JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online) : 2454 -7190 Vol.-21, No.-07, July (2026) pp 211-225 ISSN (Print) 0973-8975

TIME -RESOLVED MACHINE LEARNING FRAMEWORK FOR VOC DETECTION USING ZNO NANOSTRUCTURED SENSORS

D. Paramita[1], R. Sunipa[2], B. Utpal[3]

[1]Department of Electronics and Communication Engineering, Netaji Subhash Engineering College, West Bengal, India-700152.

[2] Krishnagar Women's College, Krishnagar, West Bengal, India-741101.

[3]Department of Computer Science and Engineering, University of Kalyani, Nadia, India-799046.

Email: [1]paramita.bhattacharya@nsec.ac.in, [2]suniparoy4@gmail.com, [3]utpal0172@gmail.com,

Corresponding Author: **D. Paramita**

https://doi.org/10.26782/jmcms.2026.07.00013

(Received: April 27, 2026; Revised: July 01, 2026; Accepted: July 13, 2026)

Abstract

The presence of volatile organic compounds (VOCs) is generally understood through their effects on the environment and human health. This paper presents a ZnO nano-foam-like thin film sensor to detect acetone and formaldehyde at 190- 2020 ppm in controlled laboratory settings. A quasi-steady time-resolved machine learning methodology is used to overcome the poor selectivity of ZnO-based sensors. A short portion of the sensor response is used to extract features instead of just the steady-state values, giving a more accurate representation of the dynamic sensing behavior. Various machine learning algorithms are evaluated for VOC classification. Among these, Artificial Neural Networks (ANN) have the best accuracy of 98%, whereas Gradient Boosting performs more balanced in terms of precision and recall. The findings indicate that time-resolved features could be added to enhance the classification performance without the necessity to use sophisticated sensor arrays. This paper demonstrates the integration of nanostructured sensing materials with machine learning for VOC classification using experimentally acquired sensor data. The experimentally acquired sensor data was used to validate the proposed framework under controlled laboratory conditions, demonstrating its feasibility for VOC classification.

Keywords: ZnO, Formaldehyde, Acetone, Supervised Machine Learning, Gradient Boost, ANN.

I. Introduction

Sensitive detection of volatile organic compounds (VOCs) is critical for environmental monitoring, indoor air assessment, and the diagnosis of early diseases. [III][IV][XV]. Metal oxide gas sensors, especially the sensors based on zinc oxide (ZnO), are popular as they are very sensitive, inexpensive, and easy to manufacture. These sensors have been found to have good responses to different gases such as H_2, CO_2, SO_2, as well as the VOCs [VI], [XVI]. Nevertheless, their low selectivity, particularly when chemically similar VOCs are present, is a major challenge.

The use of machine learning (ML) methods has increasingly been utilized to boost the performance of gas sensing by making it possible to extract and classify features. The bulk of current research uses steady-state response values as input characteristics [VIII] that are easy to analyze, but do not reflect the dynamic nature of sensor signals. To enhance selectivity, several works have employed sensor arrays and electronic nose (e-nose) systems with ML algorithms that include k-nearest neighbors (kNN), random forest (RF), and PCA-based approaches [I], [IX], [X], [XI], [XIII], [XVII]. Other more recent systems have involved more sophisticated sensing devices, including graphene-based field-effect transistor (GFET) and ensemble learning algorithms, including Gradient Boosting and XGBoost, among others, which have shown better classification results [II], [V], [VII], [VIII], [XI], [XIV]. Table 1 provides a comparative overview of existing strategies.

Table 1: A COMPARISON OF REPORTED EXTRACTED FEATURES AND OUTPUT ACCURACIES FOR SENSOR

Sensor Type	Objective	Features	Dataset Size	Data Processing	Model	Accuracy (%)	Ref
Resistive Gas Sensors	VOC Detection	NM	30 measurements	PCA, LDA	XGBoost	88.6	[XIII]
Resistive Gas Sensors	Harmful Gas Detection	dOT, Sensitivity	Application-specific dataset	PCA	GBM/RF	96	[XIV]
Resistive Metal Oxide Sensors	VOC Detection	NM	300 measurements	PCA (TV=75.5%)	LDA	76	[VIII]
SnO_2 Sensors	VOC Detection	dOT	64 measurements	NM	RF	85.93	[XII]
Commercial VOC Sensors	VOC Discrimination	Normalized Response	Multiple measurements under varying conditions	PCA (TV=82%)	LDA	78	[IX]
CNT Sensors	Food Classification	CWT, FFT	12 cycles/food sample	PCA	cf-RF	91	[X]

GFET Sensor	Gas Detection	Carrier Mobility	100 measurements/chemical	NA	ANN	98	[VII]
ZnO Sensor (Proposed)	VOC Detection	Sensitivity, dOT	324	PCA	ANN	98	Present

As can be seen in Table 1, most of the existing approaches use multi-sensor arrays or application-specific datasets, which adds complexity and decreases scalability to the system. Moreover, the majority of methods assume steady-state or sparse feature presentation, which are inadequate to describe the temporal dynamics of sensor responses. This indicates one important knowledge gap of the current approaches, either relying on complicated hardware setups or not exploiting the dynamic nature of sensor signals in a realistic experimental setup.

To overcome this, the time-resolved machine learning framework is proposed in this work that employs only one ZnO sensor. This paper focuses on the classification of acetone and formaldehyde in the 190-2020 ppm concentration range. The feature extraction is done using a short quasi-steady time-resolved segment of sensor response, with a 10s window being demonstrated to hold enough information to accurately classify VOCs. All experimental data were acquired under controlled laboratory conditions to ensure consistent sensor measurements for the present study.

The most important contribution of this work is the creation of a machine learning-based system of VOC detection with a single ZnO sensor and a completely experimental dataset, which simplifies the system in comparison with traditional methods based on the use of multiple sensors. In this work, a quasi-steady time-resolved feature extraction strategy based on a 10 s response window is employed to capture both transient and near-stabilized characteristics of the ZnO sensor response. The study investigates the effectiveness of these temporal features for VOC classification using a single sensing element. A direct quantitative comparison with a purely steady-state feature representation was beyond the scope of the present investigation and is identified as an important direction for future study.

II. VOC Sensing Experimental Setup

The experimental framework consists of two main stages, as shown in Fig. 1. In the first stage, a ZnO-based sensor was fabricated using a sol–gel method for detecting VOCs such as formaldehyde and acetone. The sensor response was analyzed in terms of concentration, resistivity, and response magnitude. In the second stage, the acquired sensor data were processed using machine learning techniques. The workflow includes data acquisition, preprocessing, feature extraction from the quasi-steady time-resolved response, and classification using multiple models.

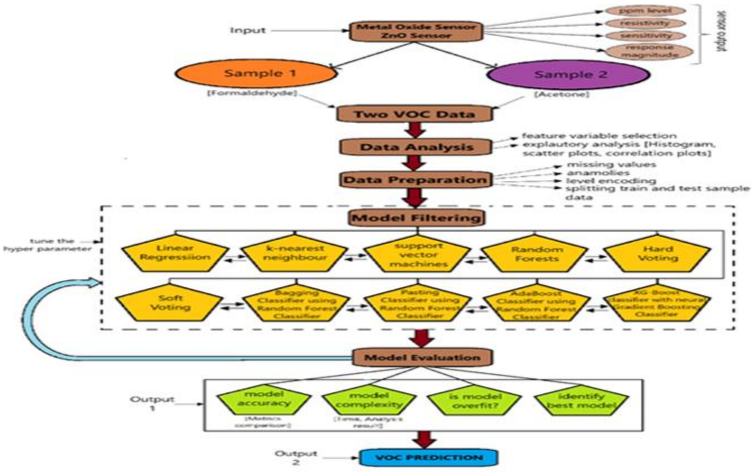


Fig 1. Experimental Data flow diagram of the model.

Sample Preparation and Gas Measurement Set-up

Details of the vapor measurement setup have been previously documented [2,3].

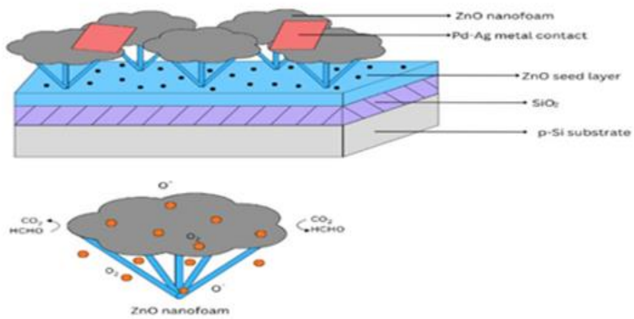


Fig. 2. Schematic diagram of the VOC Sensor.

Sensing results

The sensing performance of the ZnO sensor was evaluated by analyzing its resistance variation in air and in the presence of acetone and formaldehyde vapors. The response magnitude (RM) was calculated as:

$$\boldsymbol{RM} = \frac{Ra-Rg}{Ra} \, X \, 100\% \tag{1}$$

where Ra and Rg represent the sensor resistance in air and target gas, respectively. The response time is defined as the time required to reach 90% of the maximum response, while the recovery time corresponds to the time required to return to 90% of the baseline value. Before gas sensing, the I–V characteristics of the Pd–Ag contacts on ZnO nanofoam were examined and confirmed to exhibit ohmic behavior.

Fig. 3 shows how the magnitude of response, response time, and recovery time change with the gas concentration. In the case of formaldehyde (Fig. 3a), the magnitude of the response rises with concentration, peaking at 94.8% at 2020 ppm, with a response time that decreases as the concentration rises, from 45 s (190 ppm) to 20 s (2020 ppm). The time of recovery follows an upward trend, which implies that the desorption is slower at higher concentrations. In the case of acetone (Fig. 3b), the same tendency is observed, as the magnitude of responses increases with the concentration, and the response time decreases, and the recovery time increases. These findings suggest that an increase in the concentration of gases will improve the sensor response because more gas molecules will interact with the ZnO surface. Generally, the sensor has shown stable and reliable behavior throughout the concentration range of 190- 2020 ppm at an operating temperature of 100^0C, and has little baseline drift, which indicates good repeatability and reliability.

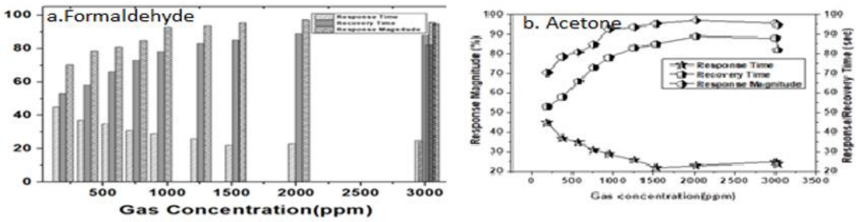


Fig. 3. Comparison charts of the magnitude of responses, response time, and recovery time of the sensor to the various concentrations of VOCs.

Sensor Data Extraction

The ZnO sensor shows a rise in electrical current when exposed to both VOCs, with a relatively high response to formaldehyde compared to acetone. Sensitivity (S) is determined as:

$$S = \frac{1}{Cv} X \frac{\Delta I}{IO} \quad (2)$$

with Cv being the VOC concentration, ΔI being the change in current, and I0 being the baseline current. Because of the low inherent selectivity, it is difficult to differentiate between VOCs with the traditional tools, which is why machine learning classification is encouraged. Before extracting features, there was baseline correction to enhance the reliability of the data with the help of:

$$YS = R(t) - R(0) \quad (3)$$

R(t) is the dynamic response, and R (0) is the baseline. In order to extract the features, a quasi-steady time-resolved method is used. A small section of the sensor response of about 10s is used instead of a single steady-state value, and the signal is sampled after it has attained about 90 percent of the maximum value. This portion records both short-lived and near-steady states of the response, which contains higher discriminatory data to classify VOC. Consequently, the suggested technique enhances the performance of the classifier because it considers the behavior of temporal responses.

Dataset and Data Classification

All the experiments were performed in Python (v3.10) in a Jupyter Notebook on a Windows 10 (x86) system with an Intel Core i5 processor and 8 GB RAM. Scikit-learn (v1.1.2) and Keras (v2.10.0) were used to implement machine learning models.

The raw data comprises 324 samples (formaldehyde: 150 samples and acetone: 174 samples) under controlled laboratory experiments of the developed ZnO-based sensor at a concentration of 190-2020 ppm. The dataset provides consistent sensor measurements obtained under controlled laboratory conditions for evaluating the proposed framework. The data were split into 70% training, 15% validation, and 15% testing data. The training data were cross-validated 5 times to enhance robustness and minimize overfitting.

To classify the data, a quasi-steady time-resolved 10s response segment was picked once the signal had attained about 90 percent of its maximum value. This segment records both transient and near-steady-state properties of the sensor response and provides temporal information that is used for machine-learning-based VOC classification. A variety of machine learning models were tested under the same conditions. Artificial Neural Networks (ANN) and Gradient Boosting were the most accurate, with 98 and 97.5%, respectively.

III. Experimental Results and Discussion

This research aims to identify volatile organic compounds (VOCs), namely acetone and formaldehyde, with a ZnO-based gas sensor in combination with machine learning algorithms. Data on sensor measurements obtained experimentally were used to train and test several classification models.

Data Preprocessing

Preprocessing of sensor data was done to enhance the quality of the data and model performance. Baseline correction, elimination of inconsistencies, and scaling of features were done as part of preprocessing.

Model Evaluation

The performance assessment of models was carried out through a systematic validation method. A 5-fold cross-validation plan was used to make sure that it is robust and generalizability. Accuracy, precision, recall, and F1-score were some of the evaluation metrics. Moreover, the performance of classification was evaluated by means of confusion matrices, ROC curves, and AUC scores.

Analysis of Results

- ***Gas Type Classification***

ROC and precision-recall (PR) curves were used to assess the classification performance of the various machine learning models, as demonstrated in Figs. 4 and 5.

Fig. 4 shows the ROC curves of various machine learning models. The strong classification capability is evident in most of the models as they are close to the top-left area. Among them, Artificial Neural Networks (ANN) and Gradient Boosting are the most successful, with AUC values that are close to unity. The other ensemble models are also competitive, including Random Forest and XGBoost, and the performance of the Decision Tree and KNN is relatively lower.

More elaboration on model performance can be seen in the precision-recall curves in Fig. 5. ANN has the best classification rate (98%), then Gradient Boosting (97.5%). Nevertheless, Gradient Boosting has a more stable balance between precision and recall at various thresholds. Random Forest is also stable and less reliable at higher recall, but Decision Tree and KNN are less reliable.

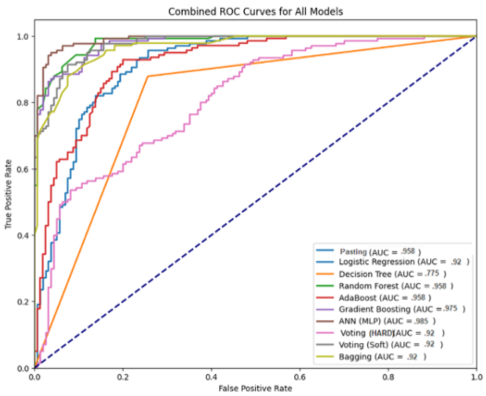


Fig. 4. Receiver Operating Characteristic (ROC) curve for different ML models.

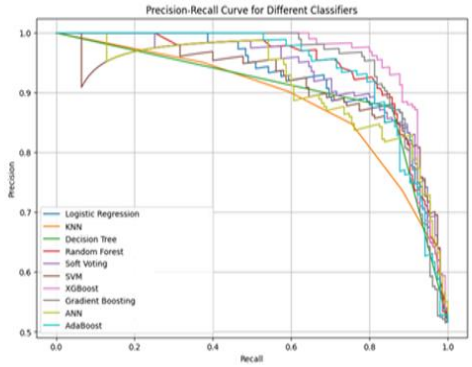


Fig. 5. Precision-Recall (PR) curve for different ML models.

On the whole, the findings suggest that ANN has the best classification accuracy, but ensemble-based models (Gradient Boosting and Random Forest) have a more balanced and reliable performance. The observed classification performance indicates that the proposed quasi-steady time-resolved extraction method effectively captures the dynamic characteristics of the sensor response. The extracted temporal features provide useful information for VOC classification under the investigated experimental conditions.

- **Model Performance Comparison:**

Various machine learning models were tested on the classification of VOCs, and their results are summarized in Table 2.

The training data confusion matrix presented in Fig. 6 reveals that most of the models have a high classification accuracy, and their predictions are clumped around the diagonal, showing that they made the right classification. Equally, the testing data confusion matrix in Fig. 7 affirms reliable model performance with low misclassification, indicative of good generalization performance with unknown samples.

Fig. 6. Confusion matrix for training data for ML models: KNN, Decision Tree, AdaBoost, Random Forest, Logistic Regression, Hard Voting, Soft Voting, XGBoost, Gradient Boosting, ANN.

The results indicate that the Artificial Neural Network (ANN) has the best test accuracy of 0.985. Nevertheless, with a rather lower recall (0.762), it results in an intermediate F1-score, which is a trade-off between the overall accuracy and sensitivity. By contrast, the Gradient Boosting classifier shows a more balanced performance, with high precision, recall, and F1-score consistently, indicating its generalization capabilities based on the time-resolved sensor features. Ensemble models like the Random Forest and the AdaBoost also have stable and competitive performance because of their ability to reduce variance. The best results are moderately high with voting and bagging-based methods, whereas simpler models like the Decision Tree and KNN have relatively lower accuracy, mostly because of overfitting and incapacity to represent the interactions between complex features.

Fig. 7. Testing data Confusion matrix in ML models: KNN, Decision Tree, AdaBoost, Random Forest, Logistic Regression, Hard Voting, Soft Voting, XGBoost, Gradient Boosting, ANN.

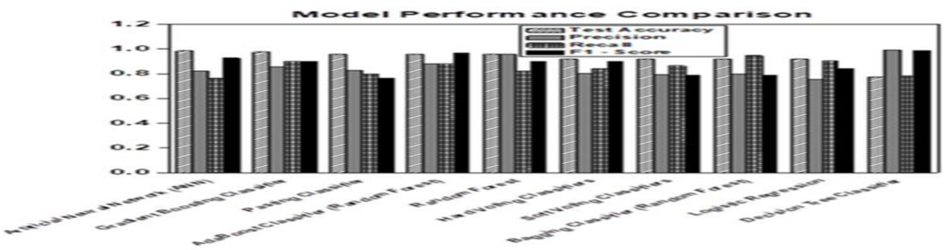


Fig. 8. Comparison chart for different ML models used in this experiment.

Fig. 8 gives a general performance of the models. Although ANN has a higher predictive accuracy, Gradient Boosting can be more predictable and reliable in all aspects of the evaluation, thus demonstrating consistent performance for the investigated VOC classification task.

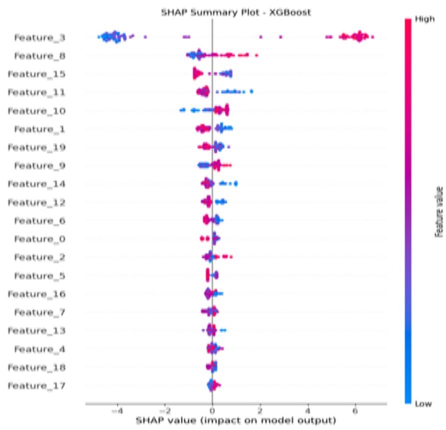
Table 2 is a summary of the performance of various machine learning models. ANN has the best classification accuracy, and Gradient Boosting has a more balanced performance in terms of precision, recall, and F1-score. Ensemble techniques like Random Forest or AdaBoost are also stable, but simple models have a lower level of accuracy.

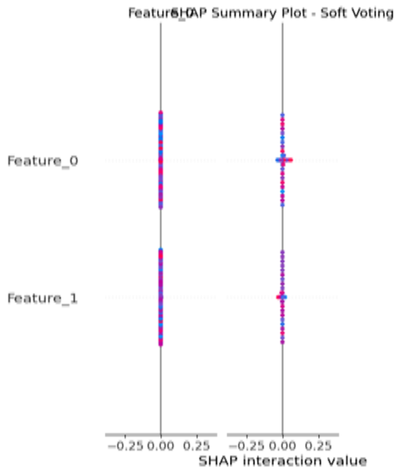
Table 2: A COMPARISON OF REPORTED EXTRACTED FEATURES AND OUTPUT ACCURACIES FOR SENSOR

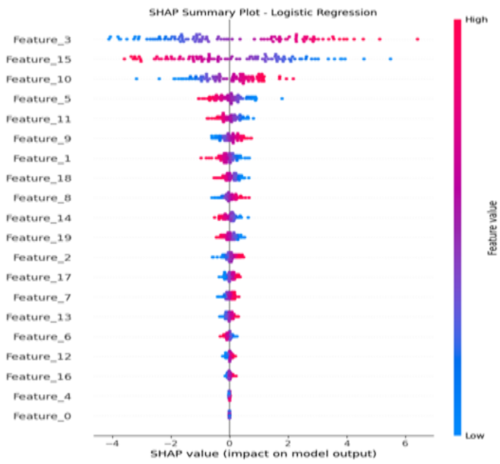
Model	Test Accuracy	Precision	Recall	F1-Score	Strengths	Weaknesses	Reason
Artificial Neural Network (ANN)	0.985	0.823	0.762	0.791	High accuracy, moderate F1-score	Recall is lower than some tree models	Captures complex relationships but may overfit
Gradient Boosting Classifier	0.975	0.858	0.898	0.878	Strong balance of precision & recall	Training time is high	Ensemble learning improves generalization
Pasting Classifier (Random Forest)	0.958	0.826	0.800	0.813	Good accuracy, robust	Moderate performance compared to boosting models	Aggregates multiple trees without boosting
AdaBoost Classifier (Random Forest)	0.958	0.881	0.879	0.880	High F1-Score, Balanced precision and recall	Sensitive to noisy data	Adjusts weights to improve weak classifiers
Random Forest	0.958	0.958	0.823	0.885	High accuracy, feature importance	Computationally expensive	An ensemble of decision trees improves generalization
Hard Voting Classifiers	0.92	0.803	0.842	0.822	Works well on varied models	Requires diverse classifiers	Uses majority vote for classification
Soft Voting Classifiers	0.92	0.795	0.864	0.828	Uses probability-weighted voting	Slightly lower performance compared to ensemble boosting methods	Weighs outputs from different models
Bagging Classifier (Random Forest)	0.92	0.796	0.946	0.865	Strong recall, reduces variance	Lower precision	Uses bootstrap aggregation for stability

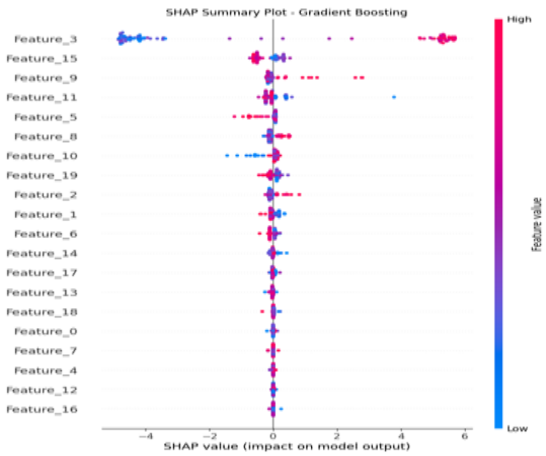
Logistic Regression	0.92	0.755	0.903	0.822	Interpretable, works well on linear data	Lower precision than tree models	Works well when relationships are linear
Decision Tree Classifier	0.775	0.992	0.785	0.878	High precision	Poor generalization	Overfits easily due to high variance

- ***SHAP Analysis:***









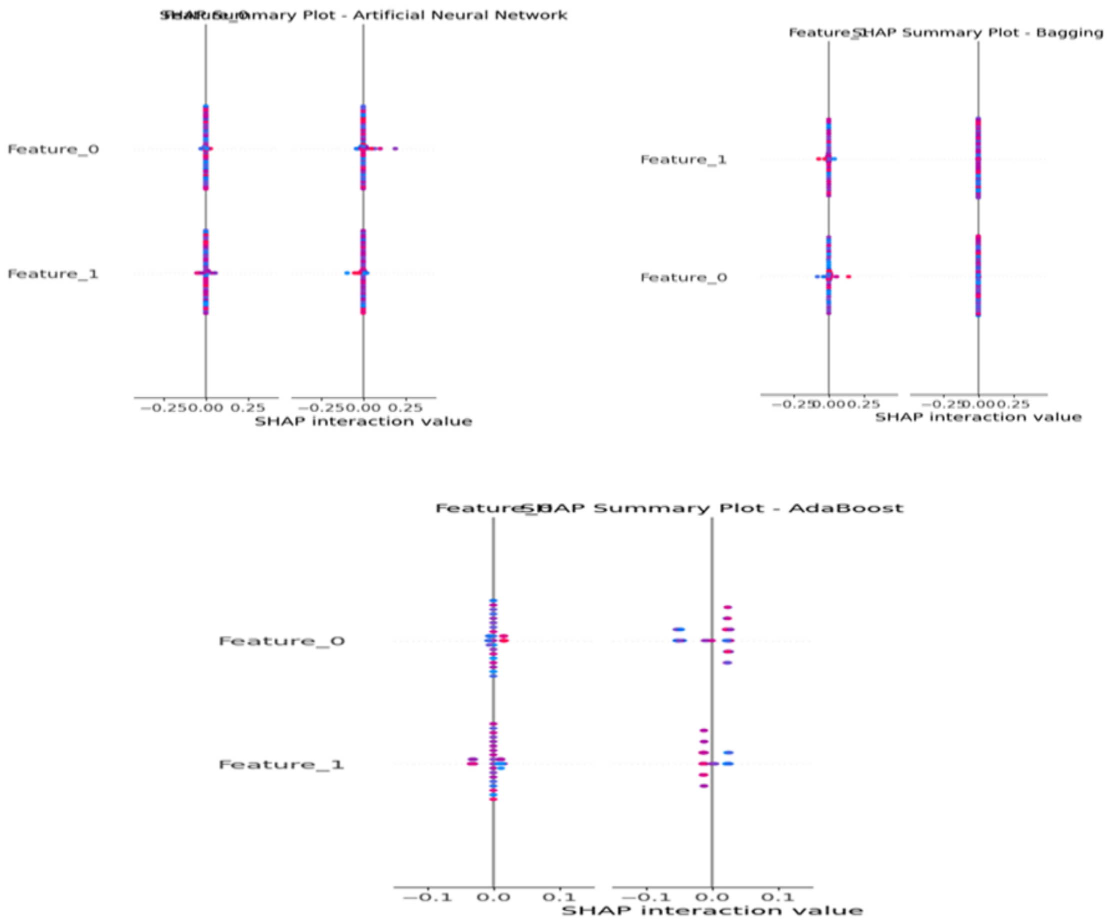


Fig. 9. SHAP analysis of different ML models.

SHAP (Shapley Additive explanations) analysis was conducted to explain the role of time-resolved characteristics in VOC classification. The SHAP summary figures in Fig. 9 reveal that a few major characteristics of the chosen 10s response window are influential in the model predictions.

The distribution of SHAP values shows that different features contribute differently to the prediction of acetone and formaldehyde. This behaviour agrees with the transient response of the ZnO sensor during gas exposure, where adsorption and desorption of gas molecules influence the sensor signal. Although SHAP explains the contribution of individual features to the model output, it does not by itself establish the underlying sensing mechanism.

Model-wise analysis shows that ensemble methods, especially Gradient Boosting, make use of more diverse contributions of features, which means that they are capable of modeling a complex nonlinear relationship. Decision Tree, on the other hand, is based on a small number of features and therefore has less generalization. Although ANN models are less straightforward to interpret, the SHAP results help identify the features that have the greatest influence on the classification outcome, making the model behaviour easier to understand.

Overall, the SHAP analysis supports the importance of the extracted time-resolved features in the classification process and provides a clearer understanding of how the models make their predictions.

V. Conclusion

This paper proposes a machine learning-based system to classify VOCs with the help of a resistive sensor based on ZnO. To measure the dynamic nature of the sensor signal, a quasi-steady time-resolved response segment (10s) is used, which allows distinguishing acetone and formaldehyde.

The top-rated models are Artificial Neural Networks (ANN) with the highest classification accuracy (98%), whereas Gradient Boosting offers a more balanced model performance based on evaluation measures. These findings demonstrate that the proposed quasi-steady time-resolved feature extraction strategy provides an effective representation of the experimentally acquired ZnO sensor response for VOC classification. The present results establish the feasibility of the proposed framework under the experimental conditions investigated. The suggested structure provides a scalable and affordable approach to VOC detection without using complicated sensor arrays.

Future work will extend the present study to additional VOCs and mixed-gas conditions. The proposed framework will also be examined under different environmental conditions to assess its performance in practical applications.

Conflict of Interest:

The authors declare that there is no relevant conflict of interest regarding this paper.

References

I. Acharyya, S., K. Mondal, and S. Basu. "Single Resistive Sensor for Selective Detection of Multiple VOCs Employing SnO_2 Hollow Spheres and Machine Learning Algorithm: A Proof of Concept." *Sensors and Actuators B: Chemical*, vol. 321, 2020, p. 128484. 10.1016/j.snb.2020.128484.

II. Banerjee, N., B. Bhowmik, S. Roy, C. K. Sarkar, and P. Bhattacharyya. "Anomalous Recovery Characteristics of Pd Modified ZnO Nanorod Based Acetone Sensor." *Journal of Nanoscience and Nanotechnology*, vol. 13, no. 10, 2013, pp. 6826–6834. 10.1166/jnn.2013.7786.

III. Banerjee, N., S. Roy, C. K. Sarkar, and P. Bhattacharyya. "High Dynamic Range Methanol Sensor Based on Aligned ZnO Nanorods." *IEEE Sensors Journal*, vol. 13, no. 5, 2013, pp. 1669–1676. 10.1109/JSEN.2013.2237822.

IV. Binson, V. A., N. A. A. Rahim, and R. Shahriman. "Non-Invasive Diagnosis of COPD with E-Nose Using XGBoost Algorithm." *Proceedings of the 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems*. IEEE, 2021, pp. 297–301. 10.1109/ACCESS51619.2021.9563303.

V. Dalis, C. "Volatile Organic Compound Assessment as a Screening Tool for Early Detection of Gastrointestinal Diseases." *Microorganisms*, vol. 11, no. 7, 2023, p. 1822. https://www.mdpi.com/2076-2607/11/7/1822.

VI. Fan, X., et al. "Exhaled VOC Detection in Lung Cancer Screening: A Comprehensive Meta-Analysis." *BMC Cancer*, vol. 24, no. 1, 2024, p. 775. 10.1186/s12885-024-12537-7.

VII. Ghosh, R., and A. Talukdar. "Lung Disease Classification from Chest X-Ray Images Using Ensemble Learning." *2024 IEEE Calcutta Conference (CALCON)*, Kolkata, India, 2024, pp. 1–5. 10.1109/CALCON63337.2024.10914228.

VIII. Hayasaka, T., et al. "An Electronic Nose Using a Single Graphene FET and Machine Learning for Water, Methanol, and Ethanol." *Microsystems & Nanoengineering*, vol. 6, no. 1, 2020, pp. 1–9. 10.1038/s41378-020-0161-3.

IX. Itoh, T., Y. Koyama, W. Shin, T. Akamatsu, A. Tsuruta, Y. Masuda, and K. Uchiyama. "Selective Detection of Target Volatile Organic Compounds in Contaminated Air Using Sensor Array with Machine Learning: Aging Notes and Mold Smells in Simulated Automobile Interior Contaminant Gases." *Sensors*, vol. 20, no. 9, 2020, p. 2687. 10.3390/s20092687.

X. Jaeschke, C., M. Klein, R. Götz, C. Schütze, and S. Zimmermann. "An Innovative Modular eNose System Based on a Unique Combination of Analog and Digital Metal Oxide Sensors." *ACS Sensors*, vol. 4, 2019, pp. 2277–2281. 10.1021/acssensors.9b01244.

XI. Mei, H., J. Peng, T. Wang, et al. "Overcoming the Limits of Cross-Sensitivity: Pattern Recognition Methods for Chemiresistive Gas Sensor Array." *Nano-Micro Letters*, vol. 16, 2024, p. 269. 10.1007/s40820-024-01489-z.

XII. Singh, S., S. S., P. Gajje, C. Adak, R. P. Shukla, and V. Kamble. "Metal Oxide-Based Gas Sensor Array for the VOCs Analysis in Complex Mixtures Using Machine Learning." *Sensors*, vol. 23, no. 8, 2023, p. 1703. 10.3390/s23081703.

XIII. Venkatesh, C., K. Ramana, S. Y. Lakkisetty, S. S. Band, S. Agarwal, and A. Mosavi. "A Neural Network and Optimization Based Lung Cancer Detection System in CT Images." *Frontiers in Public Health*, vol. 10, 2022, article 769692. 10.3389/fpubh.2022.769692.

XIV. Wang, H., S. Sun, J. Wang, and C. Zhao. "EEMD and GUCNN-XGBoost Joint Recognition Algorithm for Detection of Precursor Chemicals Based on Semiconductor Gas Sensor." *IEEE Transactions on Instrumentation and Measurement*, vol. 71, 2022, pp. 1–12. 10.1109/TIM.2022.3197762.

XV. Wilson, A. D. "Developments of Recent Applications for Early Diagnosis of Diseases Using Electronic-Nose and Other VOC-Detection Devices." *Sensors*, vol. 23, no. 18, 2023, p. 7885. https://www.mdpi.com/1424-8220/23/18/7885.

XVI. Wilson, A. D., and L. B. Forse. "Potential for Early Noninvasive COVID-19 Detection Using Electronic-Nose Technologies and Disease-Specific VOC Metabolic Biomarkers." *Sensors*, vol. 23, no. 6, 2023, p. 2887. 10.3390/s23062887.

XVII. Zhang, T., Y. Cao, M. Chen, and L. Xie. "Recent Advances in CNTs-Based Sensors for Detecting the Quality and Safety of Food and Agro-Products." *Journal of Food Measurement and Characterization*, vol. 17, no. 5, 2023, pp. 3061–3075. 10.1007/s11694-023-01850-7.