Development of a digital nose system for early detection of plant stress

Georg R. Schneider[1], Josef Scharinger[2], Eva O. Huber[1], Thomas Penkner[1], and Claudia Probst[1]

[1] The University of Applied Sciences Upper Austria, Stelzhamerstraße 23, Wels, Austria
georg.schneider@fh-wels.at
[2] Institute for Computational Perception, Johannes Kepler University, Linz, Austria

Abstract. Early forest stress detection is essential for the preservation of healthy forest ecosystems. The aim of this project is to develop an electronic nose (e-nose) using metal oxide (MOx) gas sensors that can differentiate between stressed and healthy trees by detecting volatile organic compounds (VOCs). An Arduino microcontroller was used to collect data from the gas sensors, while Python was implemented for data processing. The system applied machine learning algorithms such as Linear Discriminant Analysis (LDA), as a supervisded learning method, and Principal Component Analysis (PCA), as an unsupervisded learning method, to classify and perform dimensionality reduction on the sensor data. To enhance portability and usability, a printed circuit board (PCB) was designed, creating a compact and efficient e-nose for field testing. The sensor array was tested with various materials found in stressed trees. PCA was also applied to assess sensor sensitivity and evaluate sensor configurations. Initial results demonstrated the e-nose's ability to distinguish between diseased and healthy trees with significant accuracy. PCA showed good separation of VOC patterns but lower accuracy when detecting multiple target gases. LDA provided clearer distinctions between the two classes with minimal overlap. Although MOx sensors exhibited high sensitivity, their low selectivity for specific gases affected classification accuracy. The high sensitivity of MOx sensors often comes at the expense of selectivity. Future research will focus on identifying specific VOCs emitted by stressed trees using neural networks and improving the e-nose's ability to detect a wider range of compounds.

Keywords: electronic nose, MOx gas sensors, VOCs, European spruce, LDA, PCA, Python, machine learning

1 Introduction

Trees emit volatile organic compounds (VOCs) during normal activities, but stress from diseases or environmental changes can significantly alter VOC concentrations. Electronic nose (e-nose) technology has emerged as a promising method to monitor forest health by detecting these odor fingerprints. An e-nose employs an array of gas sensors, mimicking the human olfactory system, to identify complex odors. This technology has been widely applied in fields like food quality control and medical diagnostics and is now being explored for environmental monitoring. While gas chromatography-mass spectrometry (GC-MS) remains the gold standard for VOC detection due to its accuracy, its complexity and operational challenges make real-time monitoring impractical. Metal oxide (MOx) sensors provide a cost-effective alternative, offering high sensitivity to various gases. By detecting changes in the resistance of the metal oxide layer, MOx sensors are well-suited for environmental applications, including air quality and forest health monitoring. Their growing use across industries highlights their effectiveness and accessibility. [5, 9, 14, 7].

2 Basics and details

2.1 Basics of electronic nose systems

An electronic nose is a sensor device that mimics the human olfactory system by detecting and recognizing odors through a gas sensor array. It combines data acquisition and pattern recognition algorithms to identify unique volatile compounds emitted by different chemicals.

The basic workflow includes:

1. Sample Collection: Odor compounds are delivered to the sensor array.
2. Detection: Sensors react to changes in the medium.
3. Data Processing: Raw data is normalized, filtered, and features are extracted.
4. Pattern Recognition: Algorithms like PCA, neural networks, or LDA analyze the data to create identifiable fingerprints [1].

2.2 MOx sensor technology

Metal-oxide semiconductor (MOx) sensors are resistive sensors widely used for detecting volatile organic compounds (VOCs) due to their low cost, high efficiency, fast response, and recovery times. They operate by measuring changes in electrical resistance when exposed to gases at elevated temperatures (150–500°C).

A MOx sensor consists of a metal-oxide sensing layer, a substrate with integrated electrodes for resistance measurement, and a heating element to maintain stable operating conditions. In n-type semiconductors, resistance decreases with reducing gases (e.g., H_2, CO, alcohols) and increases with oxidative gases (e.g., NO, ozone).

These sensors are commonly applied in agriculture and forestry to detect plant infections, pest damage, and other environmental stresses. However, they are sensitive to humidity, which is managed by incorporating a heater to ensure controlled conditions. The key parameter of MOx sensors is the resistance of the sensitive layer, which changes depending on the surrounding gas. This resistance is used to identify specific gases.

Resistance (R) is measured using the voltage divider principle with a load resistor (R), a simple and effective method to convert resistance changes into voltage signals. Figure 1 shows a simplified equivalent circuit of a gas sensor. The gas sensor's circuit includes heater resistance (R_h), sensitive layer resistance (R_{sl}), and load resistance (R) [6, 7, 4].

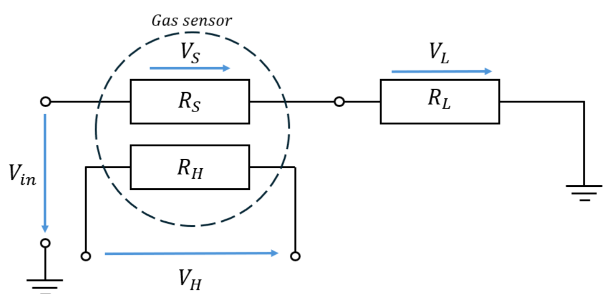

Fig. 1. Equivalent circuit of a gas sensor

The decision to use MOx sensors is based on their comparatively low cost, easy handling and uncomplicated design. Although MOx sensors are less sensitive than Quartz Crystal Microbalance (QCM) and Surface acoustic wave (SAW) sensors, the sensors currently in use achieve satisfactory results [15–17]. An enhanced version of the electronic nose with optimized sensors is currently in the development phase.

2.3 Machine learning

Machine learning involves algorithms and statistical methods that identify patterns in data and use them for predictions or decision-making [13]. It is categorized into two main types: supervised and unsupervised learning [1, 18, 19].

Supervised Learning maps inputs to outputs using labeled data (training set) to predict unseen outputs. Its objectives include:

- Classification: Predicting categorical outputs (e.g., healthy or unhealthy trees).
- Regression: Predicting continuous outputs (e.g., future values). [10–12]

Linear Discriminant Analysis (LDA) LDA is a supervised dimensionality reduction technique that identifies a linear combination of features to best separate classes in a dataset. It projects the data into a lower-dimensional space while preserving information most relevant to class discrimination.

LDA works by maximizing the distance between the means of different classes and minimizing the spread of data points within each class.

Similar to PCA, LDA creates linear combinations of features, but its focus is on maximizing class separation rather than variance [1, 8]

The sample taking set up is visualized in Figure 2 below.

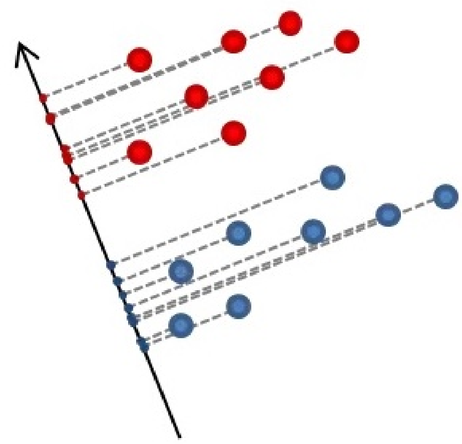

Fig. 2. Linear Discriminant Analysis [3]

Unsupervised Learning identifies hidden patterns without labeled outputs, enabling faster data processing. Its objectives include:

- Clustering: Grouping data with similar features.
- Dimensionality Reduction: Reducing input features while retaining key information [1].

Principal Component Analysis (PCA) PCA is an unsupervised learning method for dimensionality reduction. It creates new variables (principal components) as linear combinations of the original variables, retaining significant features while reducing complexity. The first principal component (PC1) captures the largest variance, followed by PC2, and so on [2]. This reduction simplifies data analysis and visualization while maintaining essential information, as can be seein in Figure 3.

[3] https://vivekmuraleedharan73.medium.com/what-is-linear-discriminant-analysis-lda-7e33ff59020a

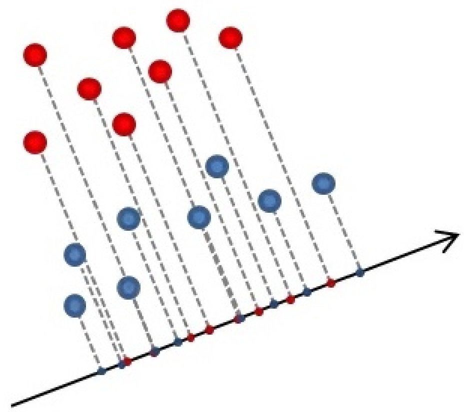

Fig. 3. Principal Component Analysis [4]

2.4 Python

Python, a versatile and user-friendly programming language, was chosen for the electronic nose project due to its suitability for scientific research and data processing. It facilitated data handling and machine learning implementation through its extensive libraries. Key libraries used include:

- NumPy and Pandas: For multi-dimensional array processing and data manipulation.
- Scikit-learn: For machine learning algorithms like classification and dimensionality reduction.
- Serial: For communication between Arduino and the PC.

Python's simplicity, flexibility, and robust library ecosystem make it ideal for analyzing data generated by the electronic nose [3].

2.5 Objective

This project aims to develop a cost-effective electronic nose system that integrates gas sensor technology with machine learning for real-time tree health monitoring. Serving as a reliable alternative to expensive GC-MS, the system has potential applications in environmental and agricultural fields.

By enabling early detection of diseases, pests, and environmental stress, this approach could enhance forest health management, promote sustainability, and improve ecosystem resilience. Additionally, the insights gained may support broader efforts in environmental conservation and sustainable agriculture.

[4] https://vivekmuraleedharan73.medium.com/what-is-linear-discriminant-analysis-lda-7e33ff59020a

3 Materials and Methods

3.1 Sensor selection

The VOC detection focused on identifying compounds like α-pinene, 3-carene, and d-limonene, which increase in concentration during tree stress. Additionally, ethanol, 2-methyl-3-buten-1-ol, and verbenol were studied as potential indicators of insect infestations, such as bark beetle attacks on spruce trees [4, 5].

To detect these VOCs, MOx sensors primarily sensitive to alcohols were selected due to their simplicity and cost effectiveness. The sensors were arranged in sets of four or pairs to enhance stability. While MOx sensors were used in this study, alternative sensor types, such as electrochemical or PID sensors, could also be applied for VOC detection. A summary of the selected sensors and their target gases is presented in Table 1.

Table 1. List of sensor types

Sensor	Sensitive to	Quantity
MQ-3	CO, Alcohol, Methane	4
MQ-135	CO, CO_2, Alcohol, Acetone	4
UST GGS-1330	CO, H_2, Methane	2
UST GGS-2330	CO, Ethanol, Methane	2
UST GGS-10330	CO, H_2, Butane	2
Figaro TGS-2600	CO, Ethanol, H_2, Methane	2
Figaro TGS-822	CO, Ethanol, Acetone	2

Table 1 presents a diverse selection of sensors used to detect various substances, ensuring a broad range of information is captured. However, this results in a large volume of collected data. To facilitate analysis, the data must be summarized before further processing.

To reduce the number of sensors, a Random Forest model was used to generate a feature importance chart. Among the sensors, only GGS-2330 stood out, showing low importance in both substance detection and tree measurements.This suggests that removing GGS-2330 would not impact the e-nose's ability to detect relevant VOCs.

Another approach to sensor reduction would be using a single unit of each sensor instead of sets, though this could affect system stability.

Additionally, ENS160 and BME280 sensors were implemented for monitoring temperature, humidity, and total VOC levels. It is important to note that Table 1 lists the gases to which these sensors are most sensitive, but other gases may also influence their responses.

3.2 Experimental setup

The experimental setup included two distinct odor measurement systems: one for tree analysis and the other for substance detection. Figure 5 shows the tree measurement setup, which includes two electronic noses: Smell Inspector by SmartNanotubes and Digi-nose. Smell Inspector was used as a reference to guarantee the accuracy of Digi-nose results. The design included push and pull pumps for air circulation. Initially, the chamber was left empty for 15 minutes to create a baseline. Following that, a tree was introduced, and its scents were investigated for 40 minutes. Twelve trees were evaluated in total, divided into three categories: healthy, dry-stressed and overwatered.

The substance measuring setup, depicted in Figure 4, had identical components pumps, a chamber, sensors, and a PC but was intended to identify specific substances. Following a 15-minute baseline air measurement, ethanol and d-limonene were added, and measurements were conducted for 10 minutes.

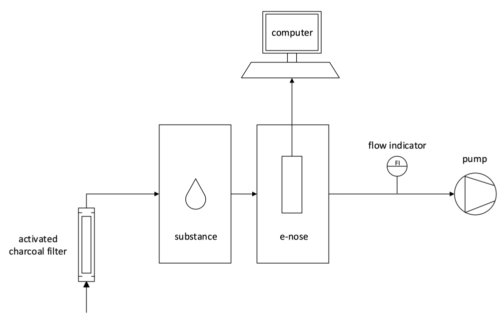

Fig. 4. Measurement setup for substance measurement

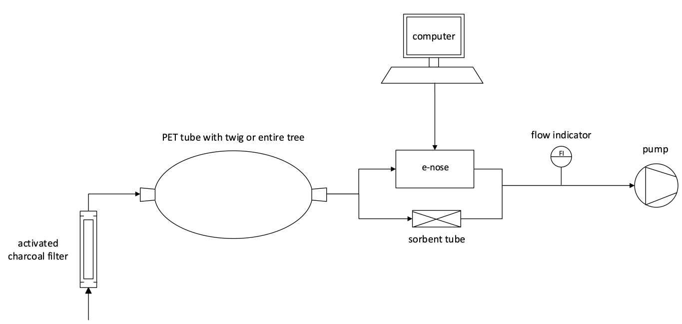

Fig. 5. Measurement setup for tree measurement

3.3 Designing the electronic nose

The sensor array was composed of eighteen gas sensors, with the quantities of each type listed in Table 1. Each sensor voltage is read in by Arduino Due and converted to resistance value by voltage divider formula. Data of the same sensor type is averaged and stored. For further processing, data is transmitted as a string. A data set is sent every three seconds from Arduino to PC with a dimensions of 1x12 in form seen in Figure 6. Data is collected and stored in a CSV file.

```
Time,Temperature,Humidity,CO2,TVOC,MQ3,MQ135,GGS1330,GGS2330,GGS10330,TGS2600,TGS822
```

Fig. 6. Structure of string data sent from Arduino to PC

3.4 Data aquisition

The primary objective of this study was to collect data from selected sensors and utilize it to validate the sensor technology. The acquired data was analyzed to assess sensor performance, accuracy, and reliability in detecting specific chemical compounds and environmental conditions. For analysis, the collection of data included measurements of α-pinene, 3-carene, d-limonene, ethanol, 2-methyl-3-buten-1-ol, verbenol, air, and additionally, ethanol-verbenol solution and mixture of all substances together. Each of these was measured 10-15 times, equaling to around 120 measurements for data analysis. Measurement of trees led to 30 measurements, where 15 measurements were healthy trees and 15 measurements were unhealthy trees.

3.5 Data processing

The data obtained from the sensors underwent preprocessing to ensure accuracy and consistency before further evaluation. The raw data was initially examined for completeness, and any missing or erroneous values were addressed using appropriate statistical methods. Since temperature, humidity, TVOC, and CO2, were not used in LDA, it was dropped, leaving only data from MOx sensors for LDA analysis. Prior to LDA, data is split into training and testing subsets, where the training set is normalized by MinMaxScaler. This way, any data leakage was prevented. In this analysis, the training set consisted of 80% of data, and the rest is used in the testing set. Data is split randomly each run. For validation of the LDA result, accuracy is computed. Due to the balanced data set, accuracy provides a reliable measure of model performance.

4 Results

After the data collection and training the verification process was conducted. A PCA algorithm provided a valuable insight into dimensionality reduction of data. However, a separation of data has yielded results with high overlapping. Therefore, an algorithm called LDA has been applied due to its ability to not only reduce dimensionality but also to separate data into classes. Because of this, all further results will focus on LDA.

4.1 Digital nose

Looking at Figure 7, one can notice quite a good separation of sick and healthy trees. The cause might be that some trees are really as sick as others or that healthy trees are really as healthy. A training set contained 80% of the data and yielded an accuracy score of 75%. The testing data set contained the rest of the data and has shown better separation, with an accuracy score of 71%. A testing set is demonstrated in Figure 7. Only a small amount of data was used for analysis, such as 27 sick trees and 14 healthy trees. Unfortunately, only this amount of data could be used at the time of analysis.

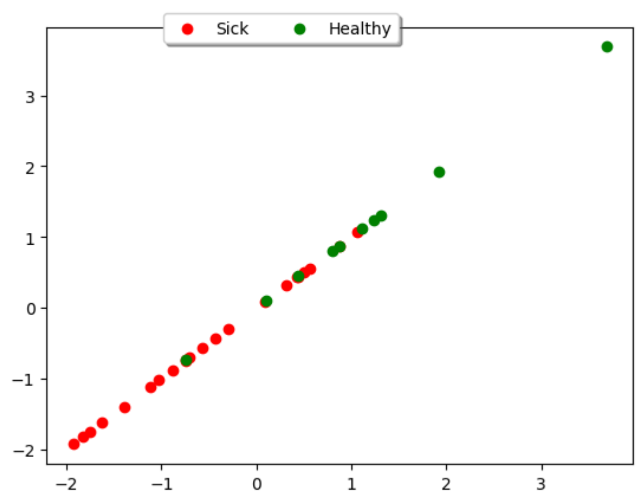

Fig. 7. LDA Analysis sick trees and health trees

Looking at Figure 8 showed also a good separation of sick, dead and healthy trees. The data set including a small amount of data was used for analysis, such as 16 sick trees, 5 dead trees and 11 healthy trees. Unfortunately, only this amount of data could be used at the time of analysis. In further were are separating the sick trees in water stress and dry stress.

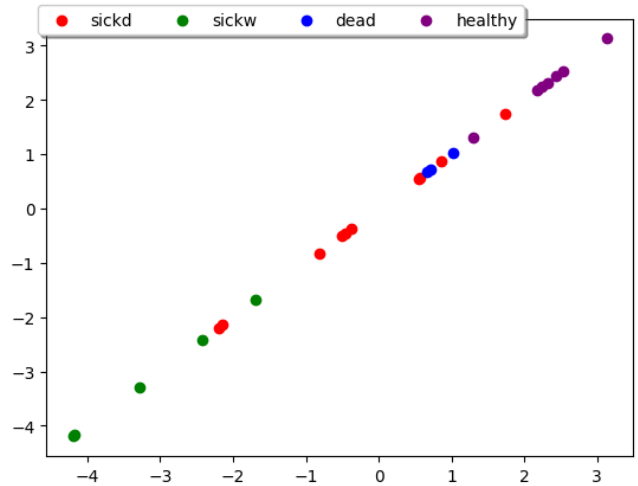

Fig. 8. LDA Analysis sick trees, health trees and dead trees

4.2 LDA analysis

Worth noting is that the data set is small. Further testing is aimed at increasing the data set for better and more efficient predictions. In the analysis of substance data, some of the substances have been able to be classified, while the classification of others was challenging. Substances like d-limonene, ethanol, and air clearly separated, also sick trees good separated. This problem occurs due to the nature of MOx sensors, where their resistance can give the same results for different odors with different concentrations. LDA analysis of substances and sick trees is presented in Figure 9 . Similar to tree analysis, 80% of the data was used for training and 20% for testing set.

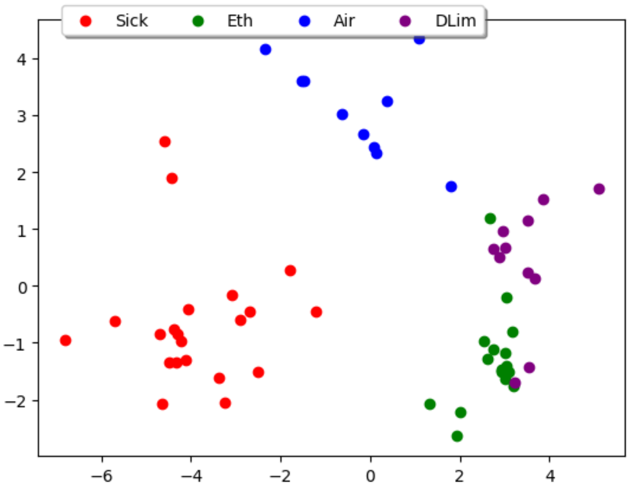

Fig. 9. LDA Analysis sick trees and pure substances

4.3 PCA analysis

We also conducted a PCA analysis with the same small data source. Further testing aims to increase the dataset for better and more efficient predictions. In the analysis of substance data, some substances could be successfully classified, while others posed a challenge. Substances such as d-limonene has a small overlapping, ethanol, and emissions from sick trees showed clear separability from air alone, but also exhibited some overlap. This issue arises due to the nature of MOx sensors, whose resistance can produce similar results for different odors at varying concentrations.

PCA analysis of substances is presented in Figure 10 and Figure 11. Similar to tree analysis, 80% of the data was used for training and 20% for testing set.

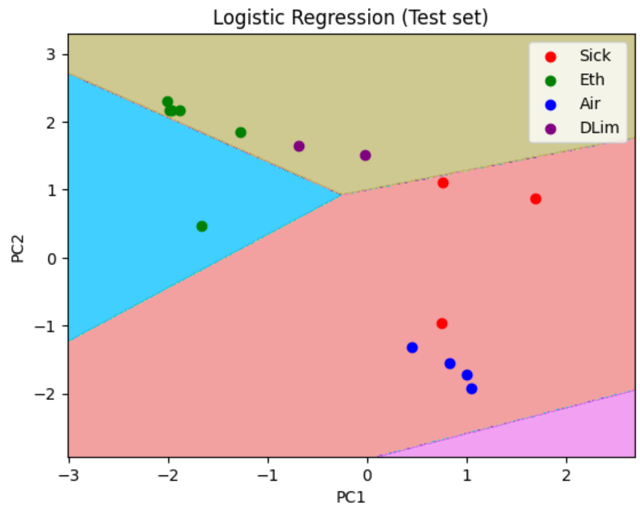

Fig. 10. PCA Analysis sick trees and pure substances with decision boundaries

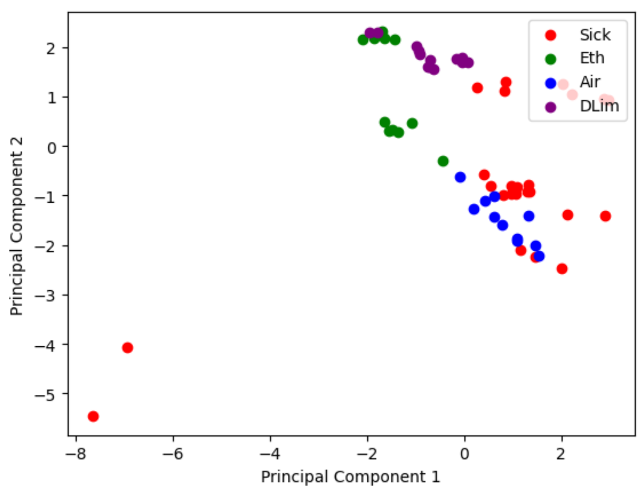

Fig. 11. PCA Analysis sick trees and pure substances

4.4 Random Forest Classifier

In an additional test called random forest and in combination with a heatmap, used the same small data source. The results also demonstrated that having a better data source is a significant advantage.

In the analysis of substance data, some of the substances have been able to be classified, while the classification of others was challenging. Substances like d-limonene, sick trees, and air clearly separated, while ethanol created overlapping.

This problem occurs due to the nature of MOx sensors, where their resistance can give the same results for different odors with different concentrations. Random Forest Classifier analysis of substances is presented in Figure 12 and Figure 13. Similar to tree analysis, 80% of the data was used for training and 20% for testing set. Training set achieved an accuracy score of 70%, while the testing set accuracy decreased to 81%.

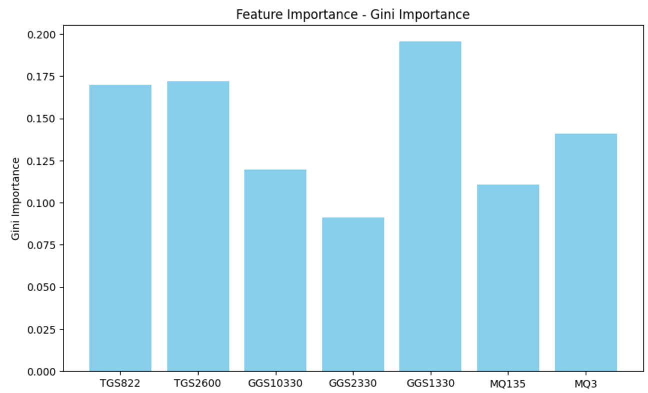

Fig. 12. Random forest classification based on different sensors

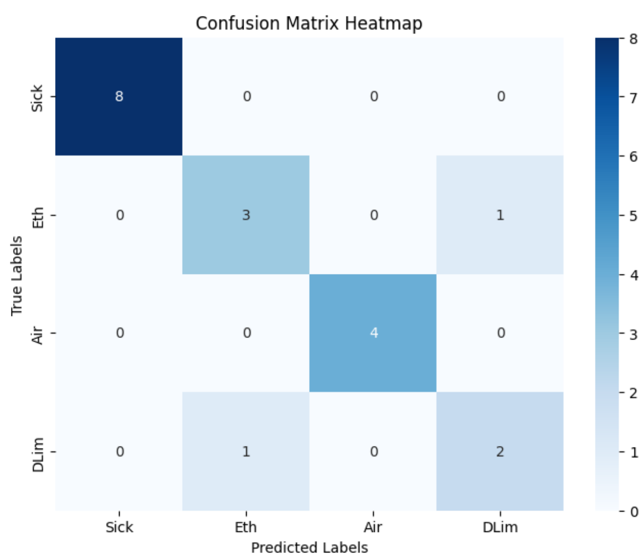

Fig. 13. Confusion matrix based on sick trees and pure substances

5 Conclusion

In this project, an electronic nose using metal-oxide gas sensors to differentiate between healthy and sick trees by distinguishing unique volatile organic compounds (VOCs) was successfully implemented. The e-nose system, together with an Arduino Due microcontroller, further improved by Python data analysis, proved an effective way for detecting and classifying VOCs related to forest health. Our methodology included developing a gas sensor array that would be capable of detecting a broad range of different odors and providing information on how healthy trees are. The procedure involved the application of PCA and LDA for data analysis and the design of a graphical user interface for real-time monitoring. Initial outcomes demonstrated the ability to show a difference between healthy and ill trees with a viable level of accuracy. While the PCA provided initial separation of data, the LDA provided clearer class separation, with overlapping in some cases.

A key issue was the low selectivity of MOx sensors. While they offer high sensitivity, they lack the ability to uniquely identify specific VOCs. This limitation reduces diagnostic accuracy, especially under real environmental conditions where external factors such as temperature and humidity can influence measurements. Therefore, integrating a temperature compensation algorithm into future system versions is necessary to reduce measurement deviations.

Another important aspect is the miniaturization and field usability of the system. The development of a compact printed circuit board (PCB) is a step in the right direction. However, further validation in field conditions is required to confirm the practical applicability of the e-nose.

Despite the overall promising results, the limited selectivity of the MOx sensors hampers the differentiation in some cases. A future study will focus on solving this problem by implementing an artificial neural network to enhance e-noses' selectivity for other gases. In conclusion, e-nose represents an optimistic and cost-effective solution to more prevalent approaches such as GC-MS for real-time monitoring of VOCs. This technique has the potential to be applied not only for forest health monitoring, but also for agricultural applications, contributing to environmental sustainability while making monitoring options more accessible. Further improvements in sensor selectivity and machine learning algorithms will be necessary to fully explore the potential of this newly developed technology.

6 Outlook

Future work shall focus on the development of a temperature compensation algorithm to utilize temperature data and adjust MOx sensor output data. This would improve the e-nose's ability to operate in different environmental conditions. While LDA has shown promising results, the application of other algorithms, such as Random Forest, could help address the problem of overfitting, as it is less prone to it. Another approach would be to increase the dataset through additional measurements or by data augmentation (such as applying noise to the existing data).

Finally, incorporating supplementary evaluation metrics, such as the F1-score and detailed confusion matrix analyses, will provide a more comprehensive assessment of classification performance, further bolstering the system's reliability in both forestry and agricultural applications.

References

1. Bishop, C.M., Nasrabadi, N.M.: Pattern Recognition and Machine Learning. Springer, New York (2006)
2. Jaadi, Z.: Principal Component Analysis (PCA): A Step-by-Step Explanation. https://builtin.com/data-science/step-step-explanation-principal-component-analysis (2025). Accessed 13 Mar 2025
3. Digitalogy: NodeJs vs Python: Which is Better for Backend Development? https://www.digitalogy.co/blog/nodejs-vs-python-for-backend-development (2023). Accessed 13 Mar 2025
4. Eranna, G.: Metal Oxide Nanostructures as Gas Sensing Devices. CRC Press, Boca Raton (2011)
5. Gardner, J.W., Persaud, K.C.: Electronic Noses and Olfaction 2000: Proceedings of the 7th International Symposium on Olfaction and Electronic Noses, Brighton, UK, July 2000. CRC Press, Boca Raton (2001)
6. Jaaniso, R., Tan, O.K.: Semiconductor Gas Sensors. Elsevier, Amsterdam (2013)
7. Khorramifar, A., Karami, H., Lvova, L., Kolouri, A., Łazuka, E., Piłat-Rożek, M., Łagód, G., Ramos, J., Lozano, J., Kaveh, M., et al.: Environmental engineering applications of electronic nose systems based on MOX gas sensors. Sensors 23(12), 5716 (2023)
8. Wang, Y., Zhang, W., Gao, R., Jin, Z., Wang, X.: Recent advances in the application of deep learning methods to forestry. Wood Science and Technology 55(5), 1171–1202 (2021)
9. Yadav, A., Yadav, K., Ahmad, R., Abd-Elsalam, K.A.: Emerging frontiers in nanotechnology for precision agriculture: Advancements, hurdles and prospects. Agrochemicals 2(2), 220–256 (2023)
10. Zampieri, G., Vijayakumar, S., Yaneske, E., Angione, C.: Machine and deep learning meet genome-scale metabolic modeling. PLoS Computational Biology 15(7), e1007084 (2019)
11. Ben Lazreg, M., Goodwin, M., Granmo, O.-C.: Information abstraction from crises related tweets using recurrent neural network. In: IFIP International Conference on Artificial Intelligence Applications and Innovations, pp. 441–452. Springer, Cham (2016)
12. Eliassi-Rad, T.: Building Intelligent Agents that Learn to Retrieve and Extract Information. The University of Wisconsin-Madison, Madison (2001)
13. Ismail, I.A., Chukwuka, E.E., Abiola, M.K.: Improving business insights using machine learning algorithms. GSJ 12(2) (2024)
14. Nijhof, B., Castells-Nobau, A., Wolf, L., Scheffer-de Gooyert, J.M., Monedero, I., Torroja, L., Coromina, L., van der Laak, J.A.W.M., Schenck, A.: A new Fiji-based algorithm that systematically quantifies nine synaptic parameters provides insights into Drosophila NMJ morphometry. PLoS Computational Biology 12(3), e1004823 (2016)
15. R. Suzuki, T. Emura, Y. Tokura, N. Kawamura, and Y. Hori, A Quartz Crystal Microbalance Based Portable Gas Sensing Platform for On-Demand Human Breath Monitoring, *IEEE Access*, vol. 8, pp. 146166–146171, 2020. https://doi.org/10.1109/ACCESS.2020.301385710.1109/ACCESS.2020.3013857
16. F. Kus, C. Altinkok, E. Zayim, S. Erdemir, C. Tasaltin, and I. Gurol, Surface acoustic wave (SAW) sensor for volatile organic compounds (VOCs) detection with calix[4]arene functionalized gold nanorods (AuNRs) and silver nanocubes (AgNCs), *Sensors and Actuators B: Chemical*, vol. 330, p. 129402, 2021. https://doi.org/10.1016/j.snb.2020.12940210.1016/j.snb.2020.129402

17. D. Nuhoglu, C. Tasaltin, and I. Gurol, QCM and SAW gas and VOC sensors based on metal oxide composites (principles, fabrication, sensing materials, and performances), in *Complex and Composite Metal Oxides for Gas, VOC, and Humidity Sensors*, B. C. Yadav and P. Kumar, Eds., Elsevier, 2024, pp. 223–252. https://doi.org/10.1016/B978-0-323-95385-6.00016-710.1016/B978-0-323-95385-6.00016-7 @
18. Y. Feng, Y. Wang, B. Beykal, M. Qiao, Z. Xiao, and Y. Luo, A mechanistic review on machine learning-supported detection and analysis of volatile organic compounds for food quality and safety, *Trends in Food Science & Technology*, vol. 143, p. 104297, 2024.
19. S. Singh, S. Sajana, P. Varma, G. Sreelekha, C. Adak, R. P. Shukla, and V. B. Kamble, Metal oxide-based gas sensor array for VOCs determination in complex mixtures using machine learning, *Microchimica Acta*, vol. 191, no. 4, p. 196, 2024.