http://www.pjbs.org

PJBS

ISSN 1028-8880

Pakistan Journal of Biological Sciences

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

OPEN ACCESS Pakistan Journal of Biological Sciences

ISSN 1028-8880
DOI: 10.3923/pjbs.2026.179.192

Research Article
Finding Hidden Huanglongbing using an Electronic Nose

[1]Anthoni Sulthan Harahap, [1]Siti Subandiyah, [2]Iman Rahman, [3]Nerida Donovan and [2]Kuwat Triyana

[1]Department of Plant Protection, Faculty of Agriculture, Universitas Gadjah Mada, Bulak Sumur, Yogyakarta 55281, Indonesia
[2]Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Sekip Utara BLS 21, Yogyakarta 55281, Indonesia
[3]Department of Primary Industries and Regional Development, New South Wales (NSW), Menangle, New South Wales 2568, Australia

Abstract

Background and Objective: The Huanglongbing (HLB) is one of the most destructive diseases affecting citrus worldwide. A major challenge in its management is its ability to remain asymptomatic for extended periods, delaying timely detection and control. This study aims to develop and evaluate a compact electronic nose (e-nose) system equipped with metal-oxide semiconductor (MOS) sensors for the early detection of Candidatus Liberibacter asiaticus infection in citrus leaves through Volatile Organic Compound (VOC) analysis under field-like conditions. **Materials and Methods:** A total of 454 Purworejo Siamese citrus leaf samples were collected from two orchards. The infection status of each sample was confirmed using conventional Polymerase Chain Reaction (PCR) prior to headspace VOC extraction. The cross-sensitive MOS sensor array converted VOC interactions into electrical signals, which were subsequently preprocessed, feature-extracted and analyzed using machine learning pipelines. Model selection and optimization were performed on baseline-shifted data. **Results:** A stratified 5-fold cross-validation using the Extra Trees algorithm successfully discriminated between PCR-confirmed Candidatus Liberibacter asiaticus-infected leaves and healthy controls, achieving an accuracy of 84.57% (95% confidence interval: 80.98%-88.15%). These results were obtained under field-like conditions and were further validated using headspace gas chromatography-mass spectrometry (HS-GC/MS), which revealed distinct VOC profiles for each group. **Conclusion:** This study demonstrates the potential of the electronic nose (e-nose) as a rapid, in-field screening tool capable of prioritizing samples for laboratory confirmation, thereby supporting effective HLB management.

Key words: Huanglongbing, plant disease detection, electronic nose, asymptomatic, hyperparameter tuning

Citation: Harahap, A.S., S. Subandiyah, I. Rahman, N. Donovan and K. Triyana, 2026. Finding hidden huanglongbing using an electronic nose. Pak. J. Biol. Sci., 29: 179-192.

Corresponding Author: Siti Subandiyah, Department of Plant Protection, Faculty of Agriculture, Universitas Gadjah Mada, Bulak Sumur, Yogyakarta 55281, Indonesia Tel.: +62 812-2657-7651

Funding: This research was funded by Agricultural Human Resources Development and Extension, the Ministry of Agriculture, Republic of Indonesia and the Australian Centre for International Agricultural Research, Australia (ACIAR HORT/2019/164).

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

The HLB is one of the worst citrus diseases in Asia, Africa and the Americas[1]. Candidatus *Liberibacter asiaticus* (CLas)[2], associated with HLB spreads via plant material and psyllid vectors like the Asian citrus psyllid (*Diaphorina citri* Kuwayama, Hemiptera: Psyllidae). Asymptomatic identification is crucial for stopping the spread and decreasing economic losses[3] because symptoms expression is typically delayed by months or years. Historically, targeted symptomatic leaves, where high CLas titres enabled detection using molecular assays[4].

Detection is reported to be more difficult in the early stages of infection when CLas is present below detectable levels using published molecular assays[4]. Despite its role in spreading HLB to new regions and hindering eradication and monitoring, the economic effects of asymptomatic infection have rarely been studied[4]. Visual, physiological, biochemical and sensor-based methods can detect CLas. Visual symptoms are affected by nutritional stress and other factors and cannot detect asymptomatic infection[5]. The ELISA and quantitative PCR (qPCR) are sensitive but need laboratory equipment, expensive chemicals and expert personnel[6]. The iodine-starch test lacks CLas specificity[7].

Sensor-based detection can quickly detect diseased samples *in situ*. With machine learning, this method is rapid and non-destructive. Optical and electrochemical sensors have been studied[8]. A handheld fluorescence device has been shown to successfully identify Citrus greening (CLas) infection at both the symptomatic and asymptomatic stages with a high degree of accuracy, achieving a 97% success rate[9]. A commercial portable e-nose employing a 10-methyl oxide semiconductor (MOS) sensor has been reported to identify multiple symptom HLB. The analysis of the combination of visible-near infrared and thermal imaging captured using two multiple camera arrays was able to distinguish symptomatic CLas-infected plants from healthy plants with an accuracy of approximately 87%[10]. The use of highly sensitive monochrome camera vision sensor successfully distinguished between symptomatic CLas-infected samples of the blotchy mottles type due to starch accumulation and samples deficient in zinc or magnesium with an accuracy of 98.5% at field level[11]. Despite the accuracy achieved by optical sensors, but they are bulky and infrequently confirmed in orchards[6]. The use of commercial MOS electronic-nose (e-nose) arrays capable of detecting symptomatic CLas-infected leaf samples with 97.8% accuracy on symptomatic leaves, but their asymptomatic performance is uncertain. Early detection is biologically viable because VOC patterns vary quickly after infection and before symptoms appear[5].

The present study saw the testing of a portable laboratory-made MOS e-nose that has been validated for food authenticity with high accuracy[12]. The e-nose was also assessed for its capacity to recognize VOC patterns from asymptomatic CLas infections in orchard-like circumstances which has not been achieved by previous studies. Asymptomatic purworejo siamese citrus VOC signatures can be distinguished from healthy samples, offering a rapid field screen to prioritize laboratory confirmation. This study develop and evaluate a compact electronic nose (e-nose) system equipped with metal-oxide semiconductor sensors for the early detection of Candidatus Liberibacter asiaticus infection in citrus leaves through the analysis of volatile organic compounds under field-like conditions.

MATERIALS AND METHODS

This research starts from testing the source plant samples, sampling and data analysis testing conducted from December, 2022 to December, 2023.

Sample origin: All leaf samples were taken from 10-year-old live purworejo siamese citrus in Kemiri Village, Purworejo Regency, Central Java Province. To verify the HLB status of the source plant from which the sample material was derived, a conventional PCR test was performed by first extracting total DNA from 0.5 g of leaf midrib. The DNA was isolated using a modified CTAB buffer extraction method[13]. Conventional PCR assays were performed on DNA extracts using forward primer Las606 (5'-GGA GAG GTG AGT GGA ATT CCG A-3') and reverse primer LSS (5'-ACC CAA CAT CTA GGT AAA AAC C-3'). The primer sets were designed based on specific sequences located in the 3' region in all three 16S rDNAs of CLas that are conserved only in CLas[14].

Leaf samples from 19 of 31 citrus plants confirmed positive for CLas. Leaf samples from asymptomatic plants that tested positive for CLas were classed as 'CLas-infected' and plants in which CLas was not detected were called 'healthy'. Random leaf samples without evident insect or disease damage were taken from CLas-infected and healthy plants without assessing age or citrus canopy position. After collection, leaf samples were washed with sterile water, blotted dry with sterile paper and stored in polyethylene

Table 1: Distribution of citrus leaf sample data by orchard origin and shelf life

Characteristic	Asymptomatic CLas-Infected (230)	Healthy (224)	Total number
Orchard origin			
Orchard 1	106	69	175
Orchard 2	124	155	279
Sample shelf-life (days)			
1	88	128	216
2	63	27	90
3-5	79	67	148

Table 2: Specification of all chemoresistive sensors used in the e-nose system

Gas sensor	Detectable gases (cross-sensitivity)
S1	Carbon monoxide, ethanol, hydrogen, isobutane and methane
S2	Ammonia, ethanol, hydrogen, hydrogen sulphide and toluene
S3	Ethanol, hydrogen, isobutane and methane
S4	Carbon monoxide, ethanol, hydrogen, isobutane and methane
S5	Carbon monoxide, ethanol, hydrogen, isobutane, methane and propane
S6	Carbon monoxide, ethanol, hydrogen, isobutane, methane and propane
S7	Carbon monoxide, ethanol, hydrogen and methane
S8	Acetone, benzene, carbon monoxide, ethanol, isobutane, methane and n-hexane
S9	Ammonia, ethanol, hydrogen and isobutane
S10	Chlorofluorocarbons, ethanol and hydrofluorocarbons

bags at 4°C until processing using the e-nose test. Samples intended for HS-GC/MS analysis were meticulously transferred into GC/MS tubes for immediate testing.

A total of 454 leaf samples were analysed, consisting of 230 samples from asymptomatic CLas-infected plants and 224 samples from healthy plants. To avoid VOC depletion, each sample was evaluated once. Table 1 shows the distribution of tested leaves by orchard origin and shelf life.

Data was distributed by orchard origin based on e-nose-capable leaf samples. Shelf-life data was distributed using a random test sequence and e-nose daily testing capacity. The one-day shelf-life sample was examined most in this investigation. Because of the limited amounts of samples, notably shelf-life samples, all were pooled for modeling. Each sample was weighed at two grams for the e-nose test. The sample was placed in a beaker glass with a bespoke plastic top to attach to the e-nose inlet tube for VOC testing.

The HS-GC/MS sample testing was performed on four CLas-positive and four CLas-negative citrus plants. The plants' leaves were gathered, trimmed and divided into four subsamples per class. The leaf samples were then composited into healthy and asymptomatic CLas-infected samples. Two grams of each composite sample were placed in vials. The samples were promptly transferred to Gadjah Mada University's Integrated Research and Testing Institute for HS-GC/MS testing without duplication.

E-nose hardware and sampling protocol: The e-nose has 11 sensors: 10 MOS gas sensors and one temperature and humidity sensor (Table 2). These MOS gas sensors are known for their durability in detecting organic gases, although, like inorganic materials, they are sensitive to many gases[15]. The e-nose test uses non-scaled temperature-humidity sensors to monitor the sample chamber's temperature and humidity.

The MOS technology uses thin MOS plates as chemoresistors to detect changes in electrical conductivity caused by gas molecules adsorbing on their surfaces. It detects changes in the depletion layer at grain boundaries when exposed to oxidizing or reducing gases, which raises the energy barrier for free charge carriers (Fig. 1a). Consequently, sensing material resistance will change. Gas molecules that contact MOS material can donate or receive charge carriers, changing MOS resistivity. The e-nose has been utilized as a backup device in internal labs for less than three months. These settings allowed its use without calibration[16].

In the e-nose system (Fig. 1b), 11 sensors interacted with citrus leaf volatile organic compound (VOC) molecules at a specified concentration in the detection chamber. The e-nose was turned on for about 30 min before measuring. After 10 sec of delay, 60 sec of sampling and 180 sec of purging, each citrus leaf sample was evaluated for 250 sec. The VOC sampling system encapsulated identical citrus leaf samples in 30 mL clean glass vials with plastic caps during VOC emission. Before the extraction of VOCs, the samples are subjected to heating on an electric stove for a duration of 2 min. Custom acrylic tops sucked VOCs into the sensor chamber. Since the e-nose was ready, VOC collection and sensing began immediately. The procedures for sample preparation, VOC data extraction and the sensor equipment utilised were comparatively simpler than those described in the study conducted by Xu *et al.*[17]. In non-laboratory settings, VOC data extraction using more samples in less time takes 6 min.

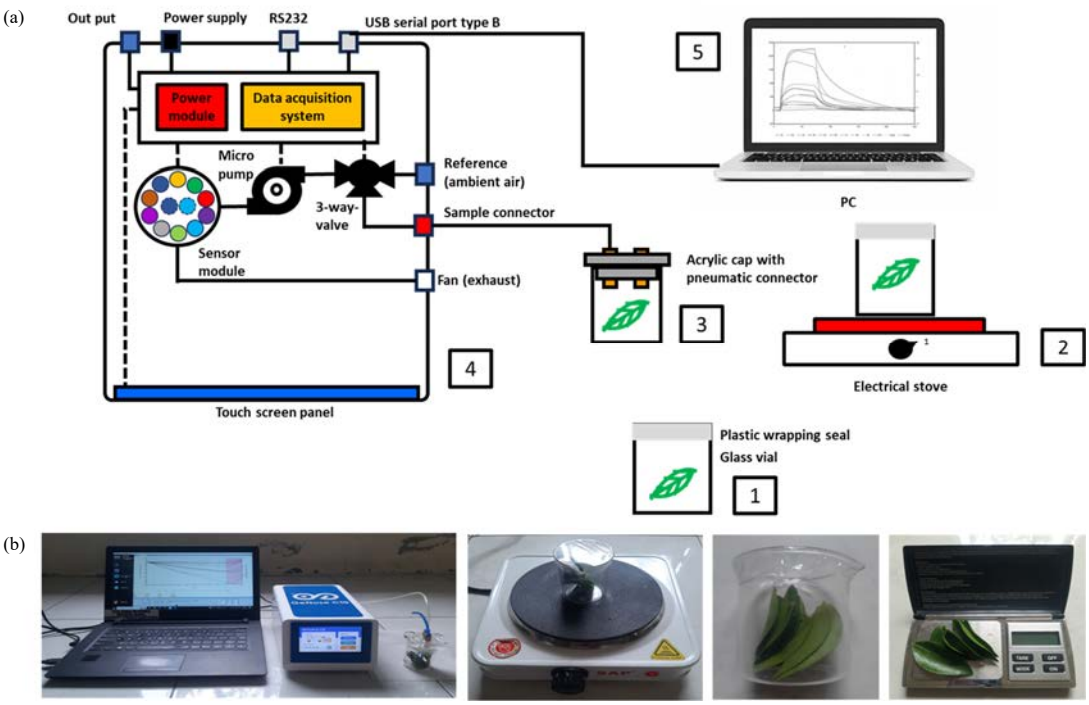

Fig. 1(a-b): (a) Citrus VOC sampling schematic: (1) Leaf samples incubated in sealed vial, (2) Heating, (3) VOC sampling, (4) E-nose sensing setup with sampling system, sensor array and DAQ system and (5) Data processing and (b) Photo of sample preparation and e-nose during citrus leaf test

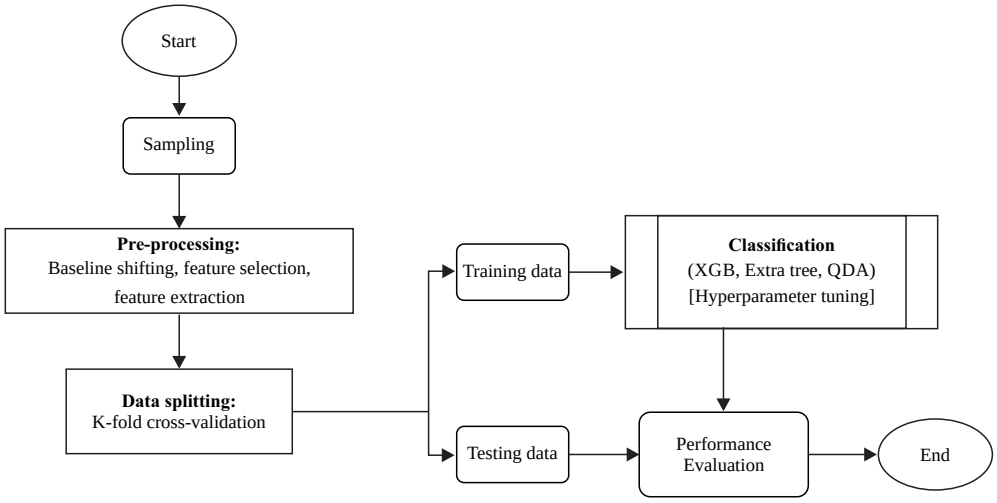

Fig. 2: Flow chart of the e-nose test

Signal acquisition and baseline shifting: The data matrix for multivariate statistical analysis of orange leaves includes e-nose S1-S10 signal profiles (collected by 10 MOS gas sensors) and temperature and humidity signal profiles (from a single sensor). The citrus leaf sample data matrix has 12 sensor responses x 2,502 rows. An electrical signal profile (Vijk) is generated for each e-nose sensor k during analysis for sample i at time j. For several seconds after sensing chamber exposure to citrus leaves, the sensor signals increased. After reaching a steady-state value, the signal declined at 60 sec.

The diverse signal profiles and fluctuations suggest that each sensor exhibits distinct and distinctive characteristics when subjected to the citrus leaf sample, contingent on the specific active materials and target gases employed. Furthermore, baseline signal levels differed between sensors,

even when the same citrus leaf samples were employed during the test. To ensure comparable trends across different measurements, the baseline-shifting process was applied according to their initial baselines. The baseline-corrected statistical analysis was conducted by the subtraction of all original sensor responses from the first response recorded during the delay phase[18]. This can be represented by the following Equation (1):

$$V'_{ijk} = V_{ijk} - \frac{\sum_{j=0}^{100-1} V_{ijk}}{100} \quad (1)$$

Equation (1) shows this, where is the shifted response or signal of the sensor in time stamp .

The different signal patterns and variations show that each sensor reacts differently to the citrus leaf sample depending on the active chemicals and target gases used. Although the test used the same citrus leaf samples, sensor baseline signal levels varied. To achieve comparable trends across measurements, baseline-shifting was applied according to initial baselines[19]. The baseline-corrected statistical analysis subtracted all original sensor responses from the first delay phase response[19].

Feature extraction and feature selection: The dimensions were reduced by extracting four features: Max, Mean, AUC and Mix based on Equations (2), (3) and (4)[20]. Once noisy residues from tube obstructions or environmental disturbances were removed, extraction began[21].

$$X_{ifk} = \frac{\max}{j \in \{100, 700-1\}} V'_{ijk} \quad (2)$$

$$X_{ifk} = \frac{\sum_{j=100}^{700-1} V'_{ijk}}{600} \quad (3)$$

$$X_{ifk} = \sum_{j=100}^{700-1} \frac{V'_{ijk} + V'_{i,j+1,k}}{2} \Delta tj \quad (4)$$

For each e-nose sensor k, an electrical signal profile was constructed across the analysis time period j for feature (f). $\Delta t = 0.1$ represents the e-nose sampling rate. Choosing the most distinct, instructional and limited features can improve data processing and storage efficiency[22].

Machine learning workflow: After feature extraction, a final matrix of 12 sensor responses × 454 samples × N features was created. Figure 2 shows the e-nose testing sequence for generating the most accurate machine learning to detect asymptomatic CLas-infected sample.

The pre-processed dataset was randomly split into training and testing sets. A classification chemometric model for the K-fold cross-validation approach (5 repetitions x5 folds, which ensured that at each validation run, the training data was left for internal validation purposes.. The model is externally validated (complete prediction) using the remaining test data. Formulae for accuracy (5), sensitivity (6) and specificity (7) of the developed model:

$$\text{Accuracy (\%)} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (5)$$

$$\text{Sensitivity (\%)} = \frac{TP}{TP + FN} \times 100\% \quad (6)$$

$$\text{Specificity (\%)} = \frac{TN}{TN + FP} \times 100\% \quad (7)$$

where, TP is true positive, TN is true negative, FP is false positive and FN is a false negative.

Headspace gas chromatography/mass spectrometry (HS-GC/MS): The HS-GC/MS testing was done on composite samples of asymptomatic CLas-infected and healthy citrus leaves. After equilibration at 110°C for 3 min, 1,100 µL of headspace was fed into a Thermo Trace 1310 GC with an ISQ-LT MS (HP-5MS UI column, 30 m × 0.25 mm × 0.25 µm An ISQ LT Single Quadrupole Mass Spectrometer mass selective detector (Thermo Fisher Scientific Inc, San Jose, CA, USA) was used for analysis. The GC oven's temperature plan started at 40°C for 5 min, heated to 180°C at 6°C per minute for 2 min and then increased to 240°C for 5 min. Ultra-high purity helium (UHP) was given at 1.00 mL/min. 230°C was the ionisation temperature, 150°C was the quadrupole temperature and 280°C was the MS transfer line temperature. Measurement delay (solvent delayed) was 3 min and the mass range was 30 to 500 AMU. Mass spectra library matching with the NIST 14 standard library and compound references identified compounds.

Statistical analysis: The dataset contained pre-processed signals from the 11 e-nose sensors during experiments. The assays were analysed using unsupervised and supervised pattern recognition methods and metric distance analysis. Principal Component Analysis (PCA) was used as an unsupervised pattern recognition method to evaluate the e-nose and classify citrus leaf samples by health state.

The performance of e-nose in detecting asymptomatic CLas infection in citrus leaf samples was further evaluated using Linear Discriminant Analysis (LDA), metric distance

analysis and Euclidean distance (Ed) and Mahalanobis distance (Md) analysis to evaluate features that could improve model accuracy. To find a model that accurately detects asymptomatic CLas infection, Extreme Gradient Boosting (XGBoost), the Extra Tree and Quadratic Discriminant Analysis (QDA) were used to test e-nose. Using SHapley Additive exPlanations (SHAP) 0.44 libraries, feature significance analysis was used to determine the best feature-model combination for identifying healthy and asymptomatic CLas-infected samples. Hyperparameter adjustment was used to optimise the model. The Python 3.11 using scikit-learn 1.4 was used for model construction and analysis.

The VOCs analysis data from HS-GC/MS results were chosen based on literature and the highest similarity index among the three matching spectrum libraries. The similarity index had to be above 700 and only one compound was considered for the final list if some were identified multiple times due to high concentration and tailing[23]. The initial normalisation of GC-MS data was calculated by summing all peak areas in a sample's total ion chromatogram (TIC), which was then set at 100%. Each peak percentage was calculated by dividing its area by the total compound area. Stacking bars to compare VOC patterns between healthy and asymptomatic samples confirmed that electronic noses can detect asymptomatic CLas-infected samples.

RESULTS AND DISCUSSION

Feature analysis: Due to their similar VOC fingerprints, the two classes overlap strongly, as does asymptomatic sugar-cane phytoplasma infection (Fig. 3a-d)[24].

In PCA, the Mean (Fig. 3b) and AUC (Fig. 3c) features explained 74.7% of the total variance, outperforming Max (Fig. 3a) or Mix (71.6 %) (Fig. 3d). Notwithstanding, not all studies have identified a strong correlation between the value of variance explained by the principal components and the classification performance. However, data dimensionality reduction by PCA is achieved by selecting features that have high variance, which in turn are commonly used for cluster analysis[25]. Therefore, both features have the potential to be used in building an accurate classifier model.

Using LDA, a supervised method, clearer class separation was achieved, with Mix feature (Fig. 3h) performing best, AUC feature (Fig. 3g) performing similarly to Mean feature (Fig. 3f) and Max feature (Fig. 3e) performing worst. The LDA feature analysis and PCA feature extraction were similar. The PCA and LDA models did not distinguish healthy and asymptomatic CLas-infected samples, but both multivariate analyses showed that the mean and AUC features could be used to build an asymptomatic CLas-infected detection model. The Max feature shows high overlapping horizontally on the same LDA score and vertically on different LD scores, while the Mean and AUC features show less overlapping (Fig. 3).

Figure 3 further reveals that the Mix feature's lowest overlapping level can effectively identify the two sample types in PCA but not LDA. The discrepancy in findings between PCA and LDA may be because, while both are commonly employed in recognition system development, PCA is used to simplify the data size without being supervised to capture the highest variance. However, LDA, a supervised classification method, seeks to find a linear combination of features that optimally differentiates predefined classes, improving its sensitivity in identifying VOC profiles with little difference between asymptomatic and healthy citrus leaves[26].

The optimal feature analysis for constructing a classification model can also be performed using either the Euclidean distance analysis or the Mahalanobis distance analysis. In the context of clustering analysis, both methods are frequently employed. The purpose of this analysis is to ascertain the distance between two data points that represent the variables to be clustered. In this study, the features demonstrating optimal potential for incorporation into a classification model are indicated by the greater distance between healthy samples and asymptomatic samples (Table 3).

Distance analysis showed Mix has the highest Euclidean (1.30) and Mahalanobis (4.99) class gaps (Table 3). The LDA plot analysis showed that the Mix feature reduced overlap between healthy sample plots and asymptomatic CLas-infected plots (Fig. 3). Mahalanobis distance (Md) consistently yielded greater remote distance analysis results for each characteristic examined than Euclidean distance (Ed). Euclidean distance analysis ignores scale and correlation, factors are independent and weighted equally. Mahalanobis distance analysis is more accurate for multivariate analysis since it considers scale and correlation[27].

Distance analysis has been reported to improve the accuracy of models for detecting plant diseases. Feature analysis using Euclidean distance has been shown to improve the accuracy of disease detection for brown spot, bacterial leaf wilt and Cercospora leaf spot, several citrus diseases[28]. Mahalanobis distance was used in discriminant analysis of Normalised Difference Spectral Index (NDSI) obtained from spectral reflectance measurements, forming a diagnostic model that could distinguish healthy and infected apple valsa canker (AVC) areas of apple branches with more than 94% accuracy[29].

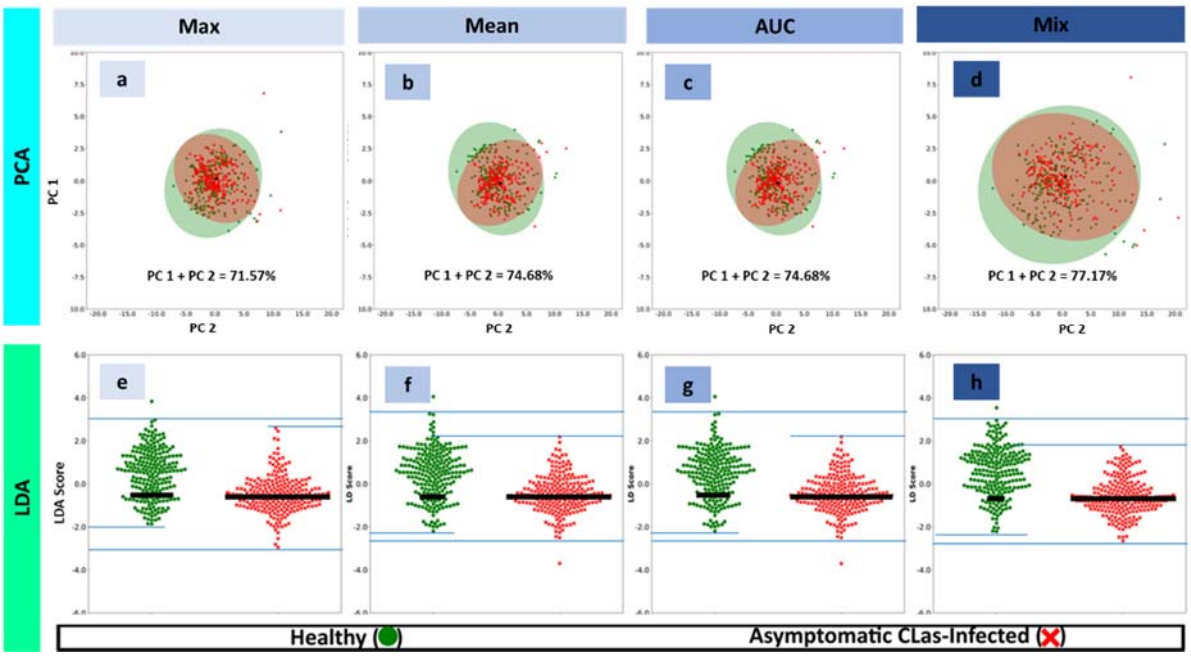

Fig. 3(a-h): (a-d) PCA and (e-h) LDA plots of feature extraction methods (max, mean, AUC and Mix), demonstrating improved class separation with the Mix feature, (a) PCA plot using max feature extraction, (b) PCA plot using mean feature extraction, (c) PCA plot using AUC feature extraction, (d) PCA plot using Mix feature extraction showing the highest explained variance despite overlap, (e) LDA plot using max feature extraction, (f) LDA plot using mean feature extraction, (g) LDA plot using AUC feature extraction and (h) LDA plot using Mix feature extraction showing the best class separation, indicated by the largest inter-class distance (green lines) and minimal overlap (shortest black lines)
Principal Component Analysis, LDA: Linear Discriminant Analysis, AUC: Area Under the Curve, Mix: Combined feature set (max, mean and AUC)

Table 3: Euclidean distance and mahalanobis distance analysis of different features

Features	Euclidean distance	Mahalanobis distance
Maximum	0.87	1.72
Mean	0.87	1.72
AUC	0.97	2.01
Mix	1.30	4.99

Best model performance for difference FE: Three classifiers (XGBoost, Extra Trees, QDA) were tested on four feature sets (Max, Mean, AUC, Mix) used in plant disease detection models[30]. Table 4 shows the 95% confidence interval (CI) best combination of each feature and model, whereas the supplementary material has all results.

The QDA model with the Mix feature fared badly despite having good distance metrics, while the XGBoost model with the AUC feature had the highest cross-validation accuracy of 83.2% (95% CI 79.3-87.1%). Features behave differently in distance analysis and machine learning models due to the basic differences in their similarity measurements. Unlike distance analysis, machine learning algorithms like XGBoost, Extra Tree and QDA can reveal more complex interactions in the VOCs dataset. This improves understanding of their disease state association[31]. A learning model's huge number of non-specific characteristics causes inconsistency in features that could increase classification accuracy when assessed using distance analysis and numerous learning models. It may even lower classifier model accuracy. Palit *et al.*[32], who studied machine learning's accuracy in analyses of Down syndrome risk factors, found this occurrence. For maize disease identification, gradient boosting captures complex, non-linear VOC interactions, hence XGBoost outperformed Extra Trees and QDA. Due to its fast and accurate training, XGBoost is said to perform better[33]. Table 4 shows that XGBoost with AUC and Extra Tree with Mean features yield the highest accuracy.

Feature importance analysis of sensor responses 1-10 from the e-nose was performed using the SHapley Additive exPlanations (SHAP) test to understand how each feature in the analysis model contributed to the predictions generated. In Fig. 4a-d below, for the Mix feature, features 0-9 are the

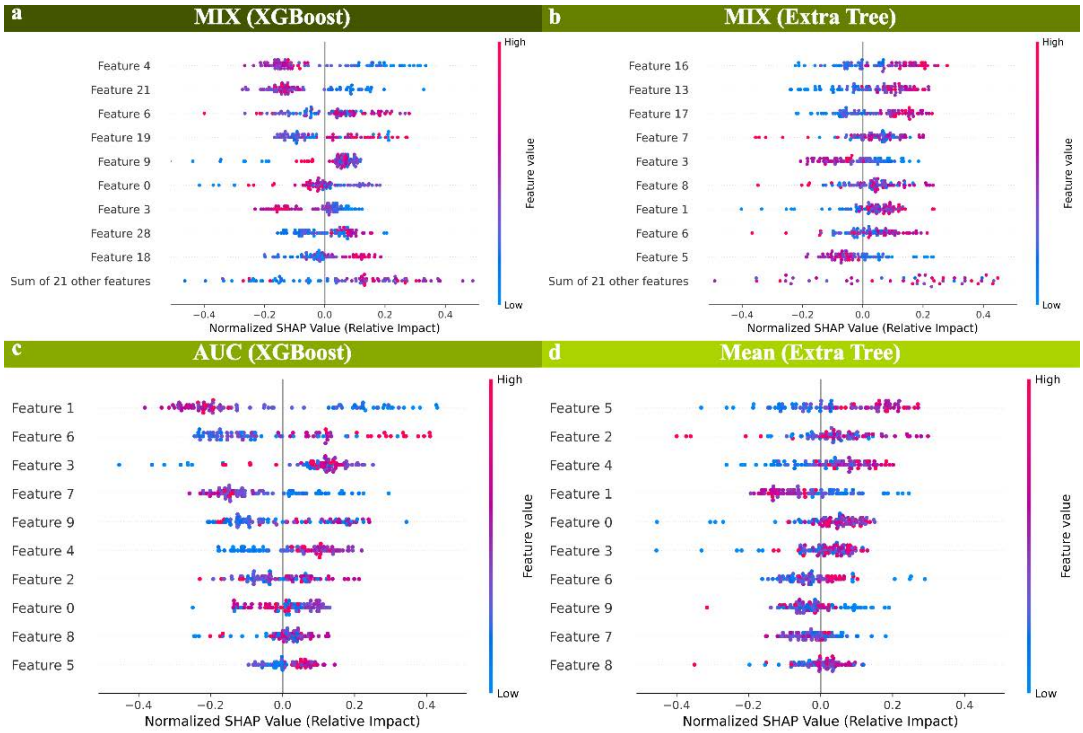

Fig. 4(a-d): Feature importance rankings of sensor responses (1-10) across feature extraction methods and classification models, highlighting the consistent contribution of sensor 7, (a) Feature 6 representing sensor 7 under FE max with a ranking of 3rd, (b) Feature 16 representing sensor 7 under FE mean with a ranking of 1st, (c) Feature 6 representing sensor 7 under FE max with a ranking of 2nd and (d) Feature 6 representing sensor 7 under FE max with a ranking of 7th
FE: Feature Extraction, AUC: Area Under the Curve

Table 4: Best model performance for difference features

Feature	Model	Accuracy (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)
Max	XGBoost	82.08	84.39	79.86
		(77.96, 85.67)	(78.82, 89.40)	(73.91, 85.31)
Mean	Extra tree	82.64	79.43	85.76
		(78.51, 86.50)	(73.40, 85.14)	(80.72, 90.43)
AUC	XGBoost	83.19	83.36	83.03
		(79.34, 87.05)	(77.17, 89.19)	(77.55, 88.02)
Mix	QDA	82.37	83.41	81.38
		(78.24, 86.23)	(77.78, 88.66)	(75.49, 87.03)

sensor responses of the Max feature, features 10-19 are the sensor responses of the Mean feature and features 20-29 are the sensor responses of the AUC feature. In Fig. 4a-c, it can be seen that XGBoost model the top three features consist of Max S5, AUC S2 and Max S7. Where this is in-line with performance in Table 4 with XGBoost show its best performance among another model in Max and AUC feature extraction. This pattern also found in Extra Tree where the top three feature is consisting of Mean S7, Mean S4 and Mean S8.

The SHapley Additive exPlanations (SHAP) test was used to analyze the importance of e-nose sensor responses 1-10 to see how each feature in the analysis model affected predictions. Features 0-9 are Max feature sensor responses, 10-19 are Mean feature sensor responses and 20-29 are AUC feature sensor responses for the Mix feature in Fig. 4. As illustrated in Fig. 4a-c, the top three features of the XGBoost model are Max S5, AUC S2 and Max S7. As shown in Table 4, XGBoost outperforms other models in Max and AUC feature extraction. The top three features in Extra Tree are Mean S7, Mean S4 and Mean S8.

Information or attributes used to identify the sensor pattern can affect sensor responses. Multiple models show S7 as a top contributor to model decision. 1st in Mix Extra Tree (Mean S7), 2nd in AUC XGBoost (Max S7), 3rd in Mix XGBoost (Max S7). The model is constantly affected by weather S7 response in the Maximum, AUC, or Mean values. Mix and AUC

Table 5: Best hyperparameter tuning performance

	AUC XGBoost		Mean extra tree	
Matrices	Original (%)	Hyperparameter tuning (a) (%)	Original (%)	Hyperparameter tuning (b) (%)
CI Accuracy	83.19	84.34	82.64	84.57
	(79.34, 87.05)	(80.17, 88.15)	(78.51, 86.50)	(80.98, 88.15)
CI Specificity	83.36	84.46	79.43	82.19
	(77.17, 89.19)	(78.53, 89.77)	(73.40, 85.14)	(76.09, 87.43)
CI Sensitivity	83.03	84.23	85.76	86.89
	(77.55, 88.02)	(78.76, 89.12)	(80.72, 90.43)	(81.93, 91.35)
Test accuracy	75.00	77.00	77.00	79.00
Test specificity	69.00	67.00	64.00	67.00
Test sensitivity	80.00	87.00	89.00	91.00

Model specification: (a) Tune to Accuracy Score. Booster = 'gbtree'; learning_rate = 0.4, max_depth = 4, (b) Tune to Accuracy Score. Class Weight = {0:1,1:5}, criterion: gini, n estimator = 300

XGBoost model also benefit from S2 AUC. S4 additionally shows Mix Extra Tree and AUC XGBoost's big influence. From this investigation, S7 has the greatest impact on model classification, followed by S2 and S4.

In addition to employing the optimal results from the feature extraction process, the performance accuracy of a model can be further enhanced through a tuning process. In the context of the three combinations of machine learning and features that demonstrated optimal performance, it was observed that only two of these combinations were optimised

Optimization of the best performing model: The hyperparameter tweaking method optimises machine learning model hyperparameters. Instead of learning model parameters during training, hyperparameters are defined beforehand[34]. This tuning technique determines the best hyperparameter configuration for model performance. This study sets the hyperparameter tuning model specifications differently for the optimized models, the AUC XGBoost and the Mean Extra Tree, as shown in Table 5 footnotes:

Table 5 shows that the Extra Tree machine learning model using feature mean performs best after hyperparameter adjustment with an accuracy of 84.57% (80.98%, 88.15%). The XGBoost feature AUC machine learning model follows with 84.34% (80.17%, 88.15%) accuracy. Hyperparameter tuning improves accuracy by identifying several parameter combinations that give optimal accuracy before training. In this study, the Extra Tree model's best hyperparameter configuration allocated equal weights to the healthy and asymptomatic classes, with a proportion of 3 to each (Class Weight = {0:1,1:35}). The model divides nodes using information measurement, which can improve decision Trees by incorporating data uncertainty (entropy) and the number of decision Trees to be created in an ensemble of 300 (n_estimator = 300). Decision Trees underpin each boosting iteration in the best hyperparameter configuration for the XGB model. Every training iteration adjusts the model parameters by 0.4 (learning_rate = 0.4) and each decision tree can only be 4 deep. Effective model hyperparameter adjustment improves accuracy, efficiency and generalisation[34].

The MOS-based e-nose with modest 84% accuracy detects unseen infections in situ, a unique feature. Field investigations for SARS-CoV-2 diagnosis must have 80% sensitivity[35]. The study incorporated samples with varying shelf-life, collected from two orchards. Despite the limited number of specimens within each storage category, the investigation demonstrated accurate detection in samples ranging from 1 to 3-5 days of shelf-life, thereby indicating the potential applicability of the findings under field conditions. This discovery lays the groundwork for field detection methods that speed up plant disease identification. The e-nose utilised in this investigation is portable, weighs just 5 kg and comes in a single device outside the sample container and hose for flowing VOCs into the sensor chamber. This study used a long-lasting, low-cost e-nose sensor array. The VOCs or VOC signals can be acquired in 6-7 min, compared to 20 min with Xu *et al.*[17] system.

Analysis based on Headspace-Gas Chromatography/Mass Spectrometry (HS-GC/MS): Infected plants create different Volatile Organic Compounds (VOCs) than healthy plants. Some investigations have found that diseased plants release unique chemical profiles of VOCs in larger proportions than healthy plants Other investigations have demonstrated that various VOCs are emitted in identical proportions by diseased and healthy plants, suggesting that they cannot be utilised as indicators of plant illness[36].

Research into using plant disease detection based on VOC patterns using chromatography techniques aims to distinguish plants infected with disease symptoms from healthy ones[36,37]. The HS-GC/MS confirmed class differences through 83 VOCs that were detected (Fig. 6) revealed lower

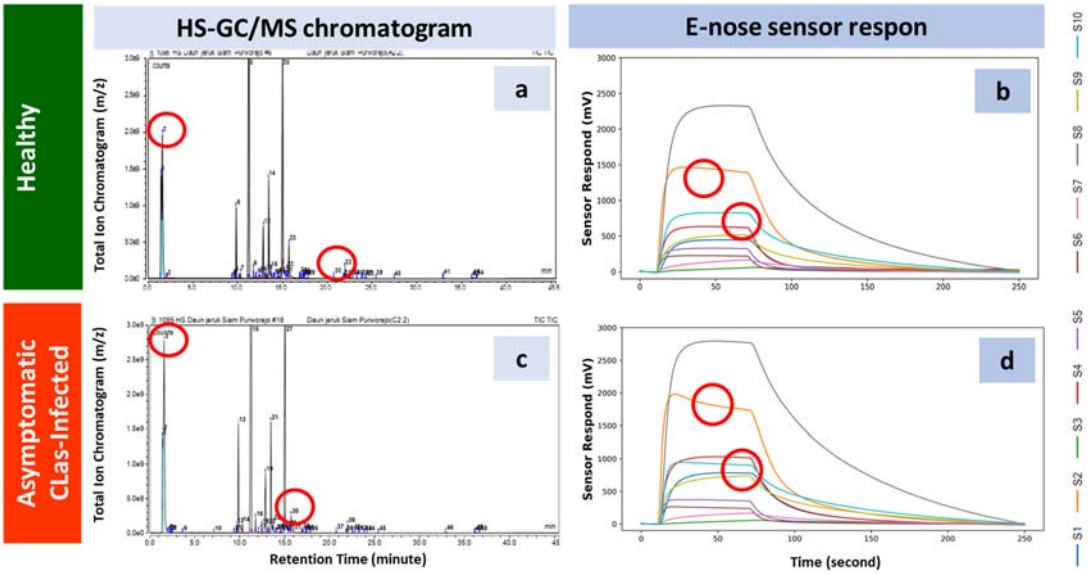

Fig. 5(a-d): HS-GC/MS chromatograms and e-nose sensor responses illustrating peak differences between healthy and asymptomatic CLas-infected samples, (a) HS-GC/MS chromatogram of the healthy sample showing baseline VOC profile, (b) E-nose sensor response pattern corresponding to the healthy sample, (c) HS-GC/MS chromatogram of the asymptomatic CLas-infected sample showing distinct peak variations and (d) E-nose sensor response of the asymptomatic CLas-infected sample, highlighting altered sensor signals

Red circles indicate key peak differences between sample classes, HS-GC/MS: Headspace Gas Chromatography-Mass Spectrometry, E-nose: Electronic nose, CLas: Candidatus Liberibacter asiaticus and VOCs: Volatile Organic Compounds

Table 6: HS-GCMS analysis of selected typical VOCs from sample leaves

			Area in chromatogram (x 103) (%)	
Numbers	Compound	Formula	Healthy	Asymptomatic CLas-infected
1	1,6-Nonadien-3-ol, 3,7-dimethyl-	$C_{11}H_{20}O$	69.60	0.00
2	2-Octen-1-ol, 3,7-dimethyl-	$C_{10}H_{20}O$	17.36	0.00
3	a-acorenol	$C_{15}H_{26}O$	0.00	8.56
4	Cyclohexanol, 1-methyl-4-(1-methylethenyl)-, cis-	$C_{10}H_{18}O$	114.74	118.78
5	Cyclopentanol, 2-methyl-, trans-	$C_6H_{12}O$	134.22	0.00
6	2-(4a,8-Dimethyl-6-oxo-1,2,3,4,4a,5,6,8a-octahydro-naphthalen-2-yl)-propionaldehyde	$C_{15}H_{22}O_2$	63.73	0.00
7	1,3,6,10-Dodecatetraene, 3,7,11-trimethyl-, (Z,E)-	$C_{15}H_{24}$	20.46	0.00
8	1,2-Dihydropyridine, 1-(1-oxobutyl)-	$C_9H_{13}NO$	122.20	0.00
9	l-Alanine ethylamide, (S)-	$C_5H_{12}N_2O$	13372.28	0.00
10	N(a)-Benzoyloxycarbonyl-N(b)-trimethylammonio-l-alanine, inner sal	$C_{14}H_{20}N_2O_4$	106.43	0.00
11	1,3-Benzodioxol-2-one, hexahydro-, cis-	$C_7H_{10}O_3$	116.36	0.00
12	Thiophene-3-ol, tetrahydro-, 1,1-dioxide	$C_4H_8O_3S$	173.92	0.00
13	(R)-lavandulyl acetate	$C_{12}H_{20}O_2$	13.78	0.00
14	Cyclopentaneacetic acid, 3-oxo-2-pentyl-, methyl ester	$C_{13}H_{22}O_3$	65.89	0.00
15	Heptadecanoic acid, 16-methyl-, methyl ester	$C_{19}H_{38}O_2$	50.73	0.00
16	Hexadecanoic acid, methyl ester	$C_{17}H_{34}O_2$	603.52	466.14
17	2-Butanone, 4-(2,6,6-trimethyl-2-cyclohexen-1-ylidene)-	$C_{13}H_{20}O$	13.81	0.00
18	(1R,2S,6S,7S,8S)-8-Isopropyl-1-methyl-3-methylenetricyclo[4.4.0.02,7]decane-rel-	$C_{15}H_{24}$	0.00	369.79
19	Bicyclo[2.2.1]heptane, 2,2-dimethyl-3-methylene-, (1S)-	$C_{10}H_{16}$	94.46	237.77
20	b-Ocimene	$C_{10}H_{16}$	2787.15	5098.14
21	Caryophyllene	$C_{15}H_{24}$	0.00	897.34
22	Cyclohexane, 1-ethenyl-1-methyl-2,4-bis(1-methylethenyl)-, [1S-(1a,2b,4b)]-	$C_{15}H_{24}$	41.49	4859.86
23	Cyclopropane, trimethyl(2-methyl-1-propenylidene)-	$C_{10}H_{16}$	14.44	0.00
24	Farnesene epoxide, E-	$C_{15}H_{24}O$	63.93	0.00
25	Guaia-1(10),11-diene	$C_{15}H_{24}$	0.00	115.25
26	Humulene	$C_{15}H_{24}$	0.00	580.62
27	Linalool	$C_{10}H_{18}O$	20396.41	35311.01
28	Terpineol	$C_{10}H_{18}O$	47.60	1344.30

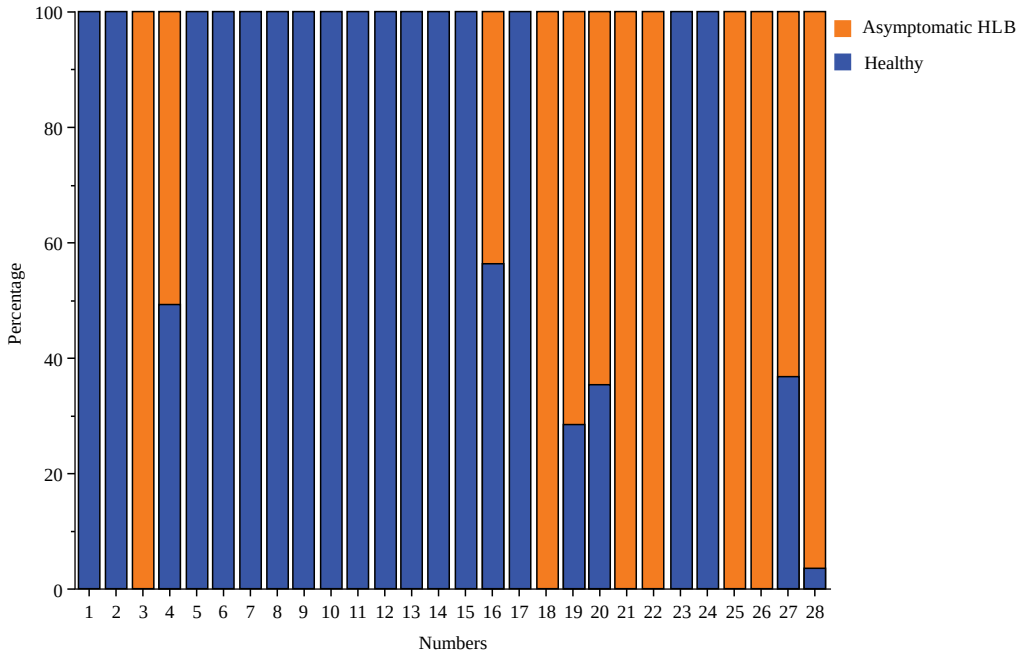

Fig. 6: Percentage stacked bar chart of selected volatile organic compound of healthy and asymptomatic CLas-infected samples

total VOC abundance in asymptomatic leaves, especially alcohols, esters and terpenes. The e-nose-based VOCs data analysis distinguished healthy citrus leaf samples fromasymptomatic CLas-infected samples with above 80% accuracy with 95% CI. HS-GC/MS VOC data analysis showed differences in VOC type and content between healthy and asymptomatic CLas-infected citrus leaf samples.

Because MOS sensors are cross-sensitive, VOC peaks and sensor channels cannot be mapped directly. SHAP's most important sensors in the classification model (Fig. 4) target alcohol, one of the most common volatile organic compounds (VOCs) in the gas chromatography (GC) profile (Table 2 and Fig. 5).

A total of 83 Volatile Organic Compounds (VOCs), including alcohol, aldehyde, aliphatic hydrocarbon, alkane, amine, cyclic organic, ester, ketone, nitrogenous compound and terpene, were identified from two types of citrus leaf samples (healthy and asymptomatic CLas-infected). The results of the analysis of randomly selected VOCs using HS-GC/MS are shown in Table 6. The differences in the relative content of VOCs are shown in the form of percentage stacked bar charts in Fig. 6. A comparison of the concentrations of VOCs in healthy and asymptomatic CLas-infected leaf samples revealed that some VOCs exhibited an increase, while others demonstrated a decrease. Notably, the presence of certain VOCs was exclusively observed in healthy citrus leaf samples, while other VOCs were exclusively detected in asymptomatic citrus leaf samples. This phenomenon of differential VOC patterns between healthy and infected samples has also been observed in studies conducted by Gottwald *et al.*[4] and Xu *et al.*[17]. Table 6 presents the HS-GCMS analysis of selected volatile organic compounds (VOCs) in healthy and asymptomatic CLas-infected leaves. Healthy leaves showed high levels of compounds such as l-alanine ethylamide, hexadecanoic acid methyl ester and b-ocimene, whereas several VOCs, including α-acorenol, caryophyllene and humulene, were prominent in asymptomatic infected leaves. Some compounds, like cyclohexanol derivatives, were present in both leaf types with minor differences. Overall, VOC profiles markedly differed between healthy and infected samples, indicating infection-induced metabolic shifts.

Figure 7a-c show that both sample types had different VOC numbers and percentages. Asymptomatic CLas-infected samples have much lower VOC levels than healthy samples. Alcohols, aliphatic hydrocarbons and esters are higher in healthy samples than in asymptomatic samples. A VOC category's content can be significant despite its numerical insignificance. To avoid overlooking the large VOC category differences between the two sample types in this phenomenon, unit content was computed by dividing the proportional amount of content in one VOC category by the overall number of VOCs. In Fig 7b-d, aldehydes and amines exhibited the highest unit content and these two VOCs, along with cyclic organics and ketones were the only VOCs found in healthy samples. Due to HLB stress, asymptomatic CLas-infected leaves had higher unit concentrations of alcohols,

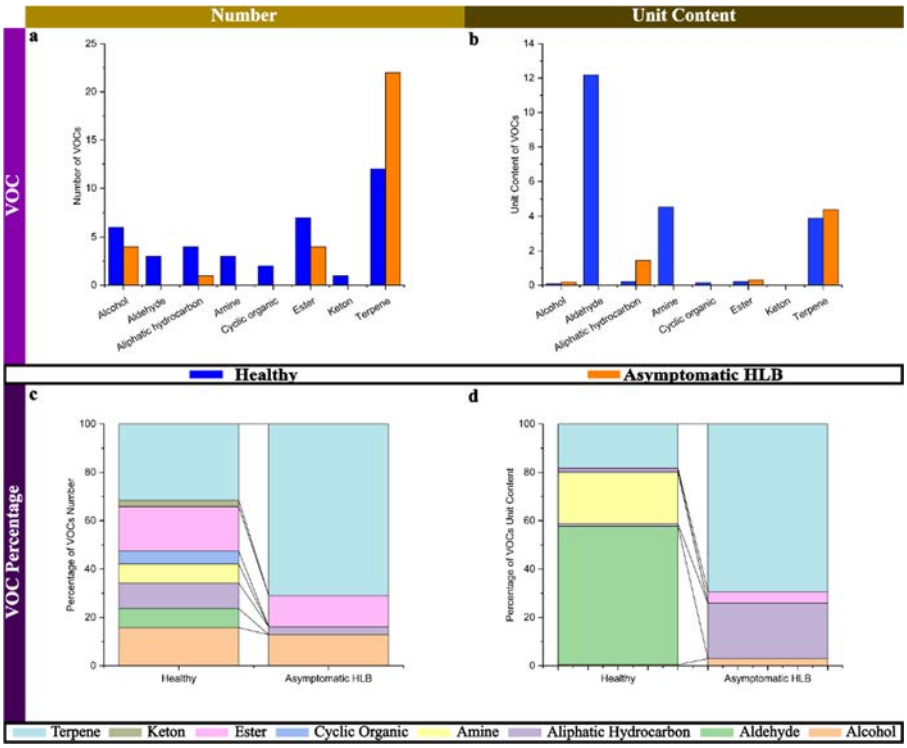

Fig. 7(a-d): (a) Number of volatile organic compounds (VOCs) in the two types of samples (healthy, asymptomatic CLas-infected), (b) Unit content of VOCs, (c) Percentage of VOCs number and (d) Percentage of VOCs unit content

aliphatic hydrocarbons, esters and terpenes than healthy samples. Citrus plant defensive and communication mechanisms change VOC content and concentration in response to external stresses[38]. Earlier research found that black zira plants (*Bunium persicum*) infected with soil-borne plant pathogenic fungus[39] formed terpene chemicals.

In summary, the findings demonstrate the presence of differential emissions from citrus leaves in response to diverse biotic stresses. These variations extend beyond simple fluctuations in the levels of individual VOCs, encompassing multiple categories of VOCs. The findings from the HS-GCMS analysis corroborate the outcomes of VOC data analysis employing e-nose, confirming that e-nose can detect asymptomatic samples The e-nose responds to the varied patterns exhibited by diverse types of VOCs in both sample types through a more streamlined process, yielding a combination of signals that serve as fingerprints for each sample. This facilitates the identification of complex VOCs[40]. The present study utilised a single portable electronic nose unit with a mass of 5 kg, in conjunction with samples obtained from two orchards. With rapid (6-7 min) measurements, the prototype halves the processing time reported by Xu *et al.*[17] and demonstrates feasibility for orchard-level surveys. After tuning, Mean-Extra Trees achieved the top CV accuracy of 84.6 % (80.9-88.2%), followed closely by AUC-XGBoost at 84.3% (80.2-88.1%).

Conventional PCR was used to confirm the presence of CLas in samples, allowing for comparison with sensor results, Real-time PCR is more sensitive than conventional PCR, capable of detecting lower quantities of nucleic acid in a sample. Therefore it should be noted that some samples recorded as healthy may have been infected with CLas below detectable levels. However the analysis was completed with the data obtained using conventional PCR. Only measuring in two orchards, single-season sampling and little external validation are additional limitations in this study. To identify VOC biomarkers, future trials should use multi-site data, longer sensor-drift monitoring or periodic recalibration and larger GC-MS panels, as well as molecular confirmation of CLas detection using more sensitive real-time PCR assays. Automating VOC enrichment and embedding the adjusted Extra-Trees model on the device micro-controller are commercial deployment priorities.

CONCLUSION

In this study, e-nose gas sensors with 10 Metal Oxide Semiconductors (MOS) gas sensors and one temperature/humidity sensor can distinguish asymptomatic CLas-infected citrus leaves from healthy controls with a cross-validated accuracy of 84.6 % (95 % CI 80.9-88.2%). The sensor-based diagnosis was supported by GC-MS, which showed

unique Volatile Organic Compound (VOC) profiles for the two leaf classes. Our discovery closes a major surveillance gap by detecting concealed, asymptomatic tissue, unlike previous MOS e-nose studies. *In-situ* operation is fast (seven minutes per sample), cheap and needs little sample preparation, making the prototype appropriate for on-orchard triage before laboratory confirmation. Our findings suggest MOS-based e-nose technology could be a useful early-warning tool for HLB management and other plant diseases with mild VOC indicators before visual signs. This study shows that e-nose can identify asymptomatic CLas-infected samples. The HLB detection in the field is possible due to the simplicity of sample collection and processing.

SIGNIFICANCE STATEMENT

Under in situ conditions, our e-nose prototype equipped with 10 MOS gas sensors and a temperature/humidity sensor successfully distinguished asymptomatic CLas-infected citrus leaves from healthy ones with a cross-validated accuracy of 84.6% (95% CI 80.9-88.2%), demonstrating a rapid, low-cost and field-deployable diagnostic approach that fills the surveillance gap for early detection before visible symptoms appear.

This study demonstrates the potential of a lab-made electronic nose (e-nose) with 10 MOS sensors to rapidly detect early-stage Huanglongbing (HLB) infections in citrus leaves. By identifying distinct VOC fingerprints before visible symptoms, the e-nose enables timely field screening, prioritizing samples for PCR confirmation. This approach offers a practical, non-destructive tool to enhance HLB monitoring and support proactive disease management strategies.

REFERENCES

1. Bové, J.M., 2006. Huanglongbing: A destructive, newly-emerging, century-old disease of citrus. J. Plant Pathol., 88: 7-37.
2. Jagoueix, S., M.J. Bove and M. Garnier, 1996. PCR detection of the two '*Candidatus*' liberobacter species associated with greening disease of citrus. Mol. Cell. Probes, 10: 43-50.
3. Cui, S., P. Ling, H. Zhu and H. Keener, 2018. Plant pest detection using an artificial nose system: A review. Sensors, Vol. 18. 10.3390/s18020378.
4. Gottwald, T.R., 2010. Current epidemiological understanding of citrus huanglongbing. Annu. Rev. Phytopathol., 48: 119-139.
5. Valdés, R.A., J.C.D. Ortiz, M.B. Beache, J.A. Cabello, E.C. Chávez, Y.R. Pagaza and Y.M.O. Fuentes, 2 016. A review of techniques for detecting Huanglongbing (greening) in citrus. Can. J. Microbiol., 62: 803-811.
6. Sankaran, S., A. Mishra, R. Ehsani and C. Davis, 2010. A review of advanced techniques for detecting plant diseases. Comput. Electron. Agric., 72: 1-13.
7. Achor, D.S., E. Etxeberria, N. Wang, S.Y. Folimonova, K.R. Chung and L.G. Albrigo, 2010. Sequence of anatomical symptom observations in citrus affected with huanglongbing disease. Plant Pathol. J., 9: 56-64.
8. Fang, Y. and R. Ramasamy, 2015. Current and prospective methods for plant disease detection. Biosensors, 5: 537-561.
9. Sankaran, S. and R. Ehsani, 2012. Detection of huanglongbing disease in citrus using fluorescence spectroscopy. Trans. ASABE, 55: 313-320.
10. Sankaran, S., J.M. Maja, S. Buchanon and R. Ehsani, 2013. Huanglongbing (citrus greening) detection using visible, near infrared and thermal imaging techniques. Sensors, 13: 2117-2130.
11. Pourreza, A., W.S. Lee and R. Ehsani, 2014. A vision based sensor for Huanglongbing disease detection under a simulated field condition. Am. Soc. Agric. Biol. Eng., 10.13031/aim.20141900251.
12. Putri, L.A., I. Rahman, M. Puspita, S.N. Hidayat and A.B. Dharmawan *et al*., 2023. Rapid analysis of meat floss origin using a supervised machine learning-based electronic nose towards food authentication. npj Sci. Food, Vol. 7. 10.1038/s41538-023-00205-2.
13. Sena-Esteves, M. and G. Gao, 2020. Introducing genes into mammalian cells: Viral vectors. Cold Spring Harbor Protoc., 2020: 297-329.
14. Fujikawa, T. and T. Iwanami, 2012. Sensitive and robust detection of citrus greening (huanglongbing) bacterium "*Candidatus Liberibacter asiaticus*" by DNA amplification with new 16S rDNA-specific primers. Mol. Cell. Probes, 26: 194-197.
15. Isaac, N.A., I. Pikaar and G. Biskos, 2022. Metal oxide semiconducting nanomaterials for air quality gas sensors: Operating principles, performance, and synthesis techniques. Microchim. Acta, Vol. 189. 10.1007/s00604-022-05254-0.
16. Romain, A.C. and J. Nicolas, 2010. Long term stability of metal oxide-based gas sensors for e-nose environmental applications: An overview. Sensors Actuat. B: Chem., 146: 502-506.
17. Xu, Q., J. Bai, L. Ma, Z. Li, B. Tan, L. Sun and J. Cai, 2023. Identification of multiple symptoms of huanglongbing by electronic nose based on the variability of volatile organic compounds. Ann. Appl. Biol., 183: 181-195.
18. Alday, P.M., 2019. How much baseline correction do we need in ERP research? Extended GLM model can replace baseline correction while lifting its limits. Psychophysiology, Vol. 56. 10.1111/psyp.13451.

19. Xie, R., M. Zhang, P. Venkatraman, X. Zhang and G. Zhang *et al*., 2019. Normalization of large-scale behavioural data collected from zebrafish. PLoS ONE, Vol. 14. 10.1371/journal.pone.0212234.
20. Yan, J., X. Guo, S. Duan, P. Jia, L. Wang, C. Peng and S. Zhang, 2015. Electronic nose feature extraction methods: A review. Sensors, 15: 27804-27831.
21. Chandrashekar, G. and F. Sahin, 2014. A survey on feature selection methods. Comput. Electr. Eng., 40: 16-28.
22. Ruano-Ordás, D., 2024. Machine learning-based feature extraction and selection. Appl. Sci., Vol. 14. 10.3390/app14156567.
23. Mills, T.C., 2014. Testing for Stability in Regression Models. In: Analysing Economic Data: A Concise Introduction, Mills, T.C. (Eds.), Palgrave Macmillan London, London, United Kingdom, ISBN: 978-1-137-40190-8, pp: 243-259.
24. Saengmanee, P., P. Burns, J. Watcharachaiyakup, U. Lertsuchatavanich and P. Wanichananan *et al*., 2024. Morphological and biochemical changes in asymptomatic and moderately symptomatic plants infected with sugarcane white leaf (SCWL) phytoplasma. J. Plant Pathol., 106: 1773-1784.
25. Yeung, K.Y. and W.L. Ruzzo, 2001. Principal component analysis for clustering gene expression data. Bioinformatics, 17: 763-774.
26. Martinez, A.M. and A.C. Kak, 2001. PCA versus LDA. IEEE Trans. Pattern Anal. Mach. Int., 23: 228-233.
27. Suárez, J.L., S. García and F. Herrera, 2021. A tutorial on distance metric learning: Mathematical foundations, algorithms, experimental analysis, prospects and challenges. Neurocomputing, 425: 300-322.
28. Mohanty, S.P., D.P. Hughes and M. Salathé, 2016. Using deep learning for image-based plant disease detection. Front. Plant Sci., Vol. 7. 10.3389/fpls.2016.01419.
29. Kurihara, J. and T. Yamana, 2022. Detection of apple valsa canker based on hyperspectral imaging. Remote Sens., Vol. 14. 10.3390/rs14061420.
30. Jackulin, C. and S. Murugavalli, 2022. A comprehensive review on detection of plant disease using machine learning and deep learning approaches. Meas.: Sens., Vol. 24. 10.1016/j.measen.2022.100441.
31. Bzdok, D., M. Krzywinski and N. Altman, 2017. Machine learning: A primer. Nat. Methods, 14: 1119-1120.
32. Palit, O.O., R.P. Dhenanta, A.I. Susanto, A.M. Syawly and A.L. Ivansyah *et al*., 2024. Accuracy of machine learning methods in the analysis of important variables of down syndrome risk factors [In Indonesian]. Indones. J. Comput. Sci., 13: 8001-8015.
33. Prakash, R.V. and G. Kirubakaran, 2023. Comprehensive analysis of corn and maize plant disease detection and control using various machine learning algorithms and internet of things. Philipp. J. Sci., 152: 2245-2251.
34. Ilemobayo, J.A., O. Durodola, O. Alade, O.J. Awotunde and A.T. Olanrewaju *et al*., 2024. Hyperparameter tuning in machine learning: A comprehensive review. J. Eng. Res. Rep., 26: 388-395.
35. Scohy, A., A. Anantharajah, M. Bodéus, B. Kabamba-Mukadi, A. Verroken and H. Rodriguez-Villalobos, 2020. Low performance of rapid antigen detection test as frontline testing for COVID-19 diagnosis. J. Clin. Virol., Vol. 129. 10.1016/j.jcv.2020.104455.
36. Ficke, A., B. Asalf and H.R. Norli, 2022. Volatile organic compound profiles from wheat diseases are pathogen-specific and can be exploited for disease classification. Front. Microbiol., Vol. 12. 10.3389/fmicb.2021.803352.
37. Wenda-Piesik, A., 2011. Volatile organic compound emissions by winter wheat plants (*Triticum aestivum* L.) under *Fusarium* spp. infestation and various abiotic conditions. Polish J. Environ. Stud., 20: 1335-1342.
38. Hazarika, S., R. Choudhury, B. Montazer, S. Medhi, M.P. Goswami and U. Sarma, 2020. Detection of citrus tristeza virus in mandarin orange using a custom-developed electronic nose system. IEEE Trans. Instrum. Meas., 69: 9010-9018.
39. Sekine, T., M. Sugano, A. Majid and Y. Fujii, 2007. Antifungal effects of volatile compounds from black zira (*Bunium persicum*) and other spices and herbs. J. Chem. Ecol., 33: 2123-2132.
40. Hartyáni, P., I. Dalmadi and D. Knorr, 2013. Electronic nose investigation of *Alicyclobacillus acidoterrestris* inoculated apple and orange juice treated by high hydrostatic pressure. Food Control, 32: 262-269.