Digi-Nose Part 2: Enhancing Accuracy and Efficiency of a Digital Nose System With Sensor Technology for Early Detection of Changes in the Forest

Leo Bilješko[1,*,†], Georg Roman Schneider[1] and Claudia Probst[1]

[1]*FH Oberösterreich University of Applied Sciences Upper Austria, Roseggerstraße 15, 4600 Wels, Austria*

Abstract

The detection of forest health has become significant for maintaining the forest environment, especially now in the time of increasing ecological stressors. The objective of this project is to design an electronic nose (e-nose) using metal-oxide (MOx) gas sensors to be able to distinguish between healthy and stressed trees by detecting unique volatile organic components (VOCs). The project involved the development and implementation of a gas sensor array, combining multiple MOx sensors, to detect VOCs. Taking advantage of the Arduino microcontroller, data was able to be received from gas sensors, while Python was utilized for data analysis. Data analysis involved machine learning methods, such as Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA), for classification and dimensionality reduction of the sensor data. Python also came in handy for the creation of graphical user interfaces. The capacity of the e-noses to differentiate between healthy and sick trees was demonstrated in the initial results, where it showed a reasonable level of accuracy. Initially, PCA provided good separation, however, with an increased number of target gases, the separation accuracy deteriorated. The LDA provided a clear separation between two classes, with slight overlaps. The e-nose was further assessed for different substances that may be present in stressed trees. Although it has shown the good separability of some substances, others overlapped. The great sensitivity of the MOx sensor comes with a cost of selectivity for different gases. Future research will focus on detecting these specific substances contained in the tree's odor using a neural network, enhancing the electronic nose's ability to detect a wider range of compounds.

Keywords

e-nose, VOCs, MOx gas sensor, machine learning

1. INTRODUCTION

In recent years, the detection of forest health has become critical for maintaining the forest environment and agricultural richness. Trees, like other living species, produce volatile organic compounds (VOCs) as part of their normal everyday activities. When stress is initiated, such as from diseases or environmental changes like drought and deluge, the concentration of VOCs that trees produce can be increased significantly. Recent studies of electronic nose technology have unlocked new alternative approaches to monitoring forest health through the detection of odor fingerprints.

Conventionally, Gas Chromatography-Mass Spectrometry (GC-MS) has been a standard for the detection of VOCs. Despite the fact that it offers high accuracy and sensitivity, GC-MS requires higher complexity and skills to operate. Due to these challenges, a real-time observation of a tree's health would be complex. Metal-oxide (MOx) sensors offer a great, accessible, and cost-effective alternative due to their wide range of sensitivity to various gases. The change of the gas or the concentration of the gas is detected by the metal-oxide layer's resistance, which makes it suitable for environmental monitoring. MOx gas sensor technology has been increasing in recent years in many applications, including air quality control, food quality monitoring, and medical applications, featuring its effectiveness in many fields [1][4].

In this study, an electronic nose system was developed for detecting the health level of a tree using MOx sensors. A system was developed to gather and analyze the odors emitted by stressed and unstressed conditions. Machine learning methods such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) have been employed to categorize different smells from trees and create a fingerprint for their detection. The process also involved the design of a graphical user interface (GUI) to enable real-time monitoring of changes.

By merging advanced sensor technology and machine learning methods, the goal is to provide a reliable, cost-effective tool for early detection of trees' health. Furthermore, this system has the potential to be applied to other agricultural practices and thereby contribute to environmental sustainability. This approach aims to offer an alternative to a more common GC-MS, taking advantage of cost-effective and more accessible MOx sensors, to enable widespread, real-time monitoring of tree health.

2. THEORY

2.1. Metal-oxide (MOx) gas sensors

MOx sensors, which are resistive sensors made from metal-oxide semiconductors, have been accepted as a promising possibility to develop low-cost, highly efficient sensors with fast response and recovery times. Due to their simple mechanism for measuring resistance and their high sensitivity to different gases, these devices are often used to detect volatile compounds (VOCs) at ppm and sub-ppm levels. These sensors have applications in the agriculture and forestry industries to detect plant infections caused by fungus, bacteria, and viruses, as well as damage caused by insects or mechanical means.

The working principle of the MOX device is based on the changes in the electrical resistance of a metal oxide semiconductor when it encounters gases at elevated temperatures (150-500 °C) [2]. The MOx sensors are susceptible to humidity and temperature, so the addition of a heater allows it to operate at controlled conditions and temperatures [2].

In this work, MOx sensor are a main type of sensors used to develop a possible solution for the detection of forest health. Their implementation will be discussed in further sections.

2.1.1. Resistance measurement

The important parameter of MOx sensors is the resistance of the sensitive layer. Depending on the type of gas surrounding the sensor, resistance varies. Based on this change, the goal is to define unique resistance for a specific gas.

The measurement of sensitive layer resistance (R_S) is typically measured using a voltage divider principle with a load resistor (R_L). This approach is favored for its simplicity and effectiveness in converting resistance changes into measurable voltage signals. Figure 1 shows a simplified equivalent circuit of a gas sensor, consisting of heater resistance R_H, sensitive layer resistance R_S, and load resistance R_L.
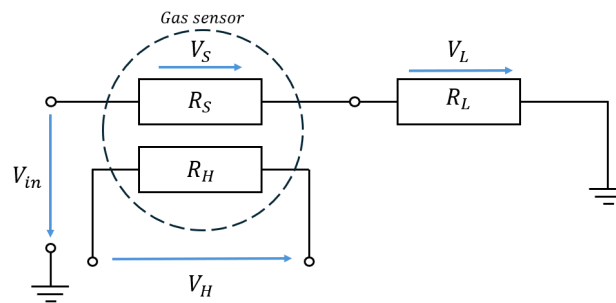

Figure 1: Equivalent circuit of a gas sensor.

2.2. Machine learning algorithms

Machine learning can be described as a set of algorithms and statistical methods that can automatically recognize patterns in data and then use the unknown patterns to accurately predict upcoming data or perform other kinds of decision-making under uncertainty. Usually, machine learning is categorized into two main types: predictive or supervised learning and descriptive or unsupervised learning.

2.2.1. Supervised learning:

In a supervised approach, the goal is to learn mapping from inputs to outputs so that it can predict the output of new, unseen inputs. An algorithm is provided with a labeled set of input-output pairs, known as the training set. An algorithm utilizes a training set to discover and learn a connection between the inputs and the outputs, giving it the ability to predict new data [3].

Supervised learning has two objectives to accomplish:

- Classification: The output variable is a class (e.g., a healthy or unhealthy tree).
- Regression: To predict the future value of the output variable.

2.2.2. Unsupervised learning:

In unsupervised learning, the goal is to find hidden patterns in the data provided. It is less defined than supervised learning since it is not given what kinds of patterns to look for and what the desired output is for each input. The advantage that unsupervised learning provides over supervised learning is that it does not require human intervention to manually label the data, which speeds up the data processing.

Unsupervised learning goals to achieve:

- Clustering: grouping the data that has similar features.
- Dimensionality reduction: reducing the number of input features while keeping as much information as possible.

2.2.3. Principal Component Analysis and Linear Discriminant Analysis

One example of unsupervised learning is principal component analysis (PCA). PCA is commonly used as a method to reduce the dimensionality of large data sets. The tradeoff for the reduction is accuracy, but the most significant data features are still retained.

The concept behind PCA is that PCA attempts to compress as much information as possible in the first principal component (PC1), then the maximum remaining information in the second (PC2), and so on. The number of principal components depends on the number of input features [3]. The construction of the first principal

component accounts for the largest possible variance in the data set. As shown below in Figure 2(a), it is the line that goes through the origin, and it is the line in which the projection of the points (red dots) is the most spread out.A second principal component is calculated in the same way; however, it must be orthogonal to the first component, and it must account for the next highest variance.

On the other hand, Linear Discriminant Analysis (LDA) is a supervised technique for dimensionality reduction that seeks to identify the linear combination of features that best distinguishes the classes in a data set. The intent is to project data into lower-dimensional space while keeping the information that is most relevant for class discrimination.

LDA works by finding a new axis that maximizes the distance between the mean values of various classes while minimizing the spread of points within each class [4]. However, with LDA, these linear discriminants are aiming to separate classes as far apart as possible, which is shown in Figure 2(b).
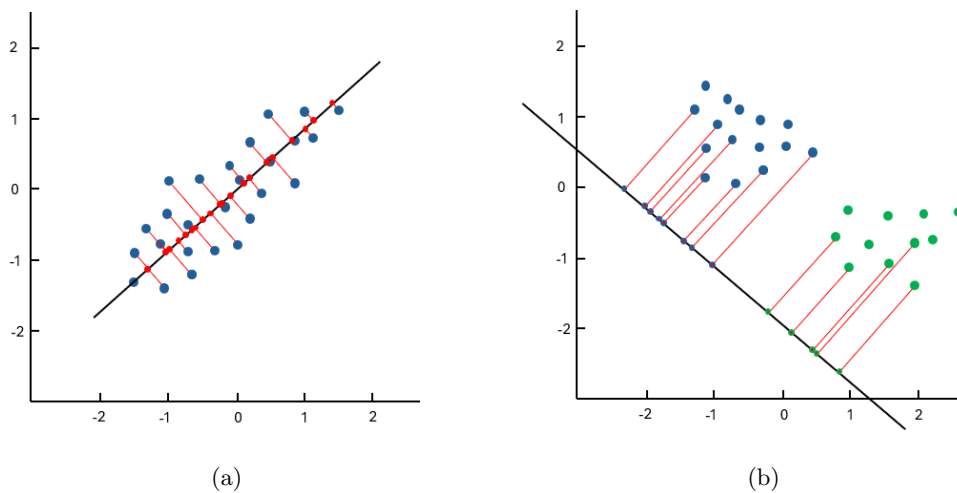

(a)   (b)

Figure 2: (a) Construction of the principal component. (b) Class separability in LDA.

3. METHODOLOGY

3.1. Sensor selection

A VOC detection focused on identifying VOCs like α-pinene, 3-carene and d-limonene whose concentration in trees increases during periods of stress [5][6]. Other VOCs, such as ethanol, 2-methyl and verbenol, were also subjects of study, as their presence could indicate insect infestations in spruce trees, such as by bark beetles. Sensors, which are mainly sensitive to alcohol, were selected to create a sensor array to detect these VOCs. A list represented in Table 1 shows sensors used and the gases they are most sensitive to. Sensors were integrated in sets of four or pairs for greater stability. All the sensors are MOx based sensors due to their simplicity and cost effectiveness, but other types of sensors could be applied, such as electrochemical or PID sensors.

Table 1 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

To try to reduce the number of sensors, Random Forest was applied to create a feature importance chart. The only sensor that stood out was GGS-2330, which has shown low importance in substance and tree measurements, as seen in Figure 3.

This indicates that taking GGS-2330 out would not affect the performance of e-nose in detection of relevant VOCs. Another way to reduce the number of sensors would be to use only one of each sensor instead of sets, but this might affect the stability of a system.

In addition to the sensors above, ENS160 and BME280 sensors for monitoring temperature, humidity, and total VOC levels were implemented.Note that a Table 1 represents the gases the sensors are most sensitive to, but there may be other gases that can affect the sensitivity.
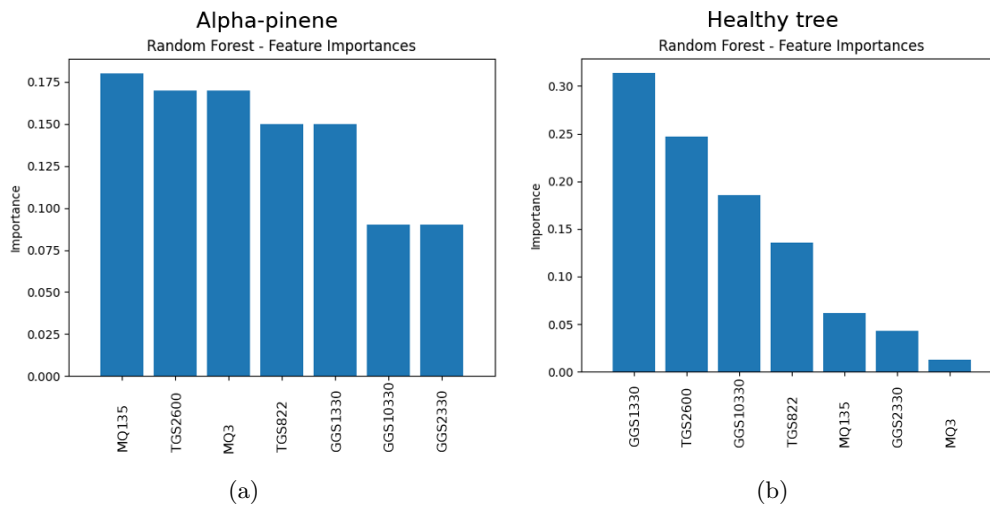
Figure 3: (a) Feature importance (α-pinene) (b) Feature importance (Healthy tree)

3.2. Experimental setup and measurement approach

The experimental setup included two distinct odor measurement systems: one for tree analysis and the other for substance detection. Figure 4(a) 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(b), 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.
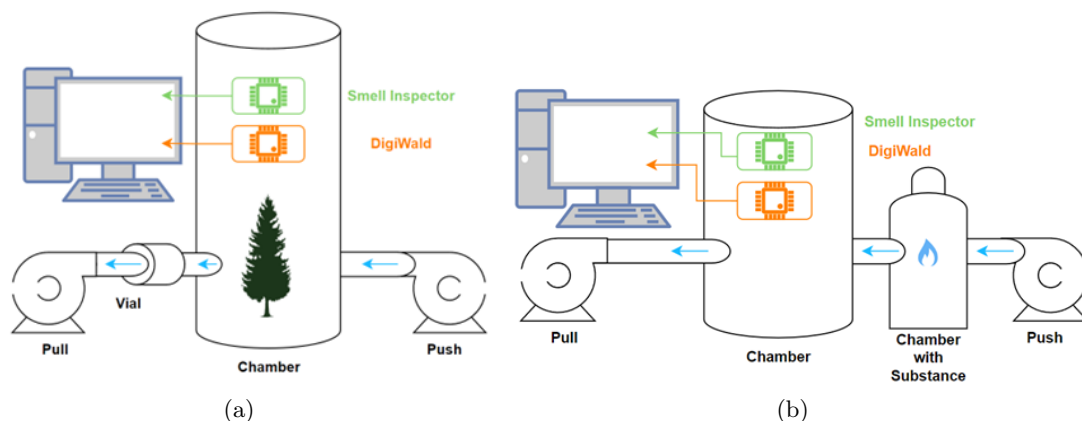
Figure 4: Measurement setup for: a) tree measurement b) substance measurement

3.3. Integrating Circuit Design, Data Acquisition, and Analysis

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 5. Data is collected and stored in a CSV file.

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

Figure 5: Structure of string data sent from Arduino to PC

For analysis, the collection of data included measurements of α-pinene, 3-carene, d-limonene, ethanol, 2-methyl, 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.

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 AND DISCUSSION

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. Digi-nose:

Looking at Figure 6(a), one can notice quite the overlapping of sick and healthy trees. The cause might be that some trees are not as sick as others or that healthy trees are not 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 6(b).
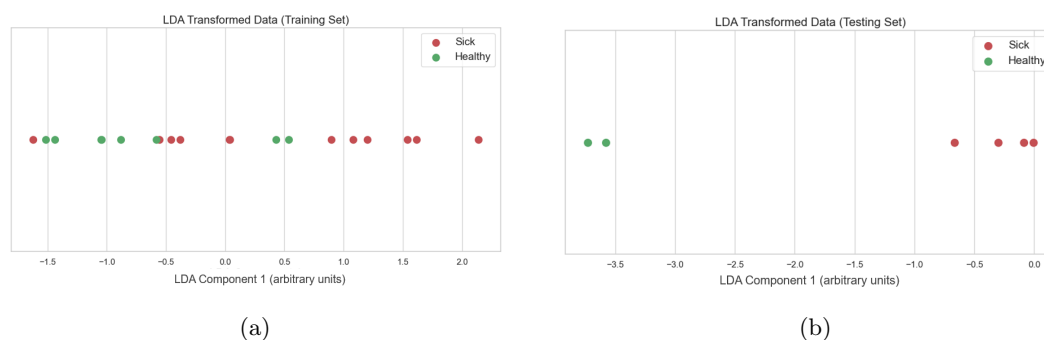

(a)   (b)

Figure 6: LDA analysis of a: a) training set b) testing set.

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, verbanol, and air clearly separated, while 2-methyl, 3-carene, α-pinene and mixture 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.

LDA analysis of substances is presented in Figure 7 and Figure 8. Similar to tree analysis, 80% of the data was used for training and 20% for testing set. Training set achieved 70% accuracy score, while the testing set accuracy decreased to 65%.
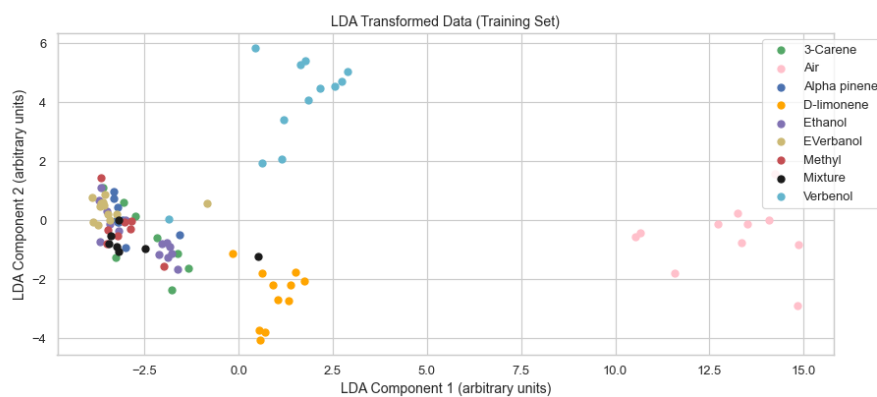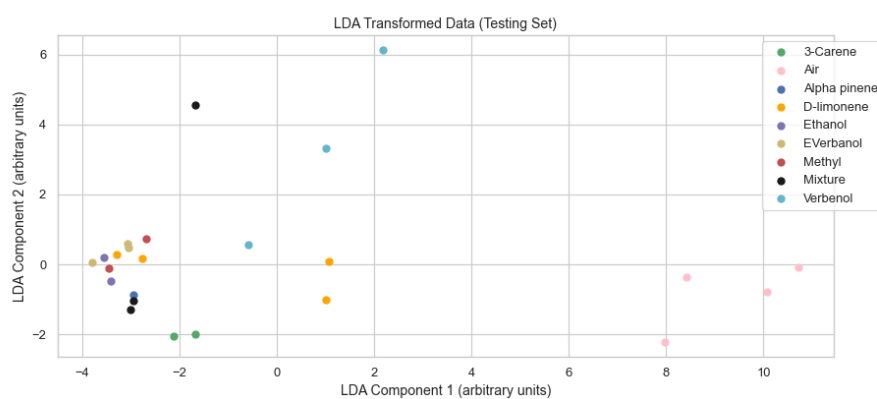
Figure 7: LDA analysis of a training set.

Figure 8: LDA analysis of a testing set.

4.2. Smell inspector:

Data from trees and substances have been gathered in the same manner as with Digi-nose, and LDA analysis has been applied. Analysis of tree data resulted in 88% accuracy in the training set, while testing accuracy was very low, around 33%. This resulted in overfitting, which may be caused by a small set of data. LDA analysis of smell inspector data can be seen in Figure 9(a) and Figure 9(b).
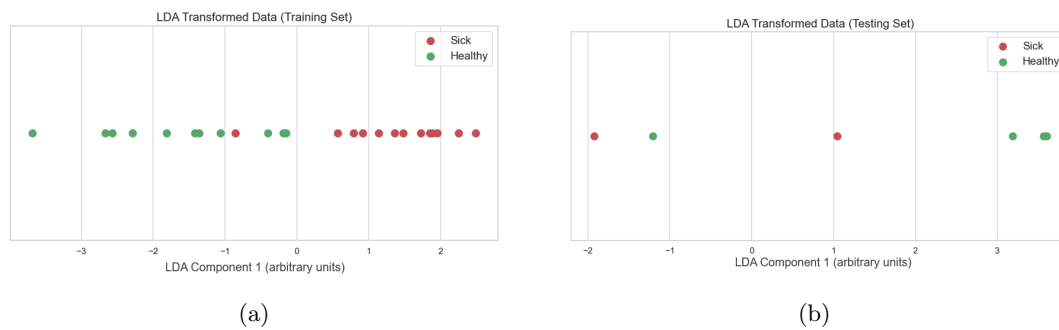
Figure 9: LDA analysis of a: a) training set b) testing set.

Substance analysis gave better results than analysis of tree data. High accuracy of 96% for the training set has shown great separability of substances, with slight overlap of substances such as mixture, α-pinene, 2-methyl. Testing accuracy dropped slightly but still remained at high 88%. LDA results are shown in Figure 10 and Figure 11.
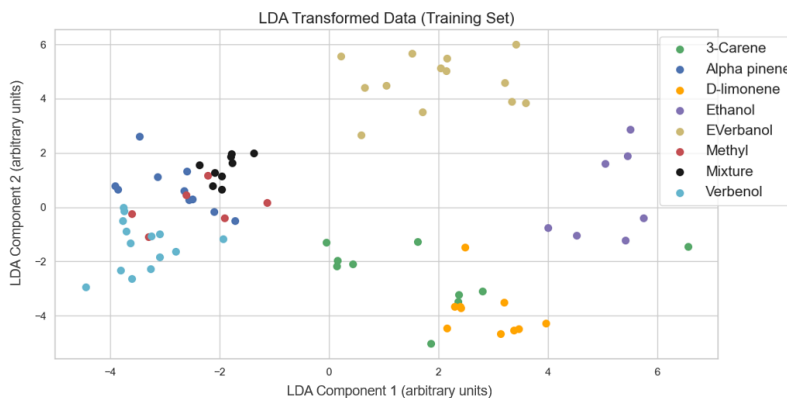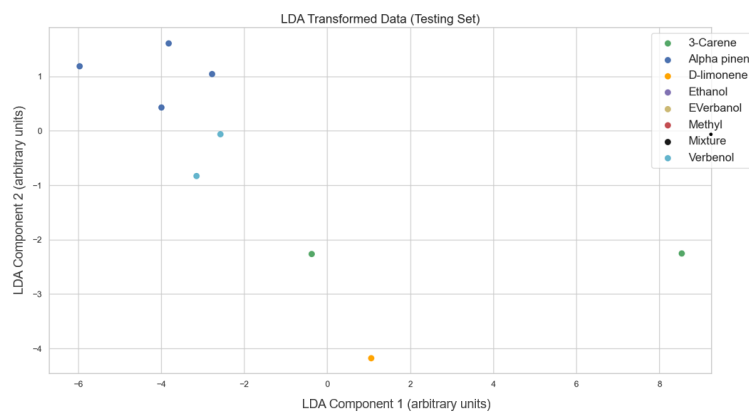

Figure 10: LDA analysis of a training set.

Figure 11: LDA analysis of a testing set.

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 reasonable level of accuracy. While the PCA provided initial separation of data, the LDA provided clearer class separation, with overlapping in some cases.

Despite the success, an obstacle such as the MOx sensors' limited selectivity for different gases has shown a level of difficulty. 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 exploit the potential of this newly developed technology.

6. OUTLOOK

Future work shall focus on the development of a temperature compensation algorithm that would utilize temperature data to adjust MOx sensors output data. This would improve e-noses ability to operate in different environmental conditions.

While LDA has shown promising results, application of other algorithms, such as Random Forest, would improve the problem of overfitting since it is less prone to it. Another approach would be to increase a data set by more measurements or by data augmentation (such as applying noise to the existing data).

Since the data is balanced, accuracy has provided a reliable measure of performance. Nevertheless, a system would benefit from additional metrics, such as the F1-score or confusion matrix, which would offer a more thorough evaluation.

References

[1] J. W. Gardner, K. C. Persaud, Electronic Noses and Olfaction 2000: Proceedings of the 7th International Symposium on Olfaction and Electronic Noses, Brighton, UK, July 2000, CRC Press, 2001.

[2] R. Jaaniso, O. K. Tan, Semiconductor gas sensors, Elsevier, 2013.

[3] C. M. Bishop, N. M. Nasrabadi, Pattern recognition and machine learning, volume 4, Springer, 2006.

[4] Y. Wang, W. Zhang, R. Gao, Z. Jin, X. Wang, Recent advances in the application of deep learning methods to forestry, Wood science and technology 55 (2021) 1171–1202.

[5] M. Antonelli, D. Donelli, G. Barbieri, M. Valussi, V. Maggini, F. Firenzuoli, Forest volatile organic compounds and their effects on human health: A state-of-the-art review, International Journal of Environmental Research and Public Health 17 (2020) 6506.

[6] J. K. Holopainen, J. Gershenzon, Multiple stress factors and the emission of plant vocs, Trends in plant science 15 (2010) 176–184.