensors are retained within each training fold.

The effect of sensor selection is analyzed in Figure 9, which shows that VOCS-P, 2M012, and MQ-137 are consistently ranked as the most informative sensors. Performance increases rapidly as top-ranked sensors are added, reaching an accuracy of 0.9072 with 6 sensors, 0.9422 with 8 sensors, and 0.9456 with all 10 sensors. The near-saturation after 8 sensors confirms that the array contains significant redundancy and suggests that reduced sensor configurations may be sufficient for practical implementations.

From a real-world application perspective, the compact descriptor representation offers a substantial advantage by transforming a high-dimensional waveform classification problem into a computationally efficient and interpretable feature space. The descriptors can be computed from baseline and response windows, so an embedded device does not need to retain the complete 9000-point waveform after feature extraction. The sensor ranking also provides a practical route for simplifying the array, reducing power consumption, and lowering maintenance cost. These properties are consistent with recent edge-oriented, low-power, and robustness-focused electronic nose developments [26,34,35]. However, deployment would require additional operational modules, including baseline stabilization, humidity and temperature monitoring, drift surveillance, periodic recalibration or transfer learning, and confidence-aware rejection of uncertain samples [35,36]. Thus, the proposed pipeline should be viewed as a lightweight decision layer that can be integrated into a broader real-time sensing workflow rather than as a complete field-deployment solution.

Despite these advantages, several limitations should be noted. The dataset contains only 90 samples, which limits the generalizability of the results. In addition, the analysis is restricted to a single public dataset and does not address calibration transfer, long-term drift, humidity and temperature disturbances, or domain adaptation, which remain central challenges in deployed electronic nose systems [35,36]. The selected descriptors are intentionally simple and may not capture more complex temporal dynamics relevant to multi-component gas mixtures. Furthermore, the current formulation treats each sample independently and does not exploit the temporal or longitudinal structure that may arise in real-world deployments.

These limitations also clarify the contribution of the study as a benchmark. The results show that the dataset is not trivially separable at 50 ppb, yet a compact and interpretable descriptor set is sufficient to capture most of the available structure. This indicates that future improvements should focus on genuinely challenging aspects such as drift robustness, uncertainty estimation, and adaptation to changing environmental conditions, rather than solely increasing model complexity.

Overall, the findings demonstrate that compact and physically interpretable descriptors provide a strong and reproducible baseline for low-concentration gas classification, while also offering practical insights for sensor selection and system design.

4. Conclusion

This benchmark study shows that a compact and interpretable time-domain representation can support strong low-concentration gas classification on the Gas Sensor Array Low-Concentration dataset. Compressing each sample from 9000 waveform values to 120 descriptors still yielded a mean accuracy of 0.9456 and a macro-averaged F1 score of 0.9437 across 100 test folds. An important finding is that classification performance increased markedly with concentration as accuracy remained limited at 50 ppb (0.9067), improved at 100 ppb (0.9333), and became almost perfect at 200 ppb (0.9967). The analysis also showed that response mean and baseline mean dominate the descriptor ranking, that the strongest signals are concentrated on VOCS-P, 2M012, and MQ-137, and that 8 carefully selected sensors nearly recover the full-array result. These findings make the compact descriptor set valuable not only as a reproducible benchmark, but also as a practical design aid for future lightweight electronic nose systems. Subsequent studies can build on this baseline by incorporating drift robustness, concentration estimation, and explicit confidence scoring for real-time sensor decisions.

Acknowledgement

The author has no acknowledgements to declare.

Author's Contributions

Bora Canbula: Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, visualization, writing-original draft, and writing-review and editing.

Ethics

There are no ethical issues after the publication of this manuscript.

References

[1]. James, D, Scott, SM, Ali, Z, O'Hare, WT. 2005. Chemical sensors for electronic nose systems. *Microchimica Acta*; 149(1-2): 1-17.

[2]. Röck, F, Barsan, N, Weimar, U. 2008. Electronic nose: current status and future trends. *Chemical Reviews*; 108(2): 705-725.

[3]. Tan, J, Xu, J. 2020. Applications of electronic nose and electronic tongue in food quality-related properties determination: a review. *Artificial Intelligence in Agriculture*; 4: 104-115.

[4]. Scott, SM, James, D, Ali, Z. 2006. Data analysis for electronic nose systems. *Microchimica Acta*; 156(3): 183-207.

[5]. Ye, Z, Liu, Y, Li, Q. 2021. Recent progress in smart electronic nose technologies enabled with machine learning methods. *Sensors*; 21(22): 7620.

[6]. Zhai, Z, Liu, Y, Li, C, Wang, D, Wu, H. 2024. Electronic noses: from gas-sensitive components and practical applications to data processing. *Sensors*; 24(15): 4806.

[7]. Chowdhury, MAZ, Oehlschlaeger, MA. 2025. Artificial intelligence in gas sensing: a review. *ACS Sensors*; 10(3): 1538-1563.

[8]. Yan, J, Guo, X, Duan, S, Jia, P, Wang, L, Peng, C, Zhang, S. 2015. Electronic nose feature extraction methods: a review. *Sensors*; 15(11): 27804-27831.

[9]. Sun, H, Tian, F, Liang, Z, et al. 2017. Sensor array optimization of electronic nose for detection of bacteria in wound infection. *IEEE Transactions on Industrial Electronics*; 64(9): 7350-7358.

[10]. Liu, B, Xie, C, Wang, Y, et al. 2021. Lung cancer detection via breath by electronic nose enhanced with a sparse group feature selection approach. *Sensors and Actuators B: Chemical*; 339: 129896.

[11]. Saha, P, Ghorai, S, Tudu, B, Bandyopadhyay, R, Bhattacharyya, N. 2014. Optimization of sensor array in electronic nose by combinational feature selection method. In: *Sensing Technology: Current Status and Future Trends II*. Berlin: Springer; 189-205.

[12]. Zhang, L, Tian, FC. 2014. A new kernel discriminant analysis framework for electronic nose recognition. *Analytica Chimica Acta*; 816: 8-17.

[13]. Zhao, W, Meng, QH, Zeng, M, Qi, PF. 2017. Stacked sparse auto-encoders based electronic nose for Chinese liquors classification. *Sensors*; 17(12): 2855.

[14]. Peng, P, Zhao, X, Pan, X, Ye, W. 2018. Gas classification using deep convolutional neural networks. *Sensors*; 18(2): 157.

[15]. Qian, J, Tian, F, Luo, Y, Lu, M, Zhang, A. 2022. A novel multisensor detection system design for low concentrations of volatile organic compounds. *IEEE Transactions on Industrial Electronics*; 69(5): 5314-5324.

[16]. Zhao, L, Tian, F, Qian, J, Li, H, Wu, Z. 2023. Feature ensemble learning for sensor array data classification under low-concentration gas. *IEEE Transactions on Instrumentation and Measurement*; 72: 2507809.

[17]. Wei, G, Li, G, Zhao, J, He, A. 2019. Development of a LeNet-5 gas identification convolutional neural network structure for electronic noses. *Sensors*; 19(1): 217.

[18]. Gamboa, JCR, Da Silva, AJ, Araujo, ICS, Albarracin, ES, Duran, CM. 2021. Validation of the rapid detection approach for enhancing the electronic nose systems performance, using different deep learning models and support vector machines. *Sensors and Actuators B: Chemical*; 327: 128921.

[19]. Guo, J, Cheng, Y, Luo, D, Wong, KY, Hung, K, Li, X. 2021. ODRP: a deep learning framework for odor descriptor rating prediction using electronic nose. *IEEE Sensors Journal*; 21(13): 15012-15021.

[20]. Huang, W, Zhang, L, Gao, W, Min, F, He, J. 2021. Shallow convolutional neural networks for human activity recognition using wearable sensors. *IEEE Transactions on Instrumentation and Measurement*; 70: 1-11.

[21]. Gao, W, Zhang, L, Huang, W, Min, F, He, J, Song, A. 2021. Deep neural networks for sensor-based human activity recognition using selective kernel convolution. *IEEE Transactions on Instrumentation and Measurement*; 70: 1-13.

[22]. Tang, Y, Zhang, L, Min, F, He, J. 2023. Multiscale deep feature learning for human activity recognition using wearable sensors. *IEEE Transactions on Industrial Electronics*; 70(2): 2106-2116.

[23]. Ziyatdinov, A, Fonollosa, J, Fernández, L, Gutierrez-Gálvez, A, Marco, S, Perera, A. 2015. Bioinspired early detection through gas flow modulation in chemo-sensory systems. *Sensors and Actuators B: Chemical*; 206: 538-547.

[24]. Jing, YQ, Meng, QH, Qi, PF, Cao, ML, Zeng, M, Ma, SG. 2016. A bioinspired neural network for data processing in an electronic nose. *IEEE Transactions on Instrumentation and Measurement*; 65(10): 2369-2380.

[25]. Sun, X, Tian, F, et al. 2017. An optimized multi-classifiers ensemble learning for identification of ginsengs based on electronic nose. *Sensors and Actuators A: Physical*; 266: 135-144.

[26]. Wang, T, Li, Y, Wang, Y, et al. 2023. 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.

[27]. Sagi, O, Rokach, L. 2018. Ensemble learning: a survey. *WIREs Data Mining and Knowledge Discovery*; 8(4): e1249.

[28]. Zhou, ZH. 2012. *Ensemble Methods: Foundations and Algorithms*. Boca Raton: CRC Press.

[29]. Hansen, LK, Salamon, P. 1990. Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*; 12(10): 993-1001.

[30]. Schapire, RE. 1990. The strength of weak learnability. *Machine Learning*; 5(2): 197-227.

[31]. Huang, GB, Zhou, H, Ding, X, Zhang, R. 2012. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics*; 42(2): 513-529.

[32]. Huang, G, Huang, GB, Song, S, You, K. 2015. Trends in extreme learning machines: a review. *Neural Networks*; 61: 32-48.

[33]. Kim, CH, Choi, SH, Choi, S, Lee, S. 2025. A reproducible benchmark for gas sensor array classification: from FE-ELM to ROCKET and TS2I-CNNs. *Sensors*; 25(20): 6270.

[34]. Luo, W, Dai, F, Liu, Y, Wang, X, Li, M, et al. 2025. Pulse-driven MEMS gas sensor combined with machine learning for selective gas identification. *Microsystems & Nanoengineering*; 11: 72.

[35]. Li, M, Wu, C, Wang, Z, Wu, Z, Huang, W, Chen, J, et al. 2026. Enhanced gas classification in electronic nose systems using an SMOTE-augmented machine learning framework. *Sensors*; 26(2): 714.

[36]. Wörner, J, Eimler, J, Pein-Hackelbusch, M. 2025. Long-term drift behavior in metal oxide gas sensor arrays: a one-year dataset from an electronic nose. *Scientific Data*; 12: 1628.