Talanta 297 (2026) 128554

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

A novel VOC mixtures classification methods based on GBLinear and TabNet and informative feature selection from gas sensors (E-Nose) data

Hamed Karami [a,*], Abdolrahman Khoshrou [b]

[a] Department of Petroleum Engineering, Knowledge University, Erbil, 44001, Iraq
[b] System Operations department, Alliander, Arnhem, The Netherlands

ARTICLE INFO

Keywords:
Advanced algorithms
Classification methodology
Feature selection
Gas mixture analysis
Electronic nose

ABSTRACT

The GBLinear and TabNet algorithms have been incorporated with essential feature selection techniques to create a new method of classifying essential oils using e-nose systems. Essential oils, known for their complex chemical compositions and a wide variety of applications in industries such as food, cosmetics, and pharmaceuticals, pose some challenges for e-nose systems due to the high variability and subtle differences in volatile compounds (VOCs). This novel approach, used for the first time for the analysis of electronic nose data, integrates efficient machine-learning models with advanced feature selection techniques and aims to increase the accuracy and interpretability of essential oil classification. This study highlights the potential of integrating interpretable machine learning models with deep learning-based architectures to address challenges in the analysis of complex gas mixtures. Not only was the classification accuracy increased by these methods, but these methods could be used in the future as promising models for analyzing complex mixtures.

1. Introduction

The electronic nose, commonly known as e-nose, is a new and developing technology that can potentially impact the world of analyzing and identifying multi-mixed volatile compounds eg; VOCs [1]. These systems embed chemical gas sensors, signal processing, and machine learning techniques and are used in different sectors such as medicine [2,3], environment [4,5], food safety and quality [6–8]. These devices use chemical sensors that capture odor profiles and translate them into measurable signals trying to replicate the biological olfactory system [9]. E-noses are now very helpful due to their nature of being able to analyze complex odors accurately in a short amount of time and not needing to be intrusive, especially when traditional means of analysis are long, costly, and troublesome [4,10]. However, the recognition of odor using e-noses is challenging largely due to the methods used to process the signals from the sensors, therefore finding suitable processing methods is a key area in research [11].

The primary step in electronic nose gas recognition is data analysis, which typically involves three stages: data preprocessing, feature extraction and selection, and pattern recognition [12]. Data preprocessing removes noise and sensor drift using filters like wavelet, Gaussian [13], Savitzky–Golay [14] or, Kalman [15], ensuring standardized data. Gas sensing signals, often high-dimensional, are reduced using techniques such as Principal Component Analysis [16] or other feature extraction methods [17]. Feature selection follows to minimize computational complexity. Finally, machine learning algorithms such as genetic algorithms, K-nearest neighbor, support vector machines, and decision trees, are employed for gas classification, with studies comparing their performance across various tasks [18].

Researchers are always on the lookout for new and improved ways to classify e-nose systems, as they are becoming increasingly advanced [19,20]. However, that presents the problem of accurately processing and understanding the vast amount of noisy data produced by the sensors [21]. This has triggered the growing need for more advanced algorithms that can better model the characteristics of e-nose data [22]. Despite the effectiveness of these methods in so many areas, they are often limited when the sensor data are high dimensional, non-linear, or have complicated interactions among features [11]. Moreover, the inherent noise and variability present in e-nose datasets further complicate these tasks and highlight the importance of exploring more advanced classification frameworks that can withstand noise while being able to preserve more intricate patterns within the data [21].

In this regard, the introduction of new machine learning structures is one of the solutions that can be effective for analyzing electronic nose

* Corresponding author. Department of Petroleum Engineering, Knowledge University, Erbil, 44001, Iraq.
E-mail address: hamed.karami@knu.edu.iq (H. Karami).

https://doi.org/10.1016/j.talanta.2025.128554
Received 15 February 2025; Received in revised form 30 April 2025; Accepted 4 July 2025
Available online 8 July 2025
0039-9140/

Table 1
The types of sensors, gas detection limits, and established chemical sensitivities of tin oxide MOS sensors incorporated within the electronic nose array.

Sensor name	Detection ranges (ppm)	Main applications (gas detection)
MQ9	10-1000 and 100-10000	CO, combustible gases
MQ4	300–100	Urban gases and methane
MQ135	10–10000	Steam ammonia, benzene, sulfides
MQ8	100–1000	Hydrogen (H_2)
TGS2620	50–5000	Alcohols, steam organic solvents
MQ136	1–200	Sulfur dioxide (SO_2)
TGS813	500–10000	CH_4, C_3H_8, C_4H_{10} (hydrocarbons)
TGS822	50–5000	Steam organic solvents
MQ3	10–300	Alcohols

datasets. GBLinear and TabNet are two advanced models that seek to address the shortcomings of classical models in data analysis. GBLinear, for example, works extremely well in high dimensional datasets for applications that require linear interactions due to its ability to capture linear relationships. This approach is well suited to the balance of enhancing a model's complexity while ensuring usability, thus making it a viable option for scenarios in which reasoning behind the decision is of value. Instead, GBLinear can still take complex relationships using linear models' simplicity and efficacy by sequentially applying linear regression to the residuals of the last fitted models [23–25]. This feature is of great importance, especially for e-nose systems because the relationship between the sensor responses and the corresponding odor classes may be far from linear. However, in terms of completeness, I must mention TabNet, which is a latest deep neural network model tailored for tabular data. The combination of its capacity for modeling complex nonlinear relationships and other features such as attention and interpretability make it an appropriate candidate for increasing the performance of e-nose. Developed by Google researchers, TabNet applies attention and feature selection for automatically searching for the features that are most informative for each prediction. Understanding which parts of the data provide the most relevant information allows for more accurate predictions and better training of the model. Also, TabNet's interpretability features can help scientists understand which particular responses of the sensors are most helpful for discriminating some specific compounds, thus aiding the design of the sensor array in a more efficient manner [26–28].

The investigational method proposed here takes advantage of the simplicity and efficiency of GBLinear, but does not neglect the importance of the deep learning aspects offered by TabNet, such as its multiple interactions between non-linear features. Together, these methods present a comprehensive framework that reconciles the often conflicting requirements of explainability and accuracy. The results of this study have major ramifications for the development of the next generation e-nose systems. All the aforementioned factors make the GBLinear-TabNet fusion a step towards greater accuracy in e-nose classification.

2. Material and methods

2.1. Sample preparation

Seven edible essential oils were prepared from various sources. These included three medicinal plant oils; mint, tarragon, and thyme, and four fruit-based oils; mango, lemon, orange, and strawberry. For each sample, both pure and industrial-grade essential oils were prepared. A total of 14 groups of essential oils were considered for the experiments.

2.2. Electronic nose instrument

A custom-built electronic nose system, developed by Karami [29], utilized a 9-sensor tin metal oxide semiconductor (MOS) array (Table 1) to analyze the aroma signature patterns of VOCs emitted by purified essential oils in this study. The system comprised five key components: an activated carbon filter, a sample headspace chamber, three one-way valves, a diaphragm pump, and a sensor array linked in series via 4.0 mm PTFE tubing. Sensor data were captured using a wireless data recorder, which transmitted the information to a personal computer (PC) for collation and statistical analysis.

The data acquisition process from the e-nose system was divided into three phases: baseline establishment, sample odor injection, and purification. In the baseline phase, clean air from the filter was introduced into the sensor chamber through a pump and one-way valves at a flow rate of 0.8 L per minute for 60 s to stabilize the sensor response. During the sample odor injection phase, the sample head was injected into the sensor chamber, maintaining the same flow rate for 150 s, until a stable voltage response was achieved. In the purification phase, clean air was reinjected into the sensor chamber by opening the solenoid valve for 60 s, allowing the sensors to return to their baseline values. To minimize instrument baseline drift, the analysis of sample headspace volatiles from purified essential oils was conducted daily during data acquisition. At least 15 replicates per sample were measured for each essential oil type, and all data recorded by a data card were transmitted wirelessly to a computer for further analysis.

2.3. Data analyze

Data from the individual sensors of the e-nose sensor array were extracted for preprocessing and subsequent analysis. The purpose of signal preprocessing was to isolate relevant information from the sensor responses and prepare the data for multivariate pattern analysis. To achieve this, sensor responses were normalized against their baseline for thrust compensation, contrast enhancement, and scaling, using the fraction method outlined by Karami, Rasekh and Mirzaee-Ghaleh [30], as shown in equation (1):

$$Y_s(t) = \frac{X_s(t) - X_s(0)}{X_s(0)} \tag{1}$$

Where Ys(t), Xs(0), and Xs(t) represent the normalized sensor response, the baseline, and the raw unprocessed sensor response, respectively.

2.3.1. GBLinear model (gradient boosted linear)

GBLinear is a variant of the gradient boosting algorithm [31] specifically designed for linear models. Unlike its tree-based counterparts, GBLinear uses linear functions as weak learners, making it particularly effective for datasets where relationships among features are linear or nearly linear. By combining the power of gradient boosting with linear regression or logistic regression, GBLinear provides both simplicity and interpretability. Its straightforward implementation in frameworks like XGBoost makes it a practical choice for various domains that prioritize interpretable results.

For multi-class classification with k classes, the model uses the softmax function to compute class probabilities and a cross-entropy loss function as follows:

$$\mathscr{L}(\Theta) = -\sum_{i=1}^{n}\sum_{j=1}^{k} y_{i,j}\log\left(\hat{y}_{i,j}\right) + \Omega(\Theta) \tag{2}$$

Where:

$y_{i,j} \in \{0,1\}$ is a one-hot encoding of the true label for sample *i*.

$\hat{y}_{i,j} = \frac{\exp\left(W_j^T x_i + b_j\right)}{\sum_{l=1}^{k}\exp\left(W_l^T x_i + b_l\right)}$ is the predicted probability for class *j,* obtained using the softmax function.

$\Theta = \{W, b\}$, where W is the matrix of weights and b is the vector of biases for all classes.

$\Omega(\Theta) = \frac{1}{2}\lambda\|W\|^2 + \alpha\|W\|_1$ includes the regularization terms.

2.3.2. TabNet model (tabular neural networks)

TabNet is a deep learning architecture that uniquely combines

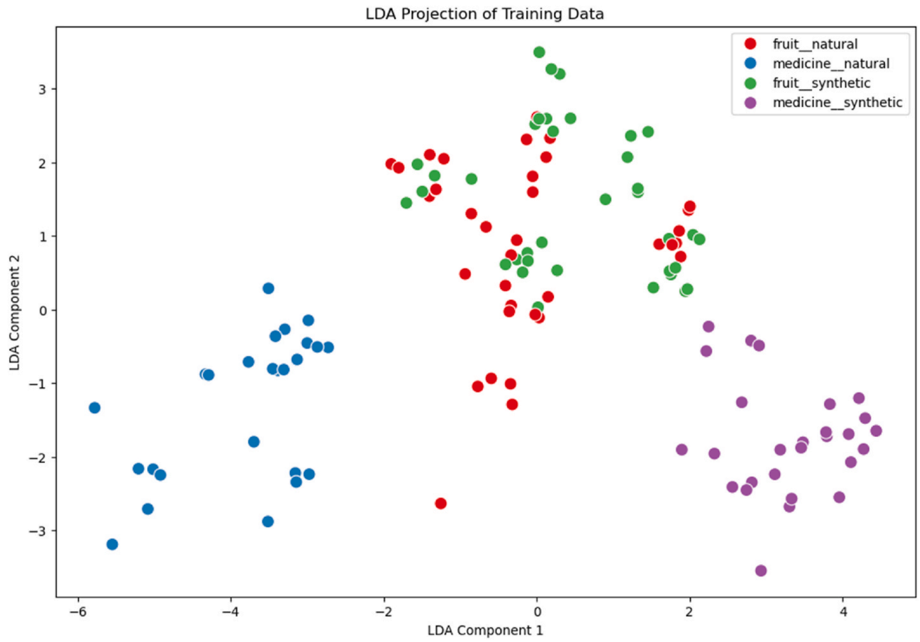

Fig. 1. Two-dimensional LDA plot.

feature selection and decision-making processes within a unified framework, leveraging sequential attention mechanisms. Unlike traditional gradient-boosted models or neural networks, TabNet dynamically selects features at each decision step, allowing the model to focus on the most relevant features for classification. This feature selection is achieved using sparse attention masks, promoting interpretability and efficiency in high-dimensional datasets. TabNet's ability to learn directly from raw tabular data without requiring extensive preprocessing makes it particularly suited for complex, heterogeneous data structures. In this study, TabNet was particularly valuable for capturing non-linear interactions and subtle patterns in the volatile organic compounds (VOCs) emitted by essential oils.

$$\mathscr{L}_{CE} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} y_{i,j}\log\left(\widehat{y}_{i,j}\right) \tag{3}$$

Where:

similar to the previous case $y_{i,j}$ is a one-hot encoding of the true labels for sample i. $\widehat{y}_{i,j} = \frac{\exp(z_{i,j})}{\sum_{c=1}^{k}\exp(z_{i,c})}$ is the predicted probability for class j for sample i.

$z_{i,j}$ is the raw logit (output before applying softmax) for class j for sample i. For more details, see Ref. [26].

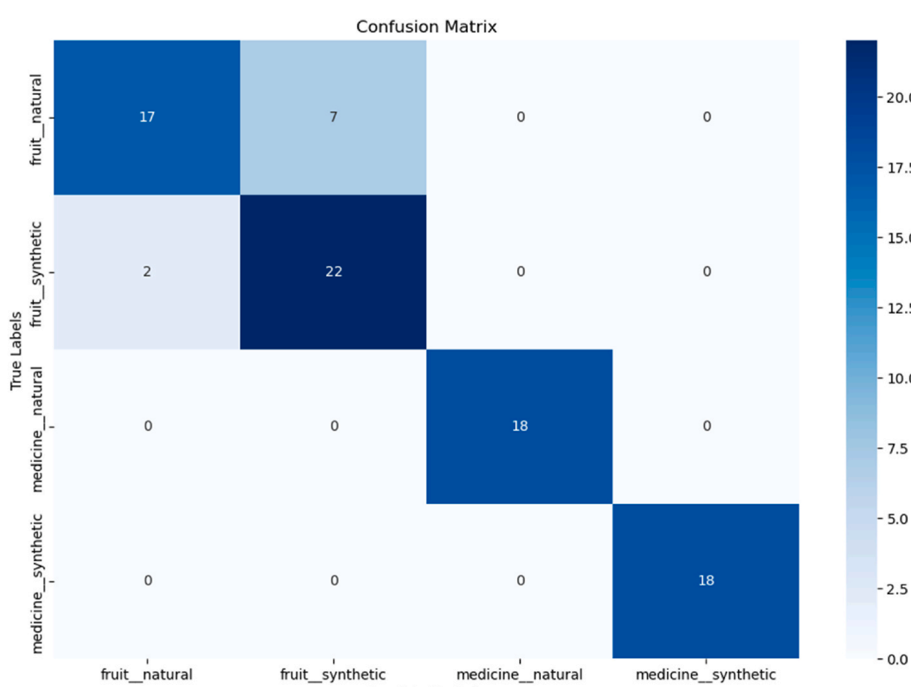

Fig. 2. Confusion matrix resulting from LDA model analysis.

Table 2
Performance parameters obtained from the LDA model.

Group	Accuracy	Precision	Recall	Specificity	AUC	F
Fruit Natural	0.893	0.708	0.895	0.892	0.800	0.791
Fruit Synthetic	0.893	0.917	0.759	0.964	0.940	0.830
Medicine Natural	1.000	1.000	1.000	1.000	1.000	1.000
Medicine Synthetic	1.000	1.000	1.000	1.000	1.000	1.000
Average per class	0.946	0.906	0.913	0.964	0.935	0.905

2.4. Evaluation criteria

To evaluate the system's performance, standard criteria such as Specificity, Recall, Precision, Accuracy, Area Under the Curve (AUC), and F-score were utilized. A confusion matrix, which incorporates true positive (TP), false positive (FP), true negative (TN), and false negative (FN) values, was used to calculate these metrics. The diagnostic criteria considered were outlined by Refs. [32,33]:

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{4}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \tag{7}$$

$$\text{AUC} = \frac{\text{Sensitivity} + \text{Precision}}{2} \tag{8}$$

$$\text{F} = \frac{2 \times \text{PR}}{P + R} \tag{9}$$

In this study, 70 % of the data was used for training and 30 % for validation, and all analyses were performed using Python (version 3.9.12).

3. Result

3.1. LDA

In this study, LDA was used to visually evaluate the grouping and overlap of different classes within the dataset, providing insights into the distinctiveness of each class based on their features. Additionally, LDA served as a benchmark classification method, offering a comparative baseline for evaluating the performance of more complex machine learning models used in subsequent analyses. As shown in Fig. 1, the natural essential oil samples from medicinal plants are clearly distinguishable from their synthetic counterparts. However, for fruit-based samples, there is noticeable overlap between the natural and synthetic types.

The confusion matrix for classifying different essential oil groups is shown in Fig. 2. For each class, the main diagonal values represent true positives (TP), while the sum of the remaining diagonal values corresponds to true negatives (TN). Additionally, the sum of the values in the respective column indicates false positives (FP), and the sum of the values in the respective row represents false negatives (FN). According to the figure, within the natural fruit essential oil group, 17 out of 24 samples were correctly classified, while 7 samples were misclassified. In the second group, representing synthetic essential oils, 22 samples were accurately identified, and 2 samples were misclassified. Finally, for the third and fourth groups, comprising pure medicinal essential oils and synthetic medicinal essential oils, all 8 samples in each group were correctly classified.

Based on equations (2)–(7), the performance parameters of the LDA method for classifying essential oils are summarized in Table 2. The confusion matrix was employed to compute the performance parameters of the detection models. As shown in Table 2, the LDA method achieved an impressive 100 % accuracy in classifying the data, highlighting its effectiveness in distinguishing between different essential oil groups. As shown in Table 2, the average precision of the LDA method achieved 90.6 % in data classification, and the values of accuracy, recall, AUC,

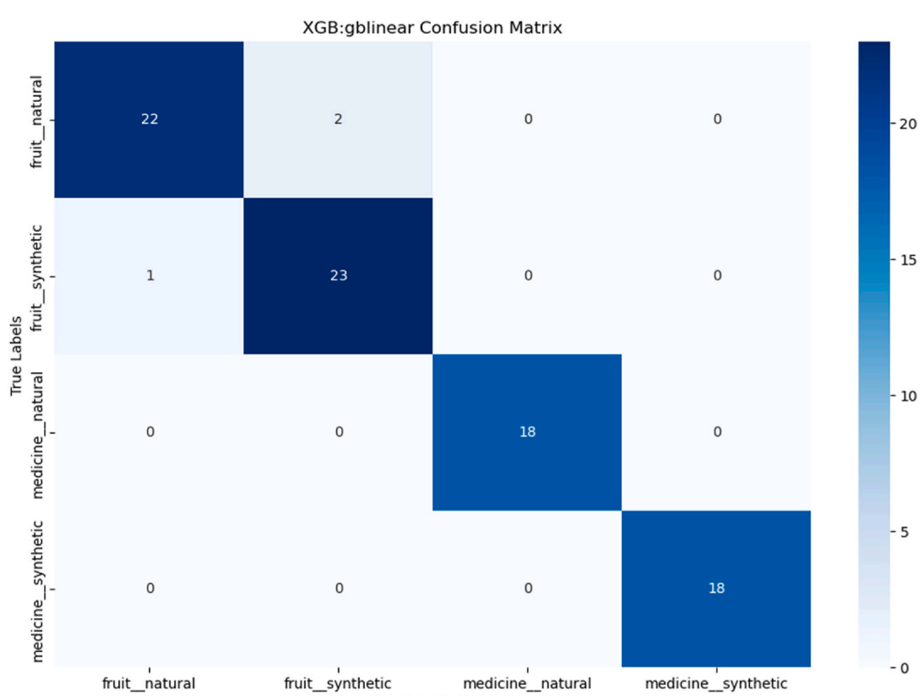

Fig. 3. Confusion matrix resulting from GBLinear model analysis.

Table 3
Performance parameters obtained from the GBLinear model.

Group	Accuracy	Precision	Recall	Specificity	AUC	F
Fruit Natural	0.964	0.917	0.957	0.967	0.942	0.936
Fruit Synthetic	0.964	0.958	0.920	0.983	0.971	0.939
Medicine Natural	1.000	1.000	1.000	1.000	1.000	1.000
Medicine Synthetic	1.000	1.000	1.000	1.000	1.000	1.000
Average per class	0.982	0.969	0.969	0.988	0.978	0.969

and F-score were equal to 94.6, 91.3, 96.4, 93.5, and 90.5 %, respectively.

3.2. GBLinear

The results of the LDA method revealed its limitations in accurately classifying fruit-based essential oils, emphasizing the importance of achieving reliable differentiation within these groups. To address this challenge, GBLinear was utilized as a more robust approach for classifying and analyzing the dataset. This method provided valuable insights into the predictive capabilities of the features and served as a benchmark for comparing its performance with other models. As illustrated in Fig. 3, GBLinear demonstrated exceptional accuracy in distinguishing between natural and synthetic essential oils derived from medicinal plants. For fruit-based samples, while some overlap was observed, only 3 sample were misclassified, showcasing a significant improvement over traditional methods commonly applied to electronic nose data. Also, according to Table 3, the values of precision, recall, and F-score were 96.9 % for the performance parameters.

3.3. TabNet

TabNet was employed to further analyze the dataset and classify essential oil groups, leveraging its advanced deep learning architecture designed for tabular data. This method combines attention mechanisms with feature selection, enabling it to capture complex patterns and relationships within the data. As shown in Fig. 4, TabNet achieved excellent results in distinguishing between natural and synthetic essential oils from both medicinal plants and fruits. Unlike other models, TabNet effectively minimized misclassification, correctly classifying all medicinal plant samples and achieving near-perfect accuracy for fruit-based samples, with only 2 minor overlaps observed. As presented in Table 4, TabNet demonstrated exceptional performance across various classification metrics. The model achieved an average precision of 97.9 %, reflecting its strong ability to correctly identify relevant samples. In terms of other performance metrics, accuracy reached 98.8 %, recall was 97.9 %, AUC scored 98.5 %, and the F-score was 97.9 %. These results highlight TabNet's superior capability in effectively handling complex datasets and ensuring high reliability in the classification of essential oil samples.

According to the results obtained, the models classified 100 % of the medicinal essential oil samples and only failed to classify fruit essential oils completely. Therefore, it is very important to focus on the sensors that are important for this sector in this section. Some sensors in the analysis may exhibit high sensitivity toward specific target analytes, making them crucial for the detection of trace compounds in food samples. Fig. 5a and b illustrate how each sensor contributes positively or negatively to the overall sensor system performance. According to Fig. 5a, the positive effect of the sensors on the classes is observed. As it is clear, the MQ136 sensor is the only sensor that has a 100 % role on the

Table 4
Performance parameters obtained from the TabNet model.

Group	Accuracy	Precision	Recall	Specificity	AUC	F
Fruit Natural	0.976	0.958	0.958	0.983	0.971	0.958
Fruit Synthetic	0.976	0.958	0.958	0.983	0.971	0.958
Medicine Natural	1.000	1.000	1.000	1.000	1.000	1.000
Medicine Synthetic	1.000	1.000	1.000	1.000	1.000	1.000
Average per class	0.988	0.979	0.979	0.992	0.985	0.979

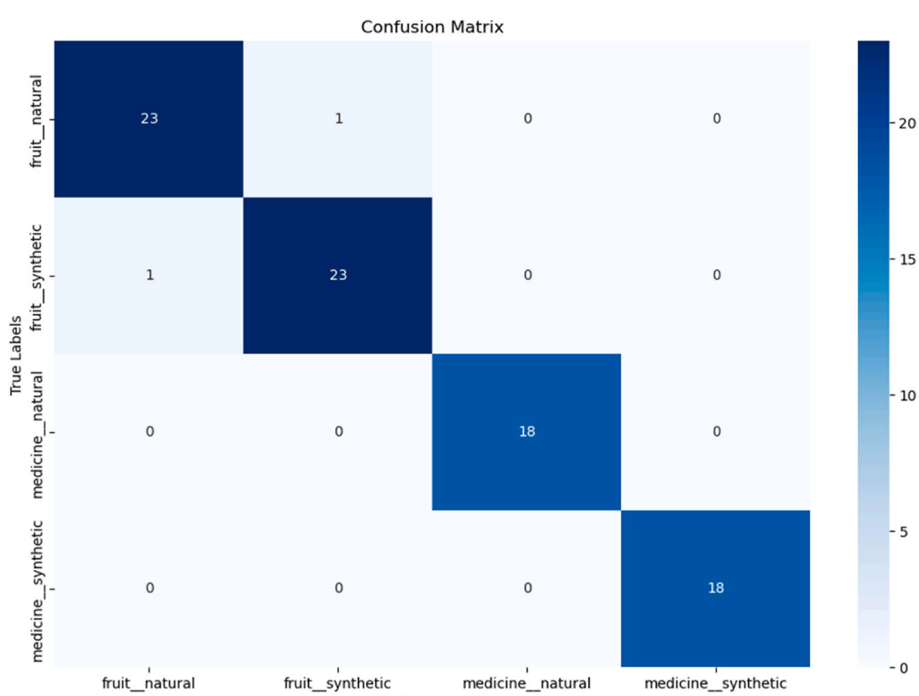

Fig. 4. Confusion matrix resulting from TabNet model analysis.

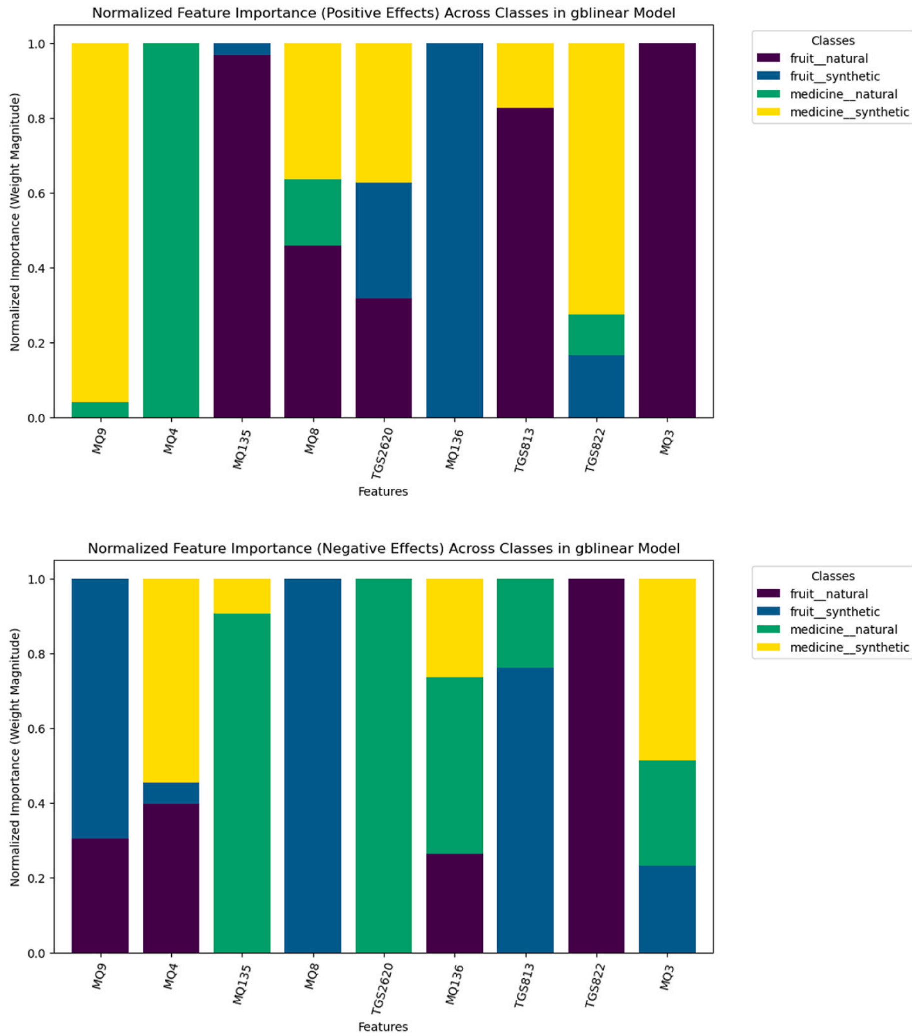

Fig. 5. a) positive and b) negative effect.

chemical fruit essential oil group. The MQ3 sensor also played the most role in the natural fruit essential oil. Therefore, these two sensors are key sensors for increasing the classification accuracy in these groups. The two sensors MQ135 and TGS813 also have a positive role in detecting natural fruit essential oil. On the other hand, the two sensors MQ9 and MQ4 also played the most role in classifying the essential oil of aromatic plants. Also, according to Fig. 5b, it can be seen that the TGS822 sensor has the most negative effect on the natural fruit essential oil group and the MQ8 sensor has the most positive effect on the chemical fruit essential oil group. Perhaps by removing these two sensors, better accuracy can be achieved in these groups. And similarly, the TGS 2620 sensor has a large negative impact on the natural medicinal essential oil group. These three sensors in Fig. 5a are sensors that have had an effect on at least three groups of essential oils. Perhaps by removing these sensors, higher accuracy can be achieved, especially in the fruit group.

This study's results reveal that the GBLinear and TabNet approach has outperformed the traditional classification methods like LDA. While employing an electronic nose for essential oil analysis, it is important to make use of models that can capture complicated non-linear relationships between the sensor responses and the volatile organic compounds (VOCs). Our research clearly illustrates how the use of these algorithms, especially in the classification of fruit-based essential oils, has had significantly better results than before.To this point, the work done on e-nose classification techniques has primarily focused on the use of the traditional machine learning methods like Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), and Random Forest (RF). For example, Rasekh and Karami [34] tried to detection of juices fraud and achieved a success rate of nearly 94 %, an outcome that was bettered with our method. Others, like. This study shows that the GBLinear model alone is superior to this method by achieving 96.9 % accuracy. Furthermore, TabNet further enhances classification performance, achieving an overall accuracy of 98.8 %. The complexity and variations of e-nose sensor response is one of the main challenges that data analyzers must deal with. Linear approaches are implemented mainly using LDA, however, fruits-based essential oils suffer from classification errors during LDA. The reason for this is the overlap between VOC profiles. As was shown in our analysis, several fruit-based samples were assigned LDA class marks in error due to composition of volatile organic compounds. That is, LDA assumptions were satisfied only to an accuracy level of 94.6 %. On the other hand, GBLinear showed improvement but only through attaining the most essential linear relationships present in the data set. While GBLinear also showed improvements counting 96,9 % in recall, the real improvement was in TabNet with 97.9 % accuracy and 99.2 % specificance. This model uses attention-based feature selection techniques deep learning classification models to overcome classification blockades. With TabNet setting, the model was able to successfully tell apart natural and synthetic fruit essential oils. In this case the model also greatly improved recall and specificity. Another important factor is the contribution of a specific sensor element within a classification category. During the conducted analysis of sensor importance, it was found that specific sensors like, MQ136 and MQ3, are much more critical than MQ9 and

MQ4 for classifying fruit essential oils and medicinal plant essential oils. Research has shown that gas sensors can often respond to multiple volatile organic compounds (VOCs), making them prone to misidentification, especially when analyzing complex mixtures [35]. This insight aligns with findings from previous research, such as Di Natale, Capuano, Quercia, Catini, Biasioli, Khomenko and Paolesse [35], which emphasized the need for selective sensor optimization to improve e-nose classification accuracy. Moreover, our results indicate that certain sensors, including TGS822 and MQ8, contributed negatively to classification performance in specific groups, suggesting that their exclusion or recalibration could further enhance model accuracy.

4. Conclusion

The combination of GBLinear and TabNet algorithms along with the selection of informative features has improved the classification of oils with the use of e-nose. TabNet with an accuracy of 98.8 %, was able to classify samples of essential oils more accurately. The accuracy of TabNet was considerably higher than that of LDA and GBLinear which was 94.6 % and 96.9 %. While all models worked well for the classification of the essential oils from the medicinal plants, it was found that TabNet algorithm did the best with incorporating essential oils derived from fruits. These results also emphasize the superiority of combining electronic noses with advanced machine learning algorithms by demonstrating their ability to fathom complex odors. The combination of GBLinear and TabNet improves accuracy and interpretability, then this method can enhance e-nose technology in the food, cosmetics, and pharmaceutical industries. As a result, this methodology contributes to algorithms combination for enhance model robustness, complete. Future research could explore the application of this methodology to a broader range of volatile compounds and investigate its potential in real-time analysis scenarios.

CRediT authorship contribution statement

Hamed Karami: Supervision, Visualization, Data curation, Resources, Writing – original draft, Validation, Software, Methodology, Conceptualization, Investigation, Formal analysis. **Abdolrahman Khoshrou:** Writing – original draft, Data curation, Software, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] J.A. Covington, S. Marco, K.C. Persaud, S.S. Schiffman, H.T. Nagle, Artificial olfaction in the 21 st century, IEEE Sens. J. 21 (11) (2021) 12969–12990.
[2] D. Karakaya, O. Ulucan, M. Turkan, Electronic nose and its applications: a survey, Int. J. Autom. Comput. 17 (2) (2020) 179–209.
[3] A.D. Wilson, M. Baietto, Advances in electronic-nose technologies developed for biomedical applications, Sensors 11 (1) (2011) 1105–1176.
[4] A. Khorramifar, H. Karami, L. Lvova, A. Kolouri, E. Łazuka, M. Piłat-Rożek, G. Łagód, J. Ramos, J. Lozano, M. Kaveh, Y. Darvishi, Environmental engineering applications of electronic nose systems based on MOX gas sensors, Sensors 23 (12) (2023) 5716.
[5] J.P. Sá, M.C.M. Alvim-Ferraz, F.G. Martins, S.I.V. Sousa, Application of the low-cost sensing technology for indoor air quality monitoring: a review, Environ. Technol. Innovat. 28 (2022) 102551.
[6] J.B.B. Rayappan, A.J. Kulandaisamy, M. Ezhilan, P. Srinivasan, G.K. Mani, Developments in electronic noses for quality and safety control, Advances in Food Diagnostics2017, pp. 63-96.
[7] H. Karami, M. Kamruzzaman, J.A. Covington, M.é. Hassouna, Y. Darvishi, M. Ueland, S. Fuentes, M. Gancarz, Advanced evaluation techniques: gas sensor networks, machine learning, and chemometrics for fraud detection in plant and animal products, Sensor Actuator Phys. 370 (2024) 115192.
[8] P. Darvishi, E. Mirzaee-Ghaleh, Z. Ramedani, H. Karami, A.D. Wilson, Detecting whey adulteration of powdered milk by analysis of volatile emissions using a MOS electronic nose, Int. Dairy J. 157 (2024) 106012.
[9] A.D. Wilson, M. Baietto, Applications and advances in electronic-nose technologies, Sensors 9 (7) (2009) 5099–5148.
[10] S. Zorpeykar, E. Mirzaee-Ghaleh, H. Karami, Z. Ramedani, A.D. Wilson, Electronic nose analysis and statistical methods for investigating volatile organic compounds and yield of mint essential oils obtained by hydrodistillation, Chemosensors 10 (11) (2022) 486.
[11] F. Wu, R. Ma, Y. Li, F. Li, S. Duan, X. Peng, A novel electronic nose classification prediction method based on TETCN, Sensor. Actuator. B Chem. 405 (2024) 135272.
[12] J.A. Covington, S. Marco, K.C. Persaud, S.S. Schiffman, H.T. Nagle, Artificial olfaction in the 21st century, IEEE Sens. J. 21 (11) (2021) 12969–12990.
[13] H.H. Afshari, S.A. Gadsden, S. Habibi, Gaussian filters for parameter and state estimation: a general review of theory and recent trends, Signal Process. 135 (2017) 218–238.
[14] X. Wang, C. Qian, Z. Zhao, J. Li, M. Jiao, A novel gas recognition algorithm for gas sensor array combining savitzky–golay smooth and image conversion route, Chemosensors 11 (2) (2023) 96.
[15] F. Auger, M. Hilairet, J.M. Guerrero, E. Monmasson, T. Orlowska-Kowalska, S. Katsura, Industrial applications of the Kalman filter: a review, IEEE Trans. Ind. Electron. 60 (12) (2013) 5458–5471.
[16] H. Karami, B. Thurn, N.K. de Boer, J. Ramos, J.A. Covington, J. Lozano, T. Liu, W. Zhang, S. Su, M. Ueland, Application of gas sensor technology to locate victims in mass disasters – a review, Nat. Hazards (2024).
[17] N.S. Aghili, M. Rasekh, H. Karami, O. Edriss, A.D. Wilson, J. Ramos, Aromatic fingerprints: VOC analysis with E-Nose and GC-MS for rapid detection of adulteration in sesame oil, Sensors 23 (14) (2023) 6294.
[18] X. Wang, Y. Zhou, Z. Zhao, X. Feng, Z. Wang, M. Jiao, Advanced algorithms for low dimensional metal oxides-based electronic nose application: a review, Crystals 13 (4) (2023) 615.
[19] M. Rasekh, H. Karami, S. Fuentes, M. Kaveh, R. Rusinek, M. Gancarz, Preliminary study non-destructive sorting techniques for pepper (Capsicum annuum L.) using odor parameter, LWT 164 (2022) 113667.
[20] N. Mohammadian, A.M. Ziaiifar, E. Mirzaee-Ghaleh, M. Kashaninejad, H. Karami, Gas sensor technology and AI: forecasting lemon juice quality dynamics during the storage period, J. Stored Prod. Res. 109 (2024) 102449.
[21] T. Liu, L. Guo, M. Wang, C. Su, D. Wang, H. Dong, J. Chen, W. Wu, Review on algorithm design in electronic noses: challenges, status, and trends, Intelligent Computing 2 (2023) 12.
[22] H.-J. He, d.S.F.M. Vinicius, W. Qianyi, K. Hamed, M. Kamruzzaman, Portable and miniature sensors in supply chain for food authentication: a review, Crit. Rev. Food Sci. Nutr. 1-21.
[23] C. Wade, K. Glynn, Hands-On Gradient Boosting with Xgboost and scikit-learn: Perform Accessible Machine Learning and Extreme Gradient Boosting with Python, Packt Publishing Ltd2020.
[24] S. Karimi, M. Asghari, R. Rabie, M. Emami Niri, Machine learning-based white-box prediction and correlation analysis of air pollutants in proximity to industrial zones, Process Saf. Environ. Prot. 178 (2023) 1009–1025.
[25] S. Dhiman, A. Thukral, P. Bedi, Impact of clinical features on disease diagnosis using knowledge graph embedding and machine learning: a detailed analysis, in: A. Verma, P. Verma, K.K. Pattanaik, S.K. Dhurandher, I. Woungang (Eds.), Advanced Network Technologies and Intelligent Computing, Springer Nature Switzerland, Cham, 2024, pp. 340–352.
[26] S.Ö. Arik, T. Pfister, Tabnet: attentive interpretable tabular learning, Proc. AAAI Conf. Artif. Intell. (2021) 6679–6687.
[27] C. Shah, Q. Du, Y. Xu, Enhanced TabNet: attentive interpretable tabular learning for hyperspectral image classification, Remote Sens. 14 (3) (2022) 716.
[28] K. McDonnell, F. Murphy, B. Sheehan, L. Masello, G. Castignani, Deep learning in insurance: accuracy and model interpretability using TabNet, Expert Syst. Appl. 217 (2023) 119543.
[29] M. Rasekh, H. Karami, A.D. Wilson, M. Gancarz, Performance analysis of MAU-9 electronic-nose MOS sensor array components and ANN classification methods for discrimination of herb and fruit essential oils, Chemosensors 9 (9) (2021) 243.
[30] H. Karami, M. Rasekh, E. Mirzaee-Ghaleh, Application of the E-nose machine system to detect adulterations in mixed edible oils using chemometrics methods, J. Food Process. Preserv. 44 (9) (2020) e14696.
[31] T. Chen, C. Guestrin, Xgboost: a scalable tree boosting system, Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (2016) 785–794.
[32] H. Karami, M. Rasekh, E. Mirzaee – Ghaleh, Comparison of chemometrics and AOCS official methods for predicting the shelf life of edible oil, Chemometr. Intell. Lab. Syst. 206 (2020) 104165.
[33] H. Karami, S. Karami Chemeh, V. Azizi, H. Sharifnasab, J. Ramos, M. Kamruzzaman, Gas sensor-based machine learning approaches for characterizing tarragon aroma and essential oil under various drying conditions, Sensor Actuator Phys. 365 (2024) 114827.
[34] M. Rasekh, H. Karami, Application of electronic nose with chemometrics methods to the detection of juices fraud, J. Food Process. Preserv. 45 (5) (2021) e15432.
[35] C. Di Natale, R. Capuano, L. Quercia, A. Catini, F. Biasioli, I. Khomenko, R. Paolesse, AR1. 3-Real time proton transfer reaction and electronic nose

simultaneous measurements on same samples, Proceedings IMCS 2018 (2018) 229–230.