ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Machine Learning Approaches: Detecting the Disease Variants in Human-Exhaled Breath Biomarkers

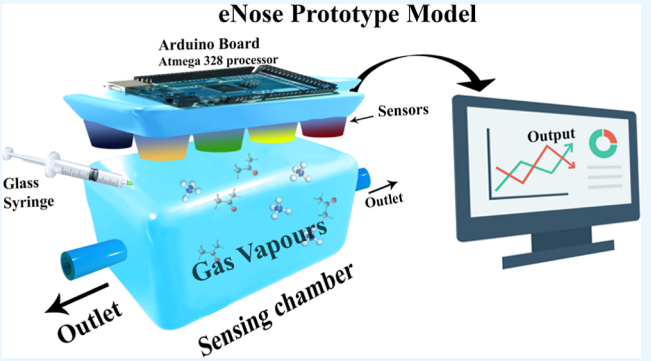
Bhuvaneswari Selvaraj, Elakkiya Rajasekar, and John Bosco Balaguru Rayappan*

Cite This: *ACS Omega* 2024, 9, 215−226

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: In recent days, the development of sensor-based medical devices has been found to be very effective for the prediction and analysis of the onset of diseases. The instigation of an electronic nose (eNose) device is profound and very useful in diverse applications. The analysis of exhaled breath biomarkers using eNose sensors has attained wider attention among researchers, and the prediction of multiple disease variants using the same is still an open research problem. In this work, an enhanced XGBooster classifier-based prediction mechanism was introduced to identify the disease variants based on the responses of commercially available metal oxide-based Figaro (Japan) sensors including TGS826, TGS822, TGS2600, and TGS2602. The implemented model secured 98.36% prediction accuracy in multiclass disease prediction and classification. The homemade one-dimensional metal oxide sensing elements such as ZnO, Cr-doped ZnO, and ZnO/NiO were integrated with the aforementioned sensor array for the specific detection of the three biomarkers of interest. This model has attained a classification accuracy of 99.77, 94.91, and 96.56% toward ammonia, ethanol, and acetone, respectively. And the multiclass disease biomarker classification accuracy of the readymade and homemade eNose prototype models was compared, and the results are summarized.



1. INTRODUCTION

In recent times, disease diagnostic techniques, especially hand-held electronic devices like the electronic nose (eNose), have emerged as an effective noninvasive tool to diagnose the onset of various diseases.[1,2] The development of such devices assists medical practitioners in narrowing down the gap in the prediction of diseases and, in turn, provides focused treatment to the patients.[3] Moreover, these devices greatly reduce the cost and time involved in the prediction and analysis of diseases.[4,5] The eNose device has achieved a prominent position among such devices in predicting and analyzing the onset of diseases, such as renal disease, diabetes, vitamin deficiencies, etc., in a noninvasive manner.[2] The recent reports on eNose-based disease detection systems by analyzing the presence of biomarkers in the exhaled breath have highlighted the need for such devices for rapid diagnosis.[6]

The eNose employs an array of cross-selective gas sensors to detect the number of VOCs in exhaled breath, but it has not yet been widely implemented in clinical practice due to the lack of specificity, standardization, a short-fall on the accuracy, replication, etc.[7] At the same time, researchers have been continuously working on the development of eNose systems for asthma biomarker prediction,[8] brain cancer detection,[9] diabetes analysis,[10] cervical cancer classification,[11] smoking detection,[12] tuberculosis,[13] prediction of pulmonary diseases,[14] and so on.

In this context, the present study has focused on the detection of primary biomarkers of renal disease (ammonia), vitamin E deficiency/cancer (ethanol), and diabetes (acetone) using an eNose. According to International Diabetes Federation (IDF) report 2019, more than 463 million people suffered because of diabetes, and this number will cross 700 million in the 2040s.[15] In India, 88 million people, i.e., 1 in 5 adults, live with this chronic medical condition.[16] As per International Society of Nephrology (ISN) report 2021, 843 million people suffered due to renal-related issues.[17] Also, 2 billion people were affected by vitamin deficiencies, which are associated with major health issues such as cancer, anemia, infertility, immune impairment, and so on.[18] Hence, early detection and diagnosis of these diseases need to be refined in order to save human lives.[19]

Received: May 29, 2023
Revised: December 6, 2023
Accepted: December 11, 2023
Published: December 22, 2023

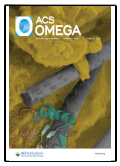


Table 1. Technical Specification of Sensors Used in the Proposed Model

s. no.	sensor code	target vapors
1	TGS 826	ammonia and amine
2	TGS 822	organic solvent and ethanol
3	TGS 2600	ethanol and methane
4	TGS 2602	ammonia, toluene, and hydrogen sulfide
5	DHT-11	temperature and relative humidity
6	ZnO	ammonia
7	Cr-doped ZnO	ethanol
8	ZnO/NiO	acetone

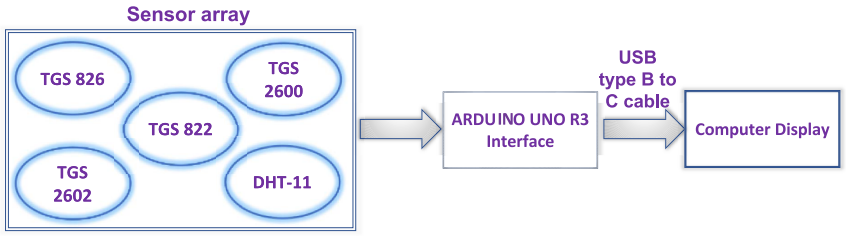


Figure 1. Schematic representation of the proposed prototype model.

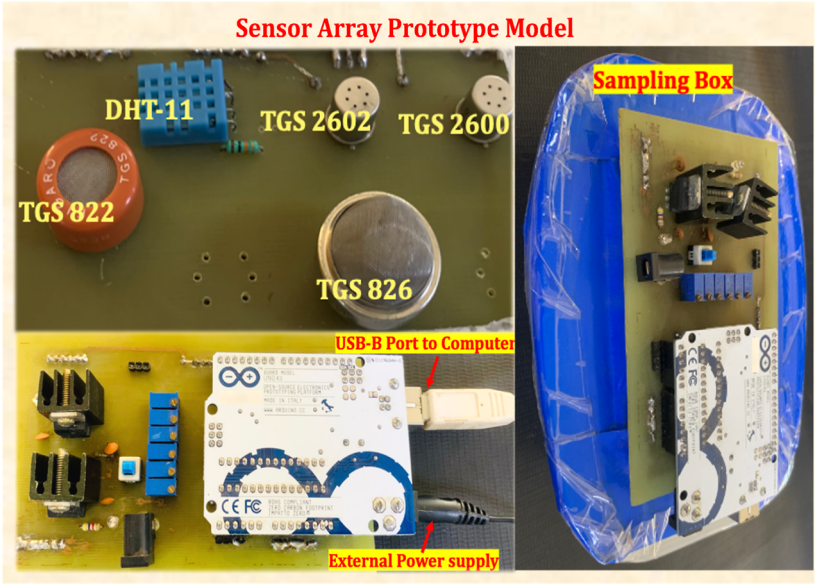


Figure 2. eNose prototype.

In this background, a custom-made eNose prototype embedded with a data analysis method has been developed for diagnosing ammonia, ethanol, and acetone biomarkers. To enhance the overall performance of the eNose device, various machine learning (ML) models have been tried. Since the development of such models poses higher complexities in handling relationships among the data points and fails to produce the expected outcome,[20,21] adapting newer advancements and fine-tuning model parameters help to optimize the model performance. In order to demonstrate the viability of the suggested sensing strategy employing metal oxide-based composite nanostructures and the ML model, the study concentrates on the detection and classification of the stated biomarkers. Hence, in this study, the complex structure of data sets has been analyzed using the enhanced extreme gradient boosting (enhanced XGBoost) classifier. In other words, in this study, the extreme gradient boosting classifier (XGBoost) model was modified based on the parameters max_depth and num_leaves (denoted as enhanced XGBoost classifier), and it was adapted to enhance the performance of the eNose prototype. The detailed analysis of various features and their relationships demands numerous mathematical techniques to explore the covariance relationship among the attributes. The implemented model has secured ~ 98% prediction accuracy in multiclass disease prediction and classification for the commercial sensors. Moreover, integration and combination of homemade and commercial sensors along with data analysis have seldom been reported. Henceforth, the homemade one-dimensional metal oxide sensing elements such as ZnO, Cr-doped ZnO, and ZnO/NiO composite nanostructures, which were reported in our previous work,[22,23] have been integrated with the aforementioned sensor array, and the multiclass disease biomarker prediction and classification accuracy of the implemented model have been observed and reported.

https://doi.org/10.1021/acsomega.3c03755
ACS Omega 2024, 9, 215−226

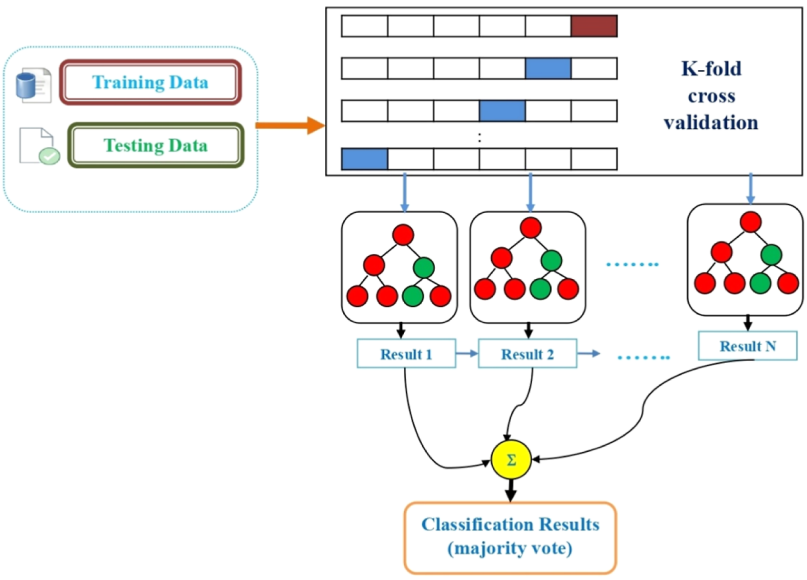


Figure 3. Enhanced XGBoost classifier architecture.

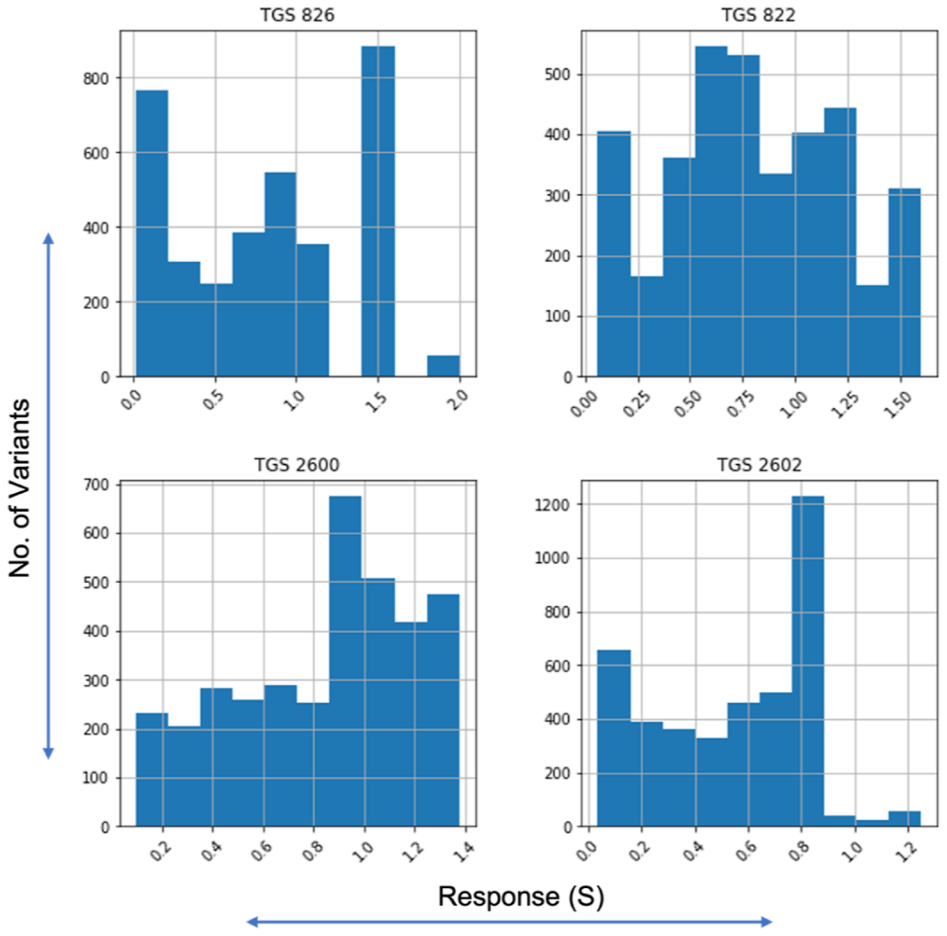


Figure 4. Data visualization of the array of readymade sensors.

2. MATERIALS AND METHODS

The eNose prototype has been designed and developed using four Figaro (Japan) sensors along with their target vapors, which are given in Table 1. In addition, the temperature and relative humidity sensor DHT-11 has been integrated with the eNose model. The schematic representation of the sensor array prototype is shown in Figure 1. In addition, three homemade sensors fabricated using ZnO, Cr-doped ZnO, and ZnO/NiO nanostructures were integrated with the aforementioned array, and the corresponding data sets have been recorded for the classification process. The data sets were stored and processed using the Atmega 328p microcontroller controller present in

https://doi.org/10.1021/acsomega.3c03755
ACS Omega 2024, 9, 215−226

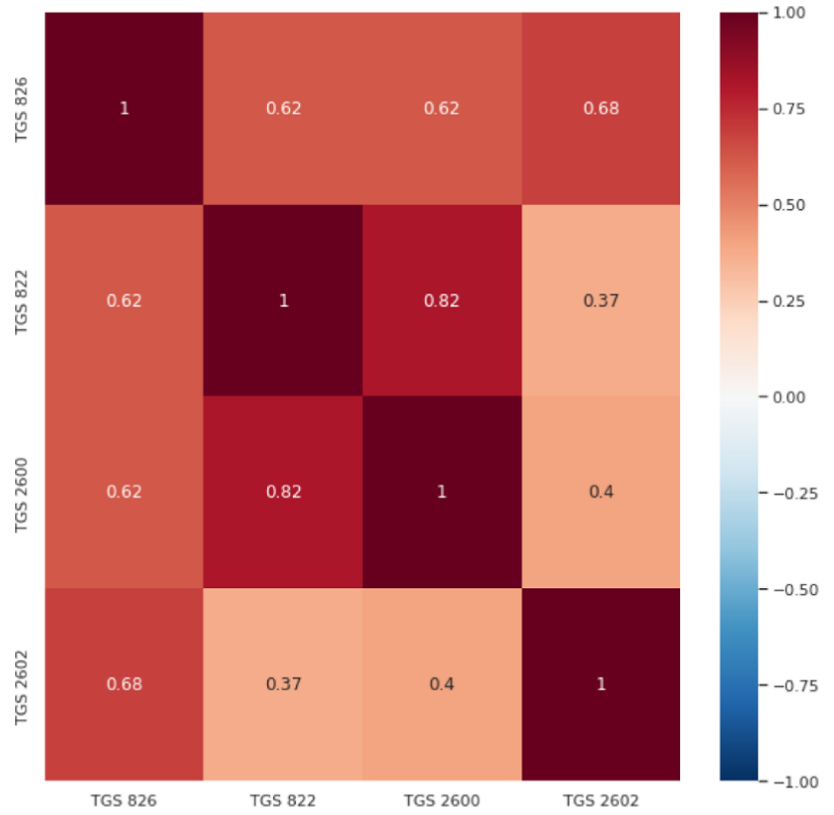


Figure 5. Feature correlations of the array of readymade sensors.

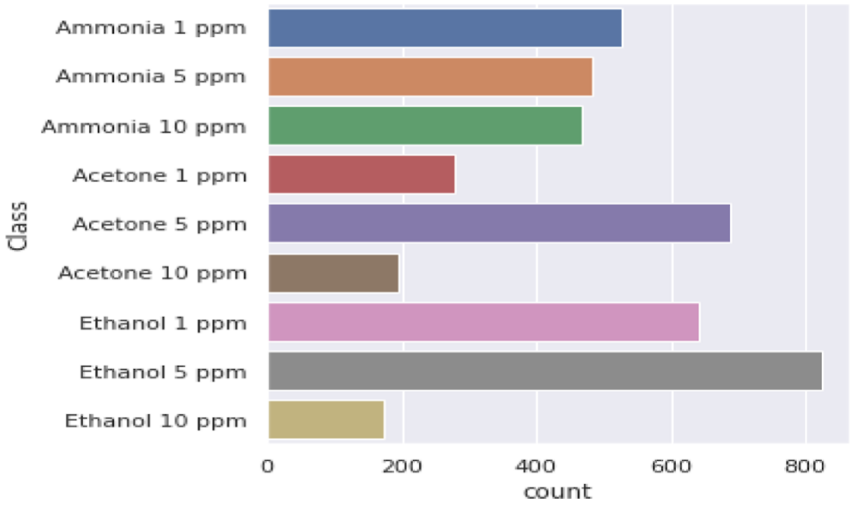


Figure 6. Class distributions of the readymade sensor data set.

the Arduino UNO R3 board. All three homemade sensors worked under room-temperature conditions (30 °C) and a relative humidity of ∼58%.

In order to train and test the sensor prototype, 1 to 10 ppm of acetone, ammonia, and ethanol vapors were introduced into the 500 mL capacity sampling box, and the headspace of the sensor array was placed for data collection. Figure 2 shows the developed eNose prototype. The primary version of the data sets has been trained by the proposed enhanced XGBoost classifier model. The streaming data values helped to identify the possibility of the occurrence of a specific disease marker. The data set consists of 42,888 entries to provide sufficient training and testing data for the proposed model development.

3. ENHANCED XGBOOST CLASSIFIER

The discrimination of disease variants from the patient's exhaled breath is a challenging task, and the earlier approaches were used for predicting diseases of particular types with a fixed set of parameters and constraints.[24−26] This work aims for multiclass classification using the enhanced XGBoost classifier and to produce better results over streaming sensor data collected from eNose. The proposed model has achieved 98% classification accuracy, which is comparatively higher than that of the earlier approaches.[24−27] The proposed enhanced XGBoost classifier utilizes the k-fold cross-validation approach for handling the various segments of training and testing data portions. Each decision tree produces various results and is

https://doi.org/10.1021/acsomega.3c03755
ACS Omega 2024, 9, 215−226

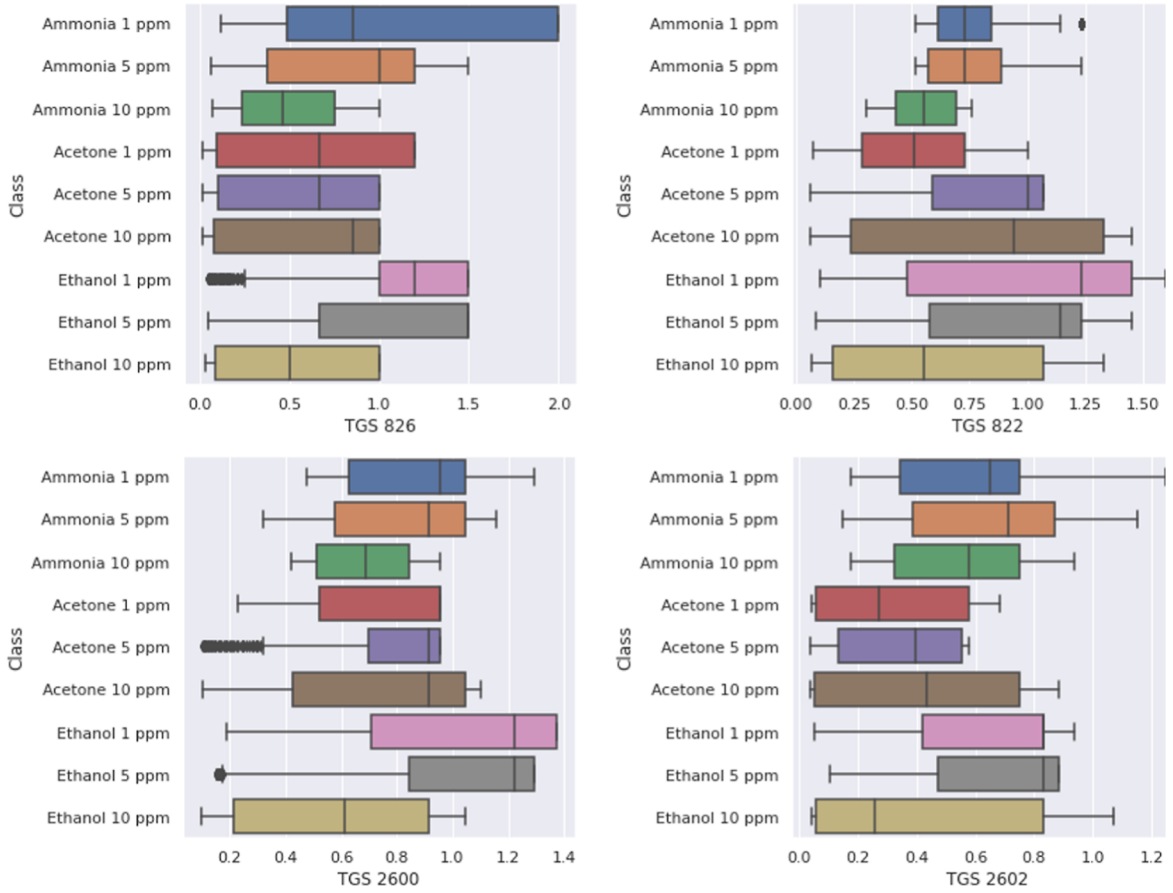


Figure 7. Box and whisker plots of the features grouped by the class.

Table 2. Sensor Arrays with Three Different Groups of Sensors

groups	readymade sensors	homemade sensors
G1	TGS 822,2600,2602 & DHT 11	ZnO
G2	TGS 826,822,2602 & DHT 11	Cr-doped ZnO (Cr−ZnO)
G3	TGS 826,822,2602 & DHT 11	NiO/ZnO

Table 3. Features of the Homemade eNose System

sensor technology	metal oxide-based chemiresistive gas sensors
measuring range	1 to 10 ppm (for the eNose system)
ML	enhanced XGBoost classifier
accuracy	99.77 (ammonia), 94.91% (ethanol), and 96.56% (acetone)
sampling	head space sampling (non-invasive)
temperature	30 °C
relative humidity (RH)	58%

compared based on a major voting scheme. The boosting mechanisms greatly improve the performance of the ML models by incorporating efficient techniques and parallel approaches. The input features for the ML were derived from the responses of metal oxide-based homemade and readymade sensors in the presence of target biomarkers, namely, ammonia, acetone, and ethanol. These input data were structured as an array of numbers. Each sample in the data set was represented by a multidimensional array, where each element of the array corresponds to a specific input feature. These sensor response data were used to train the ML model and in turn to learn patterns and relationships between the input data and the respective biomarker concentrations. A standardization approach was used to make sure that the comparisons of various characteristics are fair and accurate and also to lessen the influence of differences in feature scales. The values in the data set were transformed to a standardized range of −1 to +1 using a standard transformer.

Figure 3 demonstrates the overall architecture of the enhanced XGBoost classifier architecture.

The proposed model has used boosting, stacking, and bagging techniques, specifically combining multiple learner models, to enhance the overall performance. Additionally, the bias and variance of the model were reduced, and the predictive power was significantly improved. The following two sections narrate the process involved in implementing the enhanced XGBoost classifier model. The first section describes the classification accuracy of the eNose model with ready-made

https://doi.org/10.1021/acsomega.3c03755
ACS Omega 2024, 9, 215−226

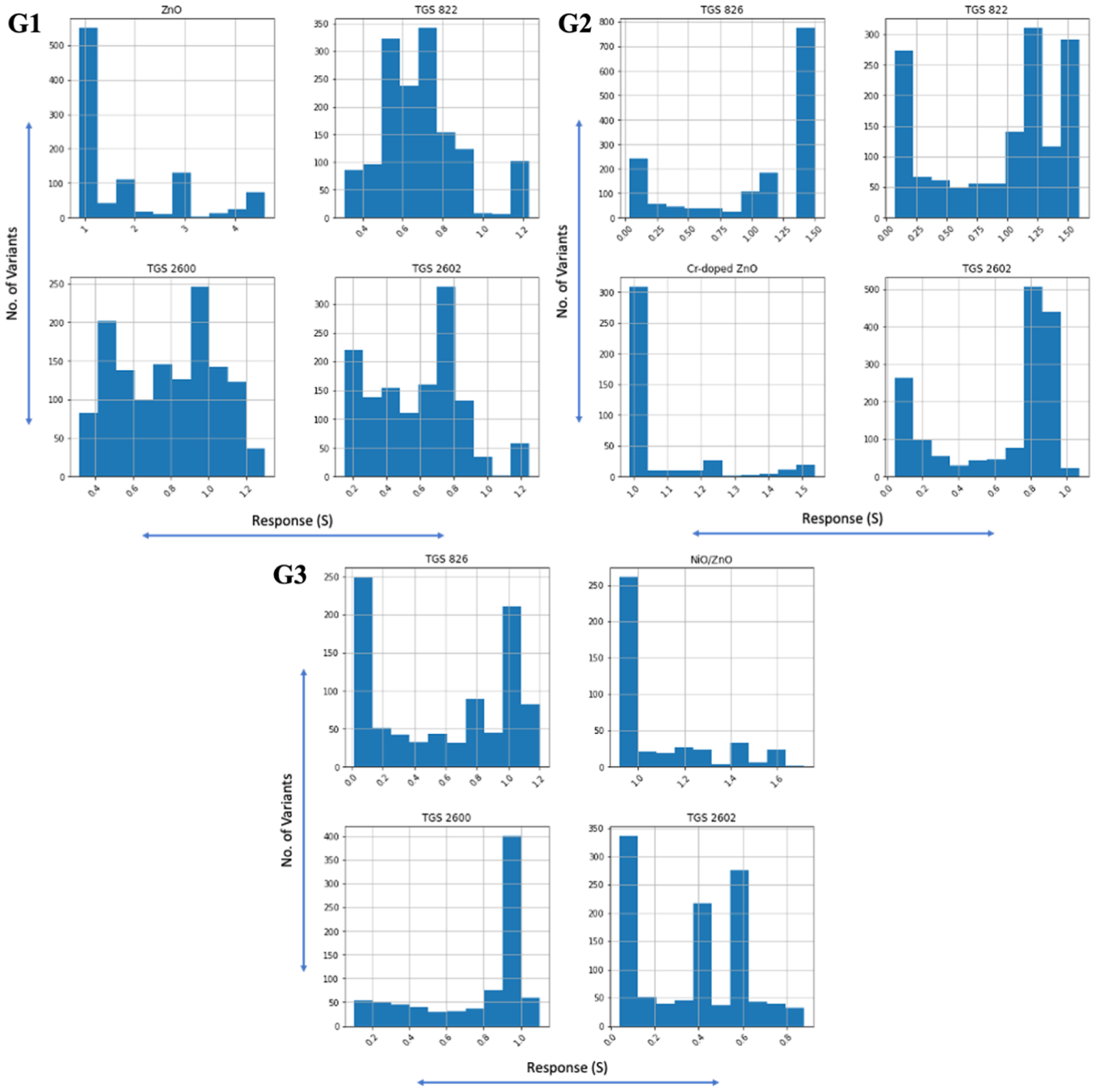


Figure 8. Feature distributions of the three eNose systems.

sensors, and the second section explains the classification accuracy of the model with homemade sensors.

4. CLASSIFICATION ANALYSIS OF THE READYMADE SENSOR ARRAY DATA SET

4.1. Data Visualization. Data visualization plays a prominent role in analyzing the data set. The bar chart is used to analyze the various dimensions of features and their relationships. This strategy has strongly supported the creation of an accurate prediction-based model, which also demonstrates the correlation of the relationships between the different variables and attributes. Figure 4 explores the distribution of the sensor array data set. The various threshold ranges of each attribute of the sensors are displayed individually to provide a comparative analysis of the data set. This plot provides additional information for the newly acquired data and produces enhanced results.

4.2. Feature Correlations. The linear dependency between two characteristic features is measured by the Pearson product-moment correlation coefficient (r) in the range of -1 to $+1$ as shown in Figure 5. The correlation matrix confirms the highest positive correlation values (0.82) between the sensors TGS 2600 and TGS 822 compared to those of the other sensors. This appears like a logical choice for a trial variable to explain the concepts of a fundamental linear regression model, and the class attribute denotes the type of class to which the particular value of features belongs to the specific data.

https://doi.org/10.1021/acsomega.3c03755
ACS Omega 2024, 9, 215−226

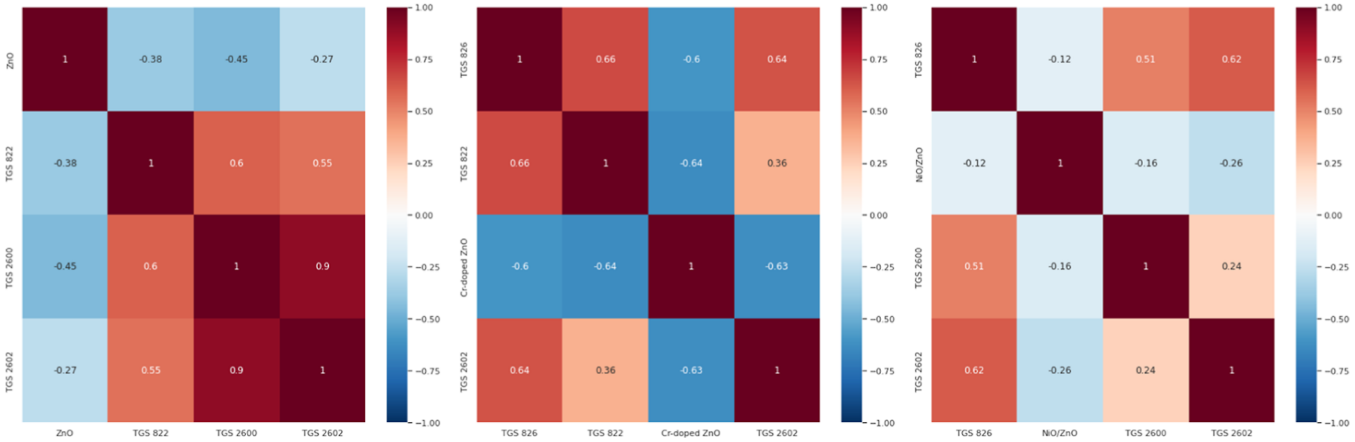


Figure 9. Feature correlation of the three eNose systems.

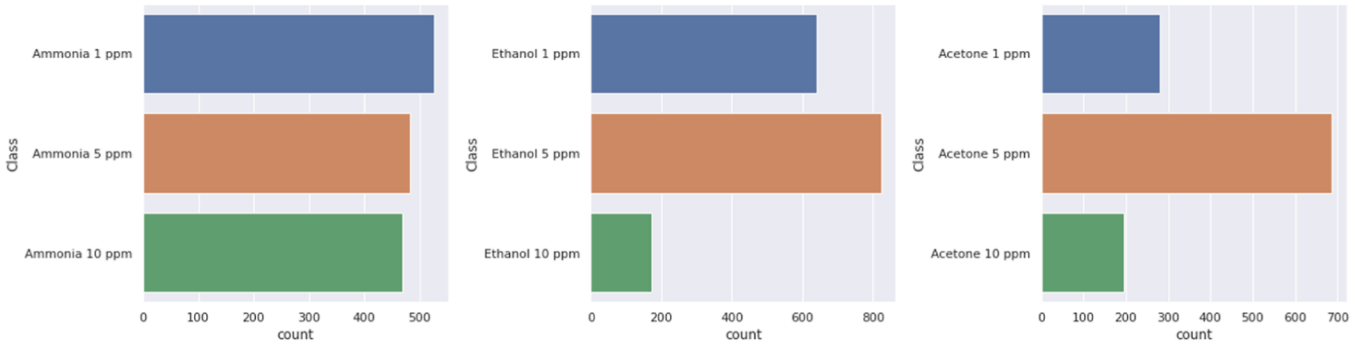


Figure 10. Class distribution of the three eNose systems.

Table 4. Classification Reports of the Confusion Matrix

biomarkers	precision	recall	F1-score
acetone 1 ppm	0.974026	0.892857	0.931677
acetone 10 ppm	0.892308	1.000000	0.943089
acetone 5 ppm	0.962441	0.995146	0.978520
ammonia 1 ppm	1.000000	1.000000	1.000000
ammonia 10 ppm	0.985816	0.985816	0.985816
ammonia 5 ppm	0.993151	1.000000	0.996564
ethanol 1 ppm	0.994792	0.989637	0.992208
ethanol 10 ppm	1.000000	0.882353	0.937500
ethanol 5 ppm	1.000000	0.995951	0.997972
accuracy	0.983632	0.983632	0.983632
macro avg.	0.978059	0.971307	0.973705
weighted avg.	0.984284	0.983632	0.983511

4.3. Class Distributions and Box Plots of the Data Set.

In the overall data set, 80% of the data have been considered training data and 20% for testing and validation. Figure 6 shows the class distributions of the training data set based on different attributes and threshold values, which provide information on imbalanced class variations. Figure 7 shows the box and whisker plots of the features, which are grouped by class. The medians confirm the symmetric and skewed distributions of the class data and help to detect and remove outliers from the sensor array data set. The whiskers provide information on the minimum and maximum data range by first or third quartile plus 1.5 times the interquartile range, which helps to attain higher accuracy of the data set.

5. CLASSIFICATION ANALYSIS OF THE HOMEMADE SENSOR ARRAY DATA SET

5.1. Feature Visualization and Correlation.

Three different room temperature-operated sensors such as ZnO,[23] Cr−ZnO, and ZnO/NiO[22] have been fabricated for the specific detection of ammonia, ethanol, and acetone. These sensors were individually integrated with the aforementioned sensor array. Thus, the eNose system consists of three different groups (G1, G2, and G3) of a sensor array. The sensor array's group details and features are given in Tables 2 and 3.

The feature distribution and correlations between the same for all three groups are displayed in Figures 8 and 9. The feature correlation provides information about the magnitude of the associations and the directions of the feature relations. It confirms the existence of both the positive and negative

https://doi.org/10.1021/acsomega.3c03755
ACS Omega 2024, 9, 215−226

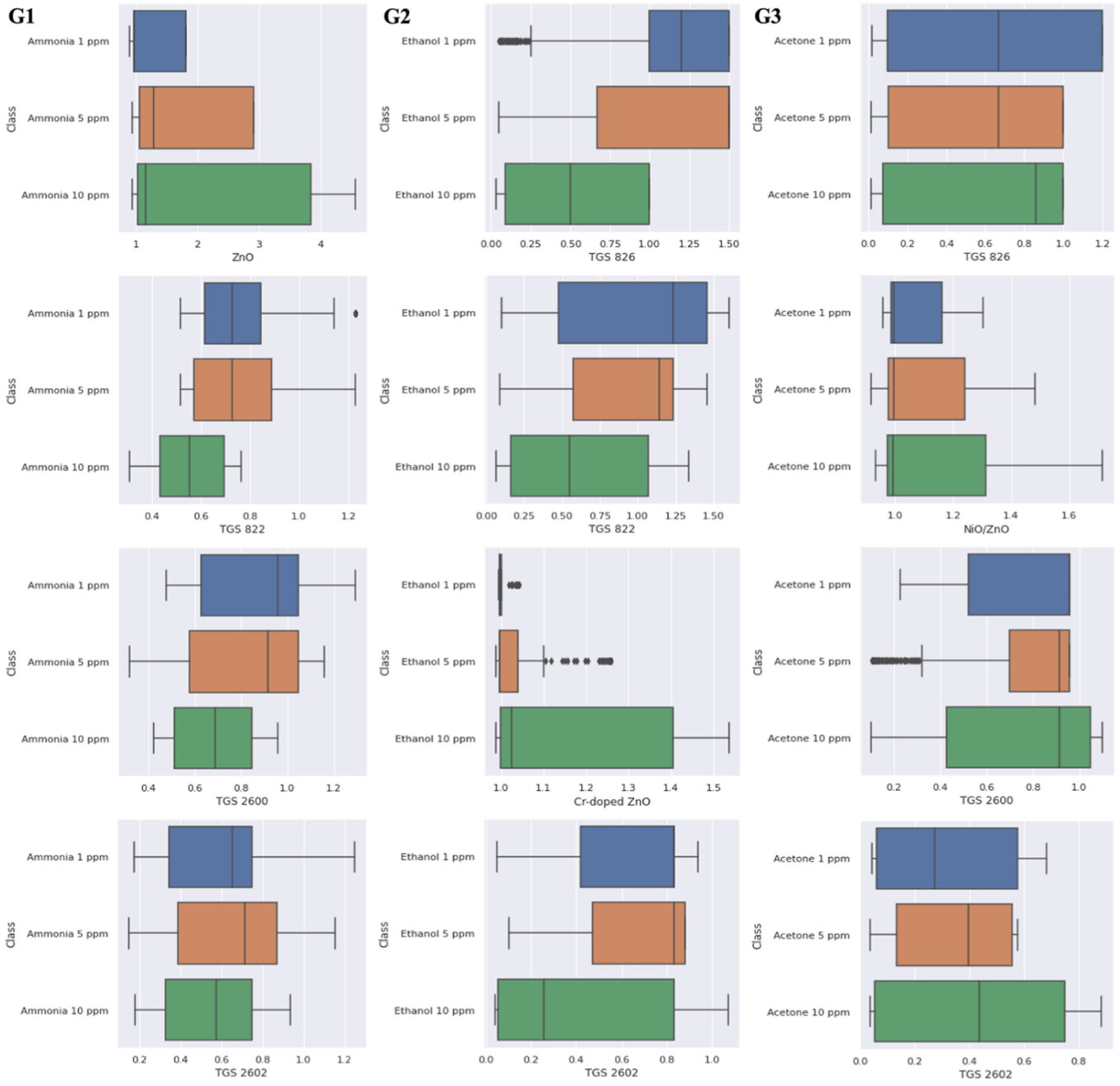


Figure 11. Box plots of the training set by the three classifiers of all three groups.

correlations. Thus, the positive correlation provides information on the dependent variable of features, but the negative correlation gives information on the independent variable of features.

5.2. Class Distributions and Box Plots of the Data Set. Figure 10 confirms the imbalanced class distributions of the sensor array data set for all three groups. A visual analysis of the performance comparison of all three groups of eNose systems is depicted by a box plot, as shown in Figure 11. The box plot reports the performance of each sensor with the average accuracy and class variance. In the box, the center line denotes the median, the box border denotes the 25th and 75th percentiles, and the whiskers are extended to the extreme data points that are not deemed outliers, but the outliers are shown individually. Additionally, box plots provide information on overlapped notches, which reduce the accuracy of the particular sensor array data system.

6. RESULTS AND DISCUSSION

This section describes the experimental results of the proposed enhanced XGBoost classifier model. The various quality measures are applied to evaluate the model performance in various dimensions.

6.1. Classification Accuracy Result of Readymade Sensor Arrays. The prediction result of the enhanced XGBoost classifier model is plotted as a confusion matrix in Figure 12. It demonstrates the summary of each class prediction result over the data set. This study has investigated the performance of the enhanced XGBoost classifier model using eqs 1−4. The precision, recall, F1 score, and accuracy are

https://doi.org/10.1021/acsomega.3c03755
ACS Omega 2024, 9, 215−226

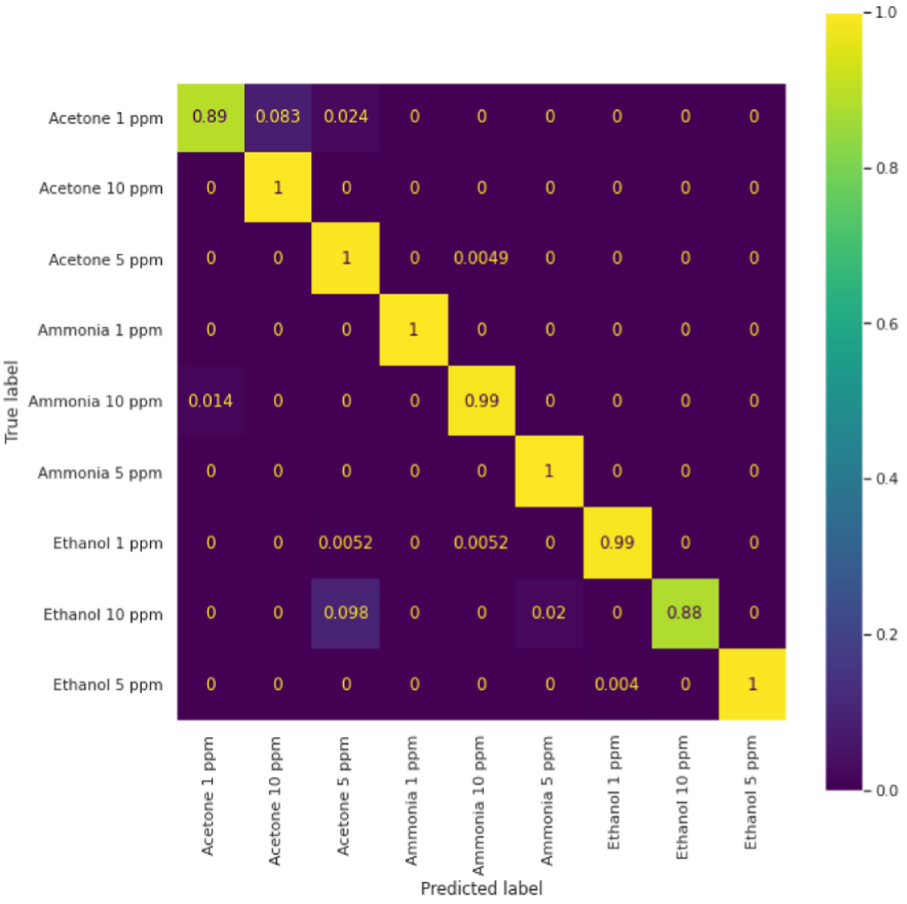


Figure 12. Confusion matrix of the enhanced XGBoost classifier results for readymade sensor data.

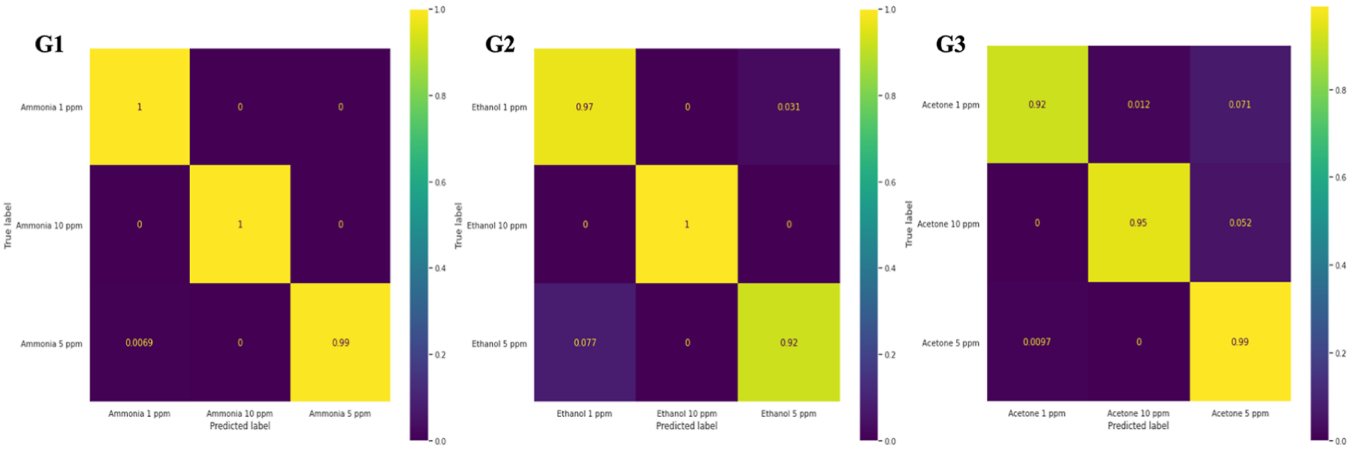


Figure 13. Confusion matrix of the enhanced XGBoost classifier model results for homemade sensor data.

determined using the same set of equations, where TP, TN, FP, and FN represent the true positive, true negative, false positive, and false negative values.[28,29] The computed metrics are given in Table 4. The proposed model exhibited 98.36% classification accuracy for the case of a ready-made sensor array.

Precision is the quality of a positive prediction made by the sensor array model. It can be measured by the number of true positives divided by the total number of positive predictions.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

Recall is the model's ability to detect positive predictions. It is evaluated from the ratio between the number of positive samples correctly classified as positive to the total number of positive samples.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

F1-score is the harmonic mean between precision and recall value. This statistical measurement is used to determine the performance accuracy.

https://doi.org/10.1021/acsomega.3c03755
ACS Omega 2024, 9, 215−226

Table 5. Prediction Accuracy of the Homemade Sensor Data

groups	biomarkers	precision	recall	F1-score
G1	ammonia 1 ppm	0.993711	1.000000	0.996845
	ammonia 10 ppm	1.000000	1.000000	1.000000
	ammonia 5 ppm	1.000000	0.993103	0.996540
	accuracy	0.997748	0.997748	0.997748
	aacro avg.	0.997904	0.997701	0.997795
	weighted avg.	0.997762	0.997748	0.997747
G2	ethanol 1 ppm	0.907767	0.968912	0.937343
	ethanol 10 ppm	1.000000	1.000000	1.000000
	ethanol 5 ppm	0.974359	0.923077	0.948025
	accuracy	0.949187	0.949187	0.949187
	macro avg.	0.960709	0.963996	0.961789
	weighted avg.	0.950947	0.949187	0.949328
G3	acetone 1 ppm	0.974684	0.916667	0.944785
	acetone 10 ppm	0.982143	0.948276	0.964912
	acetone 5 ppm	0.957944	0.990336	0.973872
	accuracy	0.965616	0.965616	0.965616
	macro avg.	0.971590	0.951760	0.961190
	weighted avg.	0.965995	0.965616	0.965382

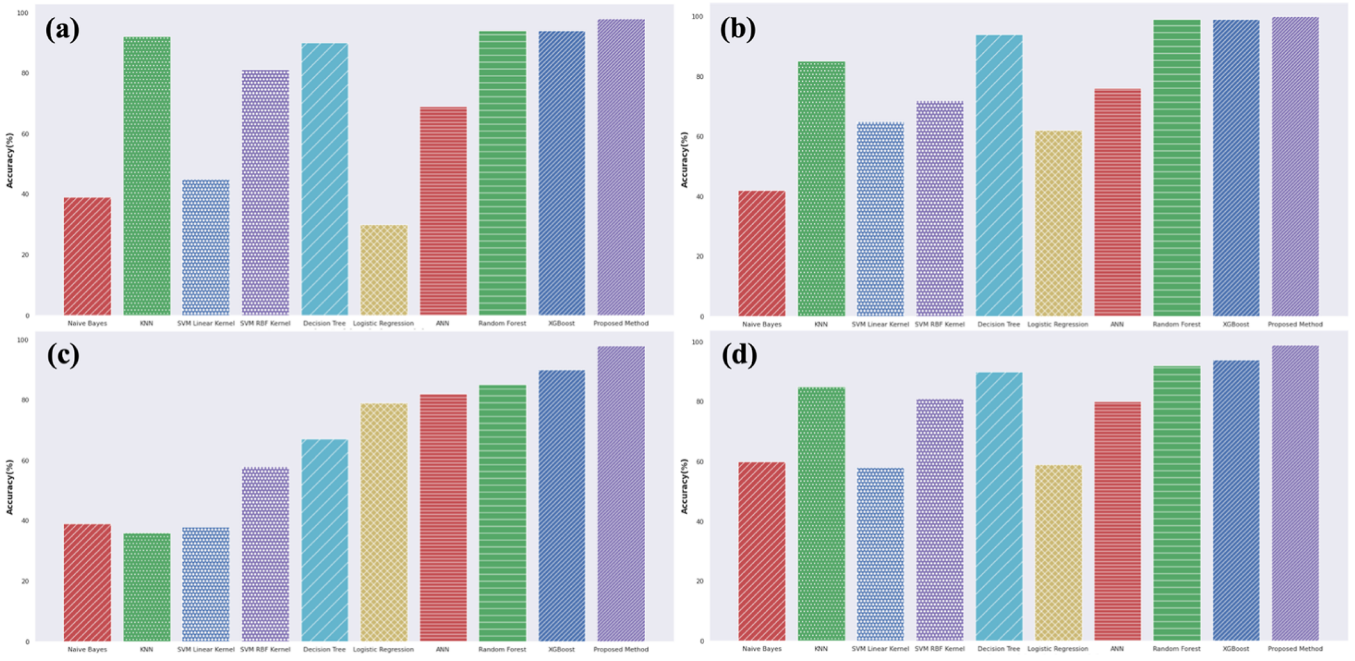


Figure 14. Performance analysis of the enhanced XGBoost classifier model for (a) readymade sensors and (b) G1, (c) G2, and (d) G3 arrays.

Table 6. Comparison of Existing Reports with the Enhanced XGBoost Classifier Model

author	method	accuracy (%)
Azraret al.[30]	KNN, decision trees, and naive Bayes	65.19,75.65,71.74
Kumari et al.[31]	SVM	78
Xu et al.[32]	RF	85
Antony and Singh[33]	NB, MLP, and RF	76.53,76.76,84.77
Krishnamurthy et al.[34]	logistic regression	83
proposed model	enhanced XGBoost classifier	98.36 (readymade sensor)
		99.77 (G1)
		94.91 (G2)
		96.56 (G3)

$$\text{F1 Score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (3)$$

And the accuracy is a standard value, i.e., the total percentage of positive predictions.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (4)$$

6.2. Classification Accuracy of the Homemade Sensor Array Data. Figure 13 depicts the confusion matrix of the enhanced XGBoost classifier results for homemade sensors. Table 5 summarizes the prediction accuracy of the homemade sensor's data set. The proposed model exhibits 99.77, 94.91, and 96.56% classification accuracy of the groups G1, G2, and G3, respectively.

https://doi.org/10.1021/acsomega.3c03755
ACS Omega 2024, 9, 215−226

The performance of the enhanced XGBoost classifier model is compared with that of the existing techniques, and the results are depicted in Figure 14. It clearly showed a better performance of the enhanced XGBoost classifier model than that of the existing models. The accuracy of the enhanced XGBoost classifier model is compared with that of the other models and is given in Table 6. The proposed technique possesses higher accuracy than the existing techniques.

Overall, the developed readymade sensor prototype responds to the multicomponents and secured an accuracy of 98.36%. The fabricated homemade sensors are selective toward specific biomarkers as discussed in our previous work.[22,23] Hence, the accuracy of the prototype is significantly enhanced toward specific biomarkers after integrating homemade sensors.

7. CONCLUSIONS

The design and development of eNose prototypes using readymade and homemade sensors for the detection of biomarkers, namely, ammonia, ethanol, and acetone, have been successfully accomplished. The enhanced XGBoost classifier approach has been adapted to perform multiclass classification of sensor array data. The implemented model has secured 98.36% prediction accuracy in multiclass disease prediction and classification of the eNose developed using readymade sensors. On the other hand, the eNose models with three different homemade sensor groups G1, G2, and G3 have exhibited a classification accuracy of 99.77, 94.91, and 96.56% toward ammonia, ethanol, and acetone, respectively. Overall, the prediction accuracies of the homemade eNose prototypes provide better performance than those of the commercially available sensors. Hence, the developed eNose systems with inbuilt homemade sensors may be deployed for real-time analysis.

AUTHOR INFORMATION

Corresponding Author

John Bosco Balaguru Rayappan − *Centre for Nanotechnology & Advanced Biomaterials (CeNTAB) and School of Electrical & Electronics Engineering, SASTRA Deemed University, Thanjavur, Tamil Nadu 613 401, India;* orcid.org/0000-0003-4641-9870; Phone: +91 4362 264 101-108 ext. 2255; Email: rjbosco@ece.sastra.edu

Authors

Bhuvaneswari Selvaraj − *Centre for Nanotechnology & Advanced Biomaterials (CeNTAB) and School of Electrical & Electronics Engineering, SASTRA Deemed University, Thanjavur, Tamil Nadu 613 401, India*

Elakkiya Rajasekar − *Department of Computer Science, Birla Institute of Technology & Science, Dubai 345055, United Arab Emirates*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c03755

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors wish to express their sincere gratitude to the Science and Engineering Research Board (SERB), Department of Science & Technology (DST) {New Delhi, India, for their financial support [ECR/2016/001805 and SR/FST/ET-I/2018/221(C)]}. We also acknowledge SASTRA Deemed University, Thanjavur, for extending infrastructure support to carry out this work.

REFERENCES

(1) Turner, A. P. F.; Magan, N. Electronic Noses and Disease Diagnostics. *Nat. Rev. Microbiol.* **2004**, *2* (2), 161−166.

(2) Wilson, A. D.; Baietto, M. Advances in Electronic-Nose Technologies Developed for Biomedical Applications. *Sensors* **2011**, *11*, 1105−1176.

(3) Batko, K.; Ślęzak, A. The Use of Big Data Analytics in Healthcare. *J. Big Data* **2022**, *9* (1), 3.

(4) Al-Masri, S.; Safi, A.; El-Farra, A. A. H.; Shehada, M. Designing a Biosignal Extraction Device for Measuring a Blood Pressure. In *2018 International Conference on Promising Electronic Technologies (ICPET)*; IEEE, 2018; pp 108−112.

(5) Razzak, M. I.; Imran, M.; Xu, G. Big Data Analytics for Preventive Medicine. *Neural. Comput. Appl.* **2020**, *32* (9), 4417−4451.

(6) Kaloumenou, M.; Skotadis, E.; Lagopati, N.; Efstathopoulos, E.; Tsoukalas, D. Breath Analysis: A Promising Tool for Disease Diagnosis—The Role of Sensors. *Sensors* **2022**, *22*, 1238.

(7) Abdel-Aziz, M. I.; Brinkman, P.; Vijverberg, S. J. H.; Neerincx, A. H.; de Vries, R.; Dagelet, Y. W. F.; Riley, J. H.; Hashimoto, S.; Montuschi, P.; Chung, K. F.; Djukanovic, R.; Fleming, L. J.; Murray, C. S.; Frey, U.; Bush, A.; Singer, F.; Hedlin, G.; Roberts, G.; Dahlén, S. E.; Adcock, I. M.; Fowler, S. J.; Knipping, K.; Sterk, P. J.; Kraneveld, A. D.; Maitland-van der Zee, A. H. ENose Breath Prints as a Surrogate Biomarker for Classifying Patients with Asthma by Atopy. *J. Allergy Clin. Immunol.* **2020**, *146* (5), 1045−1055.

(8) Wu, J.; Chen, Z. Data Decision and Transmission Based on Mobile Data Health Records on Sensor Devices in Wireless Networks. *Wirel. Pers. Commun.* **2016**, *90* (4), 2073−2087.

(9) Sadad, T.; Rehman, A.; Munir, A.; Saba, T.; Tariq, U.; Ayesha, N.; Abbasi, R. Brain Tumor Detection and Multi-Classification Using Advanced Deep Learning Techniques. *Microsc. Res. Tech.* **2021**, *84*, 1296−1308.

(10) Alam, F.; Islam, A.; Mohamed, M.; Ahmad, I.; Kamal, M. A.; Donnelly, R.; Idris, I.; Gan, S. H. Efficacy and Safety of Pioglitazone Monotherapy in Type 2 Diabetes Mellitus : A Systematic Review and Meta-Analysis of Randomised Controlled Trials. *Sci. Rep.* **2019**, *9*, 5389.

(11) Chandran, V.; Sumithra, M. G.; Karthick, A.; George, T.; Deivakani, M.; Elakkiya, B.; Subramaniam, U.; Manoharan, S. Diagnosis of Cervical Cancer Based on Ensemble Deep Learning Network Using Colposcopy Images. *BioMed Res. Int.* **2021**, *2021*, 5584004.

(12) Macalisang, J. R.; Merencilla, N. E.; Ligayo, M. A. D.; Melegrito, M. P.; Tejada, R. R. Eye-Smoker: A Machine Vision-Based Nose Inference System of Cigarette Smoking Detection Using Convolutional Neural Network. In *2020 IEEE 7th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*; IEEE 2020; pp 1−5.

(13) Hendrick, H.; Hidayat, R.; Horng, G.-J.; Wang, Z.-H. Non-Invasive Method for Tuberculosis Exhaled Breath Classification Using Electronic Nose. *IEEE Sens. J.* **2021**, *21* (9), 11184−11191.

(14) Binson, V. A.; Subramoniam, M.; Sunny, Y.; Mathew, L. Prediction of Pulmonary Diseases With Electronic Nose Using SVM and XGBoost. *IEEE Sens. J.* **2021**, *21* (18), 20886−20895.

(15) Saeedi, P.; Petersohn, I.; Salpea, P.; Malanda, B.; Karuranga, S.; Unwin, N.; Colagiuri, S.; Guariguata, L.; Motala, A. A.; Ogurtsova, K.; Shaw, J. E.; Bright, D.; Williams, R. Global and Regional Diabetes Prevalence Estimates for 2019 and Projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th Edition. *Diabetes Res. Clin. Pract.* **2019**, *157*, 107843.

(16) Kovesdy, C. P. Epidemiology of Chronic Kidney Disease: An Update 2022. *Kidney Int. Suppl.* **2022**, *12*, 7−11.

(17) Amouzegar, A.; Abu-Alfa, A. K.; Alrukhaimi, M. N.; Bello, A. K.; Ghnaimat, M. A.; Johnson, D. W.; Jha, V.; Harris, D. C. H.; Levin, A.;

https://doi.org/10.1021/acsomega.3c03755
ACS Omega 2024, 9, 215−226

Tonelli, M.; Lunney, M.; Saad, S.; Khan, M.; Zaidi, D.; Osman, M. A.; Ye, F.; Okpechi, I. G.; Ossareh, S. International Society of Nephrology Global Kidney Health Atlas: Structures, Organization, and Services for the Management of Kidney Failure in the Middle East. *Kidney Int. Suppl.* **2021**, *12*, e47−e56.
(18) Darnton-Hill, I. Public Health Aspects in the Prevention and Control of Vitamin Deficiencies. *Curr. Dev. Nutr.* **2019**, *3*, NZZ075.
(19) Budreviciute, A.; Damiati, S.; Sabir, D. K.; Onder, K.; Schuller-Goetzburg, P.; Plakys, G.; Katileviciute, A.; Khoja, S.; Kodzius, R. Management and Prevention Strategies for Non-Communicable Diseases (NCDs) and Their Risk Factors. *Front. Public Health* **2020**, *8*, 574111.
(20) Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable Ai: A Review of Machine Learning Interpretability Methods. *Entropy* **2021**, *23*, 18.
(21) Elakkiya, R.; Vijayakumar, P.; Karuppiah, M. COVID_SCREENET: COVID-19 Screening in Chest Radiography Images Using Deep Transfer Stacking. *Inf. Syst. Front.* **2021**, *23*, 1369.
(22) Selvaraj, B.; Rayappan, J. B. B.; Babu, K. J. Room Temperature ZnO/NiO Heterostructure Sensing Response: A Breath Biomarker Sensor. *J. Alloys Compd.* **2022**, *914*, 165224.
(23) Selvaraj, B.; Balaguru Rayappan, J. B.; Jayanth Babu, K. Influence of Calcination Temperature on the Growth of Electrospun Multi-Junction ZnO Nanowires: A Room Temperature Ammonia Sensor. *Mater. Sci. Semicond. Process.* **2020**, *112*, 105006.
(24) Kumari, B.; Kumar, R.; Kumar, M. PalmPred: An SVM Based Palmitoylation Prediction Method Using Sequence Profile Information. *PLoS One* **2014**, *9* (2), No. e89246.
(25) Azrar, A.; Awais, M.; Ali, Y.; Zaheer, K. Data Mining Models Comparison for Diabetes Prediction, 2018; Vol. *9*. www.ijacsa.thesai.org.
(26) Kunwar, V.; Chandel, K.; Sabitha, A. S.; Bansal, A. Chronic Kidney Disease Analysis Using Data Mining Classification Techniques. In *2016 6th International Conference—Cloud System and Big Data Engineering (Confluence)*; IEEE, 2016; pp 300−305
(27) Xu, L.; He, J.; Duan, S.; Wu, X.; Wang, Q. Comparison of Machine Learning Algorithms for Concentration Detection and Prediction of Formaldehyde Based on Electronic Nose. *Sens. Rev.* **2016**, *36* (2), 207−216.
(28) Islam, M. M.; Haque, M. R.; Iqbal, H.; Hasan, M. M.; Hasan, M.; Kabir, M. N. Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques. *SN Comput. Sci.* **2020**, *1* (5), 290.
(29) Dalianis, H. *Clinical Text Mining: Secondary Use of Electronic Patient Records*; Springer International Publishing, 2018.
(30) Azrar, A.; Ali, Y.; Awais, M.; Zaheer, K. Data Mining Models Comparison for Diabetes Prediction. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9* (8), 320−323.
(31) Kumari, B.; Kumar, R.; Kumar, M. PalmPred : An SVM Based Palmitoylation Prediction Method Using Sequence Profile Information. *PLoS One* **2014**, *9* (2), No. e89246.
(32) Xu, X.; Tan, M.; Wu, J.; Morandotti, R.; Mitchell, A.; Moss, D. J. Microcomb-Based Photonic RF Signal Processing. *IEEE Photonics Technol. Lett.* **2019**, *31* (23), 1854−1857.
(33) Antony, A.; Singh, G. Diabetes Prediction Using Medical Data. *Journal of Computational Intelligence in Bioinformatics* **2017**, *10*, 1.
(34) Thilothammal, N.; Krishnamurthy, P. V.; Runyan, D. K.; Banu, K. Does BCG Vaccine Prevent Tuberculous Meningitis? *Arch. Dis. Child.* **1996**, *75*, 144−147.

https://doi.org/10.1021/acsomega.3c03755
ACS Omega 2024, 9, 215−226