science, engineering and health studies

Research Article

Science, Engineering and Health Studies
https://lio1.tci-thaijo.org/index.php/sehs
ISSN (Online): 2630-0087

Detection of mealybug infestation on the Khasi Mandarin orange plant using electronic nose

Sudipta Hazarika[1], Rajdeep Coudhury[2], Lakhi Nath Borah[1], Arindam Phukan[1], and Utpal Sarma[2]*

[1] Department of Physics, Madhabdev University, Assam, India
[2] Department of Instrumentation & USIC, Gauhati University, Guwahati, India

***Corresponding author**:
Sudipta Hazarika
Sudipt.hazarika38@gmail.com

***Received**: 15 February 2023*
***Revised**: 3 October 2023*
***Accepted**: 5 October 2023*
***Published**: 28 December 2023*

***Citation**:
Hazarika, S., Coudhury, R., Borah, L. N., Phukan, A., and Sarma, U. (2023). Detection of mealybug infestation on the Khasi Mandarin orange plant using electronic nose. Science, Engineering and Health Studies, 17, 23040009.

ABSTRACT

Mealybugs pose a serious threat to fruit crops leading to premature leaf and fruit drops which severely affects the yield as well as the quality. Primarily pest detection is done with the help of human/animal scouting, which is cumbersome and prone to error. This paper studies the feasibility of using an electronic nose (E-Nose) for detecting mealybug infestation in Khasi Mandarin orange plants. Plants normally release volatile organic compounds (VOCs) which can act as biomarkers for specific stresses affecting the plant. These VOCs can be analyzed to diagnose the plant. VOCs emanating from leaf samples of both infested and healthy plants were analyzed using a custom-made E-Nose system containing an array of commercially available gas sensors. Dimensionality reduction techniques using principle component analysis, and linear discriminant analysis and optimized classification algorithms like support vector machine and random forest were employed to check for the discriminating capability of the E-Nose system. The technique successfully classified samples belonging to infested and healthy categories in both the classifiers with accuracies of 95.66% and 96.70%.

Keywords: pest diagnosis; crops; electronic nose; machine learning; support vector classifier

1. INTRODUCTION

Major production and economic gains in the vegetable and fruit industry are highly dependent on early-stage disease and pest diagnosis. Plants being rooted to a spot are vulnerable to a variety of biotic stresses as well as abiotic stresses. Biotic stresses on plants originate due to their interaction with living organisms, whereas abiotic stresses originate due to ambient environmental factors. These stresses adversely affect the growth and yield of the plant and a lot of research is concentrated on minimizing their adverse effects and providing sustainable management strategies. For instance, the effect of shade tolerance on the growth characteristics of moth beans and two legume species were evaluated by Na Chiangmai et al. (2013). Similarly, the yield, growth and quality of pineapples exposed to various levels of chicken manure application was evaluated by Isuwan (2014). The effect of salt stress on the growth of Dendobrium plantlets was analyzed in yet another report, which revealed high salt concentration in soil water reduced the survival rate of the plantlets (Obsuwan et al., 2021). Furthermore, proper management strategies for heavy metal and pathogenic contamination in waste water, used to irrigate vegetable crops, has been reported recently (Priyashantha and Sugirtharan, 2021). Proper and timely diagnosis provides the health status as well as the causes of disorders affecting the plant. According to Food and Agriculture Organization of the United Nations, approximately 40% of the world's annual crop production is lost to pests. Annually, plant diseases result in economic losses exceeding $220 billion globally, while invasive insects contribute at least $70 billion to these losses. Historically, the sole practical choice available to

farmers for pest management has been the utilization of pesticides. Due to their harmful environmental impact, research on feasible alternatives are being actively pursued by the scientific community in recent years. An early warning system to monitor plants for disease and pest buildup could significantly decrease farmers' dependency on chemical sprays. This work reports the use of a custom-developed electronic nose (E-Nose) system to detect the presence of mealybugs in Khasi Mandarin orange plants.

Mealybugs (*Planococcus citri* Risso), are considered to be a serious threat to many fruit cultivars and occasionally acquire epidemic status in citrus. It causes severe economic damage to citrus cultivars in India and is considered to be an economically important pest on Khasi Mandarin orange which is primarily grown in Meghalaya and Assam. Apart from Khasi Mandarin orange, these bugs also feed on plants often not related to citruses, like curry leaf plants, mango, banana, ginger, cocoa, cotton, and some flowering plants (Rao et al., 2006). They live on the aerial parts of the host plant and feed on new leaves, flowers, stem-ends of fruits, and even extract sap resulting in loss of vigor of the plant. The leaves appear to be chlorotic, wilted, and distorted, accompanied by the premature falling of fruits and leaves. Additionally, these bugs produce copious amounts of honeydew covering the feeding sites, on which a black mould grows, further degrading the health of the plant by interfering with photosynthesis. Black ants attracted to the honeydew often aid in their transportation and spread.

Extensive infestation often results in stunted growth of the plant and parts of the plant die off gradually. The pest rapidly grows in huge quantities. For instance, each female lays between 300–800 eggs into an egg sac, which are then deposited on the plants protected with fluffy white wax filaments. These eggs hatch in 1–3 weeks and the crawlers emerging from them quickly disperse over the plant to find a suitable feeding site. The most commonly employed management techniques include pruning of affected shoots, destruction of ant colonies, raking soil around the plant to expose the eggs in the soil to natural enemies, and using sticky bands around the trunk. Thus, the implementation of a timely and proper management strategy to negate the losses incurred due to pest infestation at the earliest is vital for many orchards. This would also ensure quality yield, as well as immensely help in stopping further infestations to neighboring plants.

Plant pest detection in large-scale orchards is carried out using traditional techniques, like human or animal scouting, which becomes very cumbersome, time-consuming, costly, and prone to error. It typically involves the process of checking the threshold count of pests per leaf samples collected or those captured in sticky tapes installed in the orchards. To reduce the loss incurred due to pests, it has become crucial to obtain a real-time accurate forecast of crop pests that can provide early warnings. Many techniques including the use of acoustic sensors, ultrasonic sensing, laser technology, and optical sensing have been used for pest detection (Azfar et al., 2018). A multispectral imaging technique was recently used to detect insect vectors of brown streak disease in cassava plants (Fennell et al., 2018). Computer vision and machine learning approaches have been used several times to detect pests in plants (Gondal and Khan, 2015; Gutierrez et al., 2019).

Due to the lack of an adaptive immune system (in vertebrates), plants have a plethora of complex defense mechanisms. At the front line, plants protect themselves from pests, microbes, and herbivores using physical barriers such as hairs, trichomes, thorns, and spines. Another line of defense includes the production and emission of complex specialized secondary metabolites that have toxic, repellent, and anti-nutritional effects on the stress-causing agent. These are blends of organic mixtures of carbon and hydrogen called volatile organic compounds (VOCs), which are volatile at ambient temperature and pressure. They also act as communication mechanisms used to warn and defend themselves, as well as neighboring plants (Nishad et al., 2020; War et al., 2012).

VOCs can serve as cues for assessing the health status, growth pattern, and diseases of a plant. These VOC-mediated defense mechanisms function either by emitting insect repellent VOCs or by emitting VOCs that attract pest predators. In brief, these complex chemical signals/ metabolites provide the real-time physiological health status of a plant and can be investigated for rapid diagnosis (Cui et al., 2018b). Metabolomics study to identify the specific chemical signals emitted in response to stress and the metabolic pathway associated with their emission is complex and can help in diagnosis, even before visible symptoms occur. Two types of techniques used for this purpose are chromatographic techniques, associated with mass spectrometry, and the electronic nose (E-Nose) technique. The former is used for separating and identifying the specific individual chemicals related to the stress/ attack on the plant (Barah and Bones, 2015; Maag et al., 2015). However, as useful as it may be, it is expensive, time-consuming, and requires specific expertise for identifying the chemicals emitted. In comparison, the E-Nose technique is less expensive, rapid, and compact, hence it can be used extensively on large-scale orchards. It does not facilitate the analytical determination of individual VOCs or provide a quantification of the gaseous mixture emitted, rather it recognizes gas mixtures as a whole emitted from the plant.

There are some instances confirming the promising capability of E-Nose systems to provide fast and reliable detection of pest-infested plants at an early stage. Cui et al. (2018a) reported an 86% accurate classification of aphid-stressed, whitefly-stressed tomato plants from healthy groups. In another instance, a portable device developed to draw volatiles from pest damaged products over carbon black-polymer composite sensors was reported (Lampson et al., 2014a, 2014b). With the implementation of proper pattern recognition models, an E-Nose system can be trained to classify plants under stress from the healthy ones (Ghaffari et al., 2011; Zhang et al., 2011).

In this study, the E-Nose technique has been implemented for detecting mealybug infestations in Khasi Mandarin orange, which has not been reported to date. Machine learning algorithms have been used for the classification of infected and healthy plants kept in a controlled greenhouse. The study also identifies the subset of sensors responding mostly to the classification process.

2. MATERIALS AND METHODS

A total of 18 Khasi Mandarin orange plants were grown in a greenhouse, maintaining a constant temperature of 20 °C. Mealy bugs were manually infested in 8 of these plants and were kept isolated from the rest. The plants were kept under observation for two weeks prior to sample collection for E-Nose analysis. Figure 1 shows a snapshot of an infested plant.

Figure 1. Snapshot of a mealybug infested Khasi Mandarin orange plant

Next, 15 fully expanded leaves from each tree were collected from all the plants and tested using the custom developed E-Nose system. The leaves collected were picked randomly from around the canopy of each plant in such a way that all leaves were mature and no leaf was mechanically damaged. The E-Nose system consists of an array of commercially available metal oxide gas sensors, a sample holding chamber, a data acquisition unit, and a diaphragm micro pump for inducing the sampled gas and purging the chamber with ambient air, as shown in the block diagram in Figure 2.

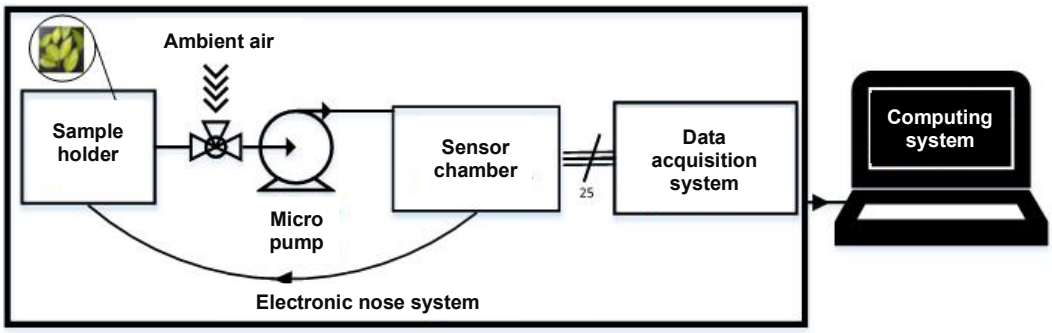

Figure 2. Block diagram of the custom designed prototype

The detailed design simulation and working of the E-Nose system can be found in our previous manuscript (Hazarika et al., 2020), where it was used to diagnose Khasi Mandarin orange plants infected with a virus called citrus tristeza virus. The sensors embedded in the sensor board and the associated circuitry was connected to the microcontroller board via a parallel bus. The microcontroller board contained an 8-bit Atmega 2560 microcontroller having an inbuilt 10-bit analog to digital converter (ADC). The microcontroller was programmed to record the responses of the sensors every second, and the data was sent to the computing system via a USB interface. Firstly, the leaf samples were cut uniformly into small pieces of approximately 1 cm^2 and fed to the sample holding chamber. The head space generated within the chamber was circulated throughout the E-Nose system using the pump. As shown in Table 1, nine sensors were used to sense the headspace circulating between the sensor chamber and the sample holder, and their corresponding digital responses were recorded and stored in the form of a (comma-separated value) csv file for analysis.

Data acquisition of each sample using the prototype consisted of two phases; i) the exposure phase and ii) the recovery phase. The sensors were exposed to the headspace sucked in from the sample holder for 4 min in the exposure phase. Afterwards, the sensor chamber and the sample holder were both evacuated and purged during the recovery phase, as shown in the flow chart in Figure 3. Between subsequent data acquisition from the two samples, both the sensor chamber and sample holding chambers were firstly purged by pumping in ambient air for 5 min and then extracting the same for another 5 min. This was done to clean the chambers of the previously acquired head space.

Table 1. Sensors used and their target gases

Sensor name	Target gases
MQ3	Alcohol, smoke
MQ6	LPG, butane
MQ4	Methane, CNG
TGS2610	Ethanol, methane, ISO-butane/propane
TGS2620	Alcohol, solvent vapors
TGS2602	Ammonia, ethanol, toluene, hydrogen sulfide
TGS822	Methane, carbon monoxide, ISO-butane, n-hexane, benzene, ethanol
TGS5141	Carbon monoxide
TGS2444	Ammonia

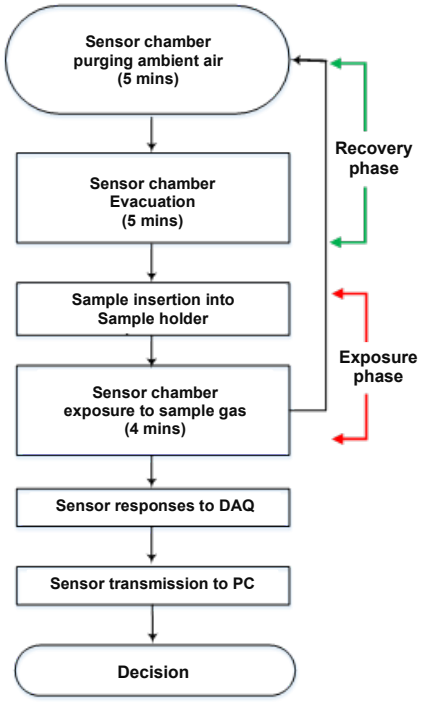

Figure 3. Flow chart of the E-Nose measurement

The data acquired from each sample analyzed using the E-Nose system was stored for pattern recognition analysis. The microcontroller in the data acquisition system was programmed to acquire data of each sample for 4 min with a sampling rate of 9 responses per second for the 9 sensors (Table 1). Thus 240 pieces of data were collected by each sensor during the exposure phase, resulting in a data set of 240×9 for each sample. These data sets generated for all 18 samples were stored and analyzed using machine learning techniques for the classification of infested and healthy samples. Feature extraction techniques like data normalization, dimensionality reduction and classification techniques were used for data analysis. Principle component analysis (PCA), and linear discriminant analysis (LDA) were used as a dimensionality reduction technique, while bootstrap aggregation (bagging) ensemble of the support vector classifier (SVC) and random forest (RF) classifier were used for classification. The entire dataset was split 80:20 for training and testing respectively. A 10-fold cross validation was used for tuning the parameters of the classifier models. Each iteration had a model trained by 9-fold of the 80% training data and tested on the untouched 20% testing data. This was done so that the model was tested with data that was not used for training. The mean of the scores obtained from each iteration was considered for evaluating the performance of the classifier models.

3. RESULTS AND DISCUSSION

The responses of the 9 sensors for a healthy and an infested sample are shown in Figure 4. The datasets containing data of the two classes of samples, infested and healthy, were analyzed using PCA, and the reduced dimensions (principal components) were plotted as shown in Figure 5.

PCA is an unsupervised technique, which reduces the dataset into components in order of the variance explained. The plot exhibited discrimination between the infested and the healthy samples, although a slight overlap appeared. The 1st principal component (PC1) explained 98% of the total variance, while the remaining components explained the rest.

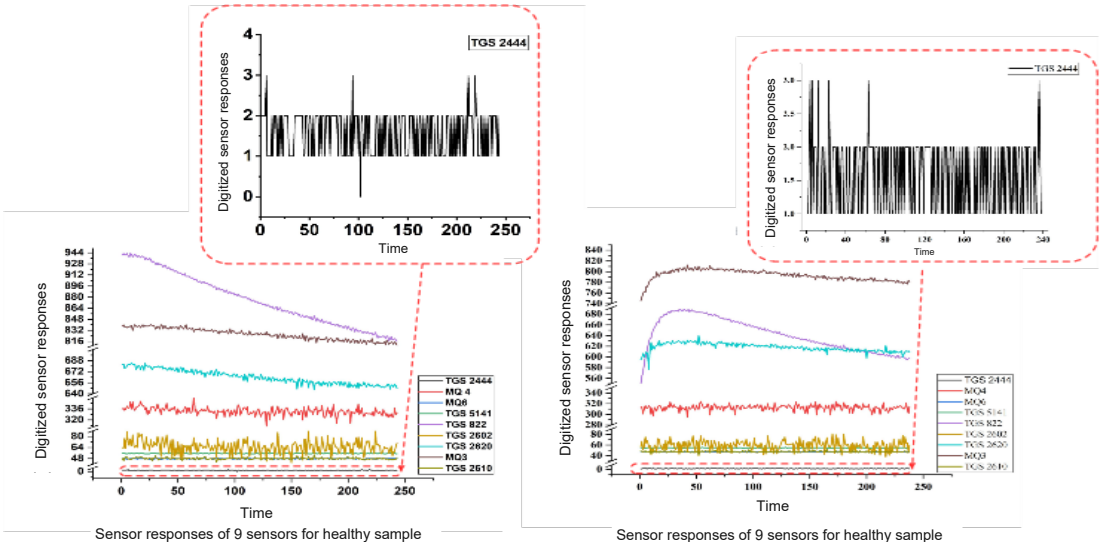

Figure 4. Responses of the 9 metal oxide gas sensors embedded in the prototype for a healthy and mealybug-infested samples

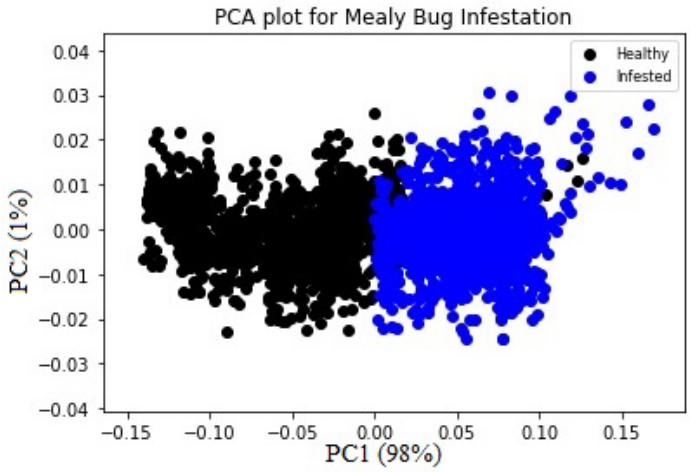

Figure 5. PCA plot showing discrimination of healthy and infested Khasi Mandarin orange plant

The classifier models used were a bagging ensemble of SVC and RF classifier. The performance of the classifiers are shown in Table 2. All the scores obtained for the classifier models were greater than 90%, which is considered to be satisfactory. The Cohen's kappa coefficient, which is a measure of the inter-rater agreement among the raters in categorical analysis is considered to have near perfect agreement for values in the range 0.81 to 1. The raters in this case being the model's predictions and the actual observations.

Table 2. Pearson correlation between physicochemical parameters with total nitrogen and nitrogen forms

Performing scores	Scores in %	
	SVC	RF
Accuracy	95.66	96.70
Precision	96.00	97.00
Recall	96.00	97.00
F1 score	96.00	97.00
Cohen's kappa score	91.35	93.40

To visualize the actual decision boundary of the model (classifier performance), the dimensionally reduced dataset using LDA, was fed to the classifier algorithm and the linear discriminants LD1, and LD2 were plotted. The LDA chooses the dimensions by minimizing the scatter of the data and maximizing the separation of the data. The classifier model demarcated the decision boundary of the training and the test data set, as shown in Figure 6. The

decision boundary in the test set follows the training model and can be seen to separates the infested and the healthy samples.

Figure 7 represents the receiver operating characteristics (ROC) curve of the SVC classifier model, which is plotted between the sensitivity and the specificity of the classifier for different threshold values. The sensitivity, denoting the percentage of correct classification of the infested samples, is also known as the true positive rate. The specificity (true negative rate) denotes the opposite. The area under the curve (AUC) score observed for each class of data was 96%. This score denotes the performance of the model in separating the two classes. In addition to resampling the dataset using bootstrap method, for training individual classifiers, the RF classifier also evaluated subsets of the features to identify the top-performing set for accurate classification. The results are presented as feature importance scores in Figure 8, highlighting the sensors contributing the most to the model's predictive capability. Understanding these scores facilitated the identification of the four most responsive sensors among the nine in the classification.

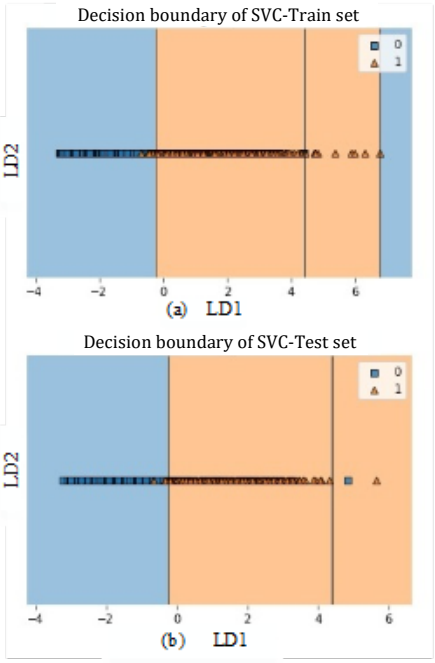

Figure 6. Decision boundary of the (A) train data set and (B) test dataset for the SVC model

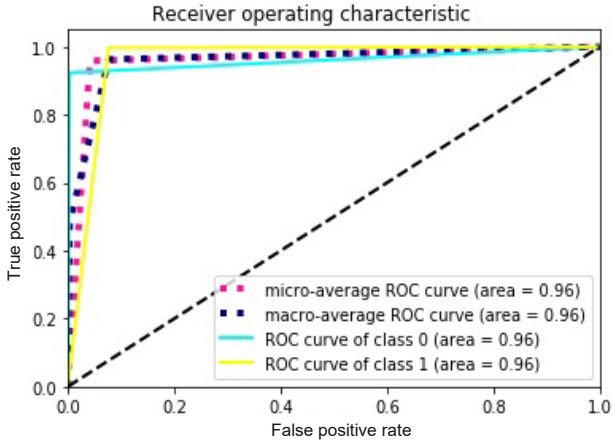

Figure 7. Receiver operating characteristic (ROC)

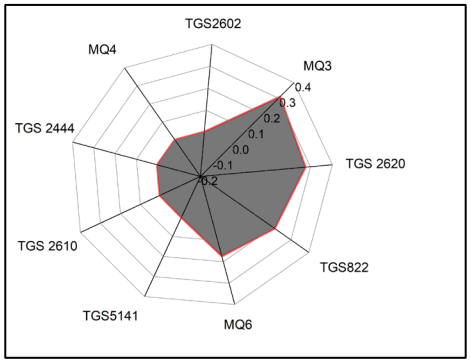

Figure 8. Radar plot of the feature importance scores of the 9 sensors used in the prototype

Leveraging this knowledge streamlines sensor selection, potentially reducing system size, costs, and complexity while maintaining research accuracy, as shown in Table 3 and Figure 9. The extracted features were individually fitted with RF classifier models to assess their performance independently. The resulting metric scores are presented in Table 3, highlighting a noticeable disparity in performance between the two subsets. Figure 9 visually reinforces these findings, illustrating that the best features led to distinct and accurate classification, whereas the subset containing the remaining features displayed lower classification accuracy.

Table 3. Performance scores of the random forest classifier model with the extracted features and the remaining features

Performance scores	**Scores in %**	
	Subset with best features	**Subset with rest of the features**
Accuracy	99.13	71.40
Precision	98.60	69.15
Recall	99.64	75.26
F1 score	99.12	72.08
Cohen's kappa score	98.26	42.88

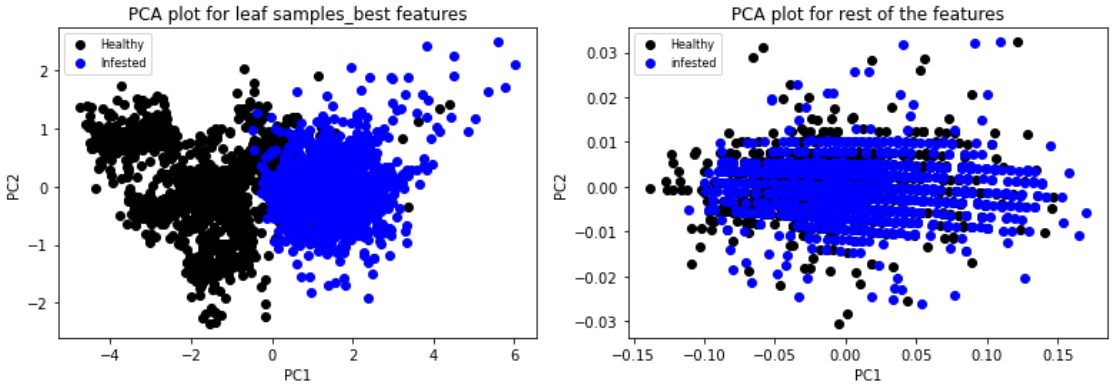

Figure 9. PCA plot showing discrimination of healthy and infested Khasi Mandarin orange plant using the subset of features containing the best features and the rest of the features

At this stage of the research, it is deemed safe to say that the E-Nose technique can be successfully implemented in predicting the yes or no status of mealybug infestation in Khasi Mandarin orange plants. However, for determining its sensitivity to varied degrees of infestation, additional data needs to be acquired and analyzed.

4. CONCLUSION

In this paper, the use of E-Nose has been reported in detecting mealybug infestations in Khasi Mandarin orange. Past research on the subject mainly focused on the traditional methods of detection and management. Here, an advanced approach for detection was investigated, which if implemented on large scale orchards will save labour, time and cost. A total of 8 Khasi Mandarin plants were infested with mealybugs and isolated from 10 healthy Khasi Mandarin plants in a greenhouse. Leaf samples from each of these plants were collected and analyzed using a custom designed E-Nose system. The data sets generated on these samples were saved and analyzed with pattern recognition and classification approaches. The classification model distinguished the infested and healthy samples with an accuracy of 95.66%. The Cohen's kappa score showed a near perfect agreement signifying the effectiveness of the

technique. Out of the nine sensors employed for the E-Nose analysis, the set of sensors mostly responding to the VOC profile alteration due to the infestation were MQ3, TGS 2620, TGS 822, and MQ6. This would help in making the prototype compact and reduce the cost. This work reports the yes–no status of mealybug infestations in a laboratory environment and additional study needs to be carried out for cultivar specific research and to analyze the sensitivity of the technique to early infestation.

REFERENCES

Azfar, S., Nadeem, A., Alkhodre, A. B., Ahsan, K., Mehmood, N., Alghmdi, T., and Alsaawy, Y. (2018). Monitoring, detection and control techniques of agriculture pests and diseases using wireless sensor network: A review. *International Journal of Advanced Computer Science and Applications*, 9(12), 424–433.

Barah, P., and Bones, A. M. (2015). Multidimensional approaches for studying plant defence against insects: From ecology to omics and synthetic biology. *Journal of Experimental Botany*, 66(2), 479–493.

Na Chiangmai, P., Pootaeng-on, Y., and Khewaram, T. (2013). Evaluation of the shade tolerance of moth bean *(Vigna aconitifolia)* and two tropical legume species. *Silpakorn University Science and Technology Journal,* 7(1), 19–31.

Cui, S., Inocente, E. A. A., Zhu, H., Acosta, N., Ling, P. P., and Keener, H. M. (2018a). Development of portable E-nose system for early diagnosis of insect-stressed tomato plants. *2018 ASABE Annual International Meeting: Paper no. 1800990*. Detroit, Michigan: American Society of Agricultural and Biological Engineers. July 29 – August 1.

Cui, S., Ling, P., Zhu, H., and Keener, H. M. (2018b). Plant pest detection using an artificial nose system: A review. *Sensors*, 18(2), 378.

Fennell, J., Veys, C., Dingle, J., Nwezeobi, J., van Brunschot, S., Colvin, J., and Grieve, B. (2018). A method for real-time classification of insect vectors of mosaic and brown streak disease in cassava plants for future implementation within a low-cost, handheld, in-field multispectral imaging sensor. *Plant Methods*, 14, 82.

Ghaffari, R., Hines, E. L., Zhang, F., Leeson, M. S., Iliescu, D. D., and Napier, R. (2011). Detection of diseases and volatile discrimination of plants: An electronic nose and self-organizing maps approach. In *Intelligent Systems for Machine Olfaction: Tools and Methodologies* (Hines, E. L. and Leeson M. S., Eds.), pp. 214–230. Hershey, PA: IGI Global.

Gondal, D., and Khan Y. N. (2015). Early pest detection from crop using image processing and computational intelligence. *FAST-NU Research Journal*, 1(1), 59–68.

Gutierrez, A., Ansuategi, A., Susperregi, L. Tubío, C., Rankić, I., and Lenža, L. (2019). A benchmarking of learning strategies for pest detection and identification on tomato plants for autonomous scouting robots using internal databases. *Journal of Sensors*, 2019, 5219471.

Hazarika, S., Choudhury, R., Montazer, B., Medhi, S., Goswami, M. P., and Sarma, U. (2020). Detection of citrus tristeza virus in mandarin orange using a custom-developed electronic nose system. *IEEE Transactions on Instrumentation and Measurement*, 69(11), 9010–9018.

Isuwan, A. (2014). Agronomic traits and fruit quality of pineapple with different levels of chicken manure application. *Silpakorn University Science and Technology Journal,* 8(1), 67–73.

Lampson, B. D., Han, Y. J., Khalilian, A., Greene, J. K., Degenhardt, D. C., and Hallstrom, J. O. (2014a). Development of a portable electronic nose for detection of pests and plant damage. *Computers and Electronics in Agriculture*, 108, 87–94.

Lampson, B. D., Khalilian, A., Greene, J. K., Han, Y. J., and Degenhardt, D. C. (2014b). Development of a portable electronic nose for detection of cotton damaged by *Nezara viridula* (hemiptera: pentatomidae). *Journal of Insects*, 2014, 297219.

Maag, D., Erb, M., and Glauser, G. (2015). Metabolomics in plant–herbivore interactions: Challenges and applications. *Entomologia Experimentalis et Applicata*, 157(1), 18–29.

Nishad, R., Ahmed, T., Rahman, V. J., and Kareem, A. (2020). Modulation of plant defense system in response to microbial interactions. *Frontiers in Microbiology*, 11, 1298.

Obsuwan, K., Deesubin, P., Tongam, A., and Juneenat, O. (2021). Influence of paclobutrazol on growth of *Dendrobium* 'Sonia Jo Daeng' under salt stress condition in tissue culture. *Science, Engineering and Health Studies*, 15, 21030001.

Priyashantha, A. K. H., and Sugirtharan, M. (2021). Heavy metals and pathogenic contamination in vegetable crops through wastewater irrigation and risk reduction in developing countries. *Science, Engineering and Health Studies*, 15, 21010008.

Rao, C. N., Shivankar, V. J., and Singh, S. (2006). Citrus mealy bug (PLANOCOCCUS CITRI RISSO) management - A review. *Agricultural Reviews*, 27(2), 142–146.

War, A. R., Paulraj, M. G., Ahmad, T., Buhroo, A. A., Hussain, B., Ignacimuthu, S., and Sharma, H. C. (2012). Mechanisms of plant defense against insect herbivores. *Plant Signaling & Behavior*, 7(10), 1306–1320.

Zhang, F., Iliescu, D. D., Hines, E. L. and Leeson, M. S. (2011). Tomato plant health monitoring: An electronic nose approach. In *Intelligent Systems for Machine Olfaction: Tools and Methodologies* (Hines, E. L. and Leeson M. S., Eds.), pp. 231–247. Hershey, PA: IGI Global.