UNIVERSITAS
GADJAH MADA
Detection of Asymptomatic Candidatus Liberibacter asiaticus-Infected Citrus using an Electronic Nose
Anthoni Sulthan Harahap, Prof. Dr. Ir. Siti Subandiyah, M.Agr.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

TABLE OF CONTENTS

Page

STATEMENT ..... **Error! Bookmark not defined.**

PREFACE ..... **Error! Bookmark not defined.**

TABLE OF CONTENTS 3

LIST OF TABLES 6

FIGURE LIST 8

LIST OF ATTACHMENTS 12

I. INTRODUCTION ..... **Error! Bookmark not defined.**

1.1. Background ..... **Error! Bookmark not defined.**

1.2. Problem ..... **Error! Bookmark not defined.**

1.3. Objectives ..... **Error! Bookmark not defined.**

1.4. Benefits ..... **Error! Bookmark not defined.**

1.5. Authenticity ..... **Error! Bookmark not defined.**

II. LITERATURE REVIEW AND THEORY BASIS**Error! Bookmark not defined.**

2.1. Huanglongbing Disease ..... **Error! Bookmark not defined.**

2.1.1. Pathogen, host plants, symptoms, *C*Las life cycle, and impact ..... **Error! Bookmark not defined.**

2.1.2. Host metabolomic responses to HLB**Error! Bookmark not defined.**

2.2. *C*Las Detection Based on VOC Monitoring**Error! Bookmark not defined.**

2.2.1. *C*las detection using HS-GC/MS . **Error! Bookmark not defined.**

2.2.2. Sensor-based *C*Las detection ..... **Error! Bookmark not defined.**

2.2.3. Type sample and sample heating prior to e-nose testing .... **Error! Bookmark not defined.**

2.3. Machine Learning Methods for VOC Classification**Error! Bookmark not defined.**

2.3.1. Unsupervised methods: PCA ..... **Error! Bookmark not defined.**

UNIVERSITAS
GADJAH MADA
Detection of Asymptomatic Candidatus Liberibacter asiaticus-Infected Citrus using an Electronic Nose
Anthoni Sulthan Harahap, Prof. Dr. Ir. Siti Subandiyah, M.Agr.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

2.3.2. Supervised discriminant methods **Error! Bookmark not defined.**

2.3.3. Support vector machine (SVM) ... **Error! Bookmark not defined.**

2.3.4. Ensemble methods..... **Error! Bookmark not defined.**

2.3.5. Partial least squares methods ..... **Error! Bookmark not defined.**

2.4. ROC Analysis for Biomarker Discovery **Error! Bookmark not defined.**

2.4.1. Univariate ROC analysis ..... **Error! Bookmark not defined.**

2.4.2. Multivariate ROC analysis ..... **Error! Bookmark not defined.**

2.4.3. Exploration of putative biomarkers **Error! Bookmark not defined.**

2.5. Theoretical Foundation ..... **Error! Bookmark not defined.**

2.6. Hypothesis ..... **Error! Bookmark not defined.**

III. RESEARCH METHODOLOGY ..... **Error! Bookmark not defined.**

3.1. Materials and Equipment ..... **Error! Bookmark not defined.**

3.2. Time and Location ..... **Error! Bookmark not defined.**

3.3. Research Design ..... **Error! Bookmark not defined.**

3.3.1. Experiment 1 (Preparation of test materials for e-nose and HS-GC/MS analysis)..... **Error! Bookmark not defined.**

3.3.2. Experiment 2 (Detection asymptomatic *C*Las infection using e-nose)..... **Error! Bookmark not defined.**

3.3.3. Experiment 3 (VOC analysis using HS-GC-MS for corroborating e-nose detection results)..... **Error! Bookmark not defined.**

3.3.4. Experiment 4 (analysis putative biomarkers for detecting asymptomatic *C*Las infection)..... **Error! Bookmark not defined.**

IV. RESULTS AND DISCUSSION..... **Error! Bookmark not defined.**

4.1. Experiment 1 (Preparation of E-nose and HS-GC/MS Test Samples) ..... **Error! Bookmark not defined.**

4.1.1. *C*Las-positive plants source ..... **Error! Bookmark not defined.**

4.1.2. Optimal sample preparation for enose testing and HS-GC/MS ..... **Error! Bookmark not defined.**

UNIVERSITAS
GADJAH MADA
Detection of Asymptomatic Candidatus Liberibacter asiaticus-Infected Citrus using an Electronic Nose
Anthoni Sulthan Harahap, Prof. Dr. Ir. Siti Subandiyah, M.Agr.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

4.2. Experiment 2 (Detection Asymptomatic *C*Las Infection Using E-nose) ..... **Error! Bookmark not defined.**

4.2.1. Windowing for the classification and performance optimisation..... **Error! Bookmark not defined.**

4.2.2. Windowing optimisation of the best performing model ..... **Error! Bookmark not defined.**

4.2.3. Hyperparameter for the classification and performance optimisation..... **Error! Bookmark not defined.**

4.2.4. Biological Basis of E-Nose Sensor Selectivity towards HLB **Error! Bookmark not defined.**

4.3. Experiment 3 (VOCs Analysis Using HS-GC-MS for Corroborating E-nose Detection Results) ..... **Error! Bookmark not defined.**

4.4. Experiment 4 (Analysis Putative Biomarkers for Detecting asymptomatic *C*Las infection)..... **Error! Bookmark not defined.**

4.4.1. Compounds variability analysis and multivariate approaches for patter recognition ..... **Error! Bookmark not defined.**

4.4.2. Receiver operating characteristics (ROC) curve analysis. .. **Error! Bookmark not defined.**

4.4.3. Putative biomarker for detecting asymptomatic *C*Las infections and their biological relevance ..... **Error! Bookmark not defined.**

4.5. General Discussion..... **Error! Bookmark not defined.**

4.5.1. Methodological considerations and sample preparation..... **Error! Bookmark not defined.**

4.5.2. E-nose performance and machine learning optimisation..... **Error! Bookmark not defined.**

4.5.3. Integrating e-nose and HS-GC/MS analyses **Error! Bookmark not defined.**

4.5.4. VOC profiling and putative biomarker discovery **Error! Bookmark not defined.**

4.5.5. Comparison with previous studies and contextualisation .... **Error! Bookmark not defined.**

UNIVERSITAS
GADJAH MADA
Detection of Asymptomatic Candidatus Liberibacter asiaticus-Infected Citrus using an Electronic Nose
Anthoni Sulthan Harahap, Prof. Dr. Ir. Siti Subandiyah, M.Agr.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

4.5.6. Limitations and future directions.. **Error! Bookmark not defined.**

V. CONCLUSIONS AND SUGGESTIONS **Error! Bookmark not defined.**

5.1. Conclusions ..... **Error! Bookmark not defined.**

5.2. Suggestions ..... **Error! Bookmark not defined.**

VI. REFERENCES ..... **Error! Bookmark not defined.**

VII. ATTACHMENTS ..... **Error! Bookmark not defined.**

LIST OF TABLES

Page

Table 1. Required materials and equipment corresponding to research activities ..... **Error! Bookmark not defined.**

Table 2. Total sample distribution by orchard origin, health status, and sample shelf-life. ..... **Error! Bookmark not defined.**

Table 3. The amount of data per treatment of e-nose and HS-GC/MS sample preparation. ..... **Error! Bookmark not defined.**

Table 4. Distribution of citrus leaf sample data by orchard origin and shelf-life for e-nose testing. ..... **Error! Bookmark not defined.**

Table 5. Specification of all chemo-resistive sensors used in the e-nose system. ..... **Error! Bookmark not defined.**

Table 6. Distribution of citrus leaf sample data by orchard origin and shelf-life after performing curation on noisy data. **Error! Bookmark not defined.**

UNIVERSITAS
GADJAH MADA
Detection of Asymptomatic Candidatus Liberibacter asiaticus-Infected Citrus using an Electronic Nose
Anthoni Sulthan Harahap, Prof. Dr. Ir. Siti Subandiyah, M.Agr.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Table 7. Performance of the dummy model with variations in the maximum, mean, and AUC (area under the curve).**Error! Bookmark not defined.**

Table 8. Presents the performance of several models using the mean extraction..... **Error! Bookmark not defined.**

Table 9. Presents the performance of the Extra tree and QDA models when varying the number of windows using the mean feature. ..... **Error! Bookmark not defined.**

Table 10. Permutation feature importance of quadratic discriminant analysis (QDA) model with mean windowing, relative humidity, and temperature..... **Error! Bookmark not defined.**

Table 11. Euclidean distance and mahalanobis distance analysis of different features. ..... **Error! Bookmark not defined.**

Table 12. Performance comparison of the best-performing machine learning models (XGBoost, Extra trees, and QDA) using four different feature extraction methods (max, mean, AUC, and mix) for classifying asymptomatic *CLas*-infected on the HLB test dataset. ..... **Error! Bookmark not defined.**

Table 13. Best classification performance achieved by extreme gradient boosting (XGBoost) and extreme randomised trees (Extra trees) models on area under the curve (AUC) and mean feature sets for asymptomatic CLas-infected citrus, comparing original versus tuned hyperparameters. ..... **Error! Bookmark not defined.**

Table 14. Optimal gas targets and biological relevance to HLB pathosystem..... **Error! Bookmark not defined.**

Table 15. HS-GC-MS analysis of selected typical VOCs from sample leaves. ..... **Error! Bookmark not defined.**

Table 16. Consensus putative biomarker for detecting asymptomatic *C*Las-infected samples based on several chemometric approaches. ..... **Error! Bookmark not defined.**

UNIVERSITAS
GADJAH MADA
Detection of Asymptomatic Candidatus Liberibacter asiaticus-Infected Citrus using an Electronic Nose
Anthoni Sulthan Harahap, Prof. Dr. Ir. Siti Subandiyah, M.Agr.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

FIGURE LIST

Page

Figure 1. The *Candidatus* Liberibacter asiaticus (*C*Las) infection cycle encompasses vector acquisition, systemic multiplication, and long-distance dissemination. **(A)** Acquisition: *D. citri* acquires the bacterium by feeding on infected citrus, which are rendered more attractive by elevated emissions of volatile organic compounds such as methyl salicylate and β-caryophyllene. Following a 16- to 18-day latent period, the vector becomes infectious and transmits CLas to healthy hosts. Symptoms typically manifest 6 to 12 months post-inoculation, while the pathogen multiplies within the phloem (Bové, 2006; Canale *et al*., 2017). **(B)** Long-distance spread: *C*Las can also be dispersed over long distances via grafting and the movement of asymptomatic infected plant

UNIVERSITAS
GADJAH MADA
Detection of Asymptomatic Candidatus Liberibacter asiaticus-Infected Citrus using an Electronic Nose
Anthoni Sulthan Harahap, Prof. Dr. Ir. Siti Subandiyah, M.Agr.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

material (Duan *et al*., 2008). This figure is adapted from Cervantes-santos *et al.* (2025).... **Error! Bookmark not defined.**

Figure 2. Schematic illustration of an antibody detection based on antibody or deoxyribonucleic acid (DNA)/ribonucleic acid (RNA). The specific combination of the analyte with immobilized antibody **(A)** or with a DNA/RNA probe **(B)** causes psychochemical changes, such as changes in mass, temperature, optical properties, or electrical potential. These changes translated into measurable signals for detection (Fang & Ramasamy, 2015) ..... **Error! Bookmark not defined.**

Figure 3. The odour sensor uses a quartz resonator (Toko, 2000). ... **Error! Bookmark not defined.**

Figure 4. Schematic illustration of the fabrication of a single-walled carbon nanotube (SWNT)-coated deoxyribonucleic acid (DNA) biosensor (Wang et al., 2019). ..... **Error! Bookmark not defined.**

Figure 4. Schematic diagram of gas sensing device using a single-walled carbon nanotube (SWNT)-single stranded deoxynucleic acid (ssDNA) biosensor (Wang et al., 2019).**Error! Bookmark not defined.**

Figure 5. **(a)** Citrus volatile organic compound (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 data acquisition system (DAQ system), and **5)** data processing. **(b)** Photo of sample preparation and e-nose during citrus leaf test..... **Error! Bookmark not defined.**

Figure 6. The signals generated by ten metal-oxide semiconductor sensors (S1-S10) and the temperature humidity sensor while detecting gas released from healthy citrus leaf sample **(a)** raw signals **(b)** normalized signals. ..... **Error! Bookmark not defined.**

Figure 7. The complete pipeline of the proposed classification framework, illustrating the pre-processing steps (baseline shifting, feature selection, and extraction), the three classifiers compared [XGBoost (extreme gradient boosting), Extra trees (extremely randomised trees), and QDA (quadratic discriminant analysis)], and the K-fold cross-validation procedure used for evaluating

UNIVERSITAS
GADJAH MADA
Detection of Asymptomatic Candidatus Liberibacter asiaticus-Infected Citrus using an Electronic Nose
Anthoni Sulthan Harahap, Prof. Dr. Ir. Siti Subandiyah, M.Agr.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

model performance on the HLB dataset.**Error! Bookmark not defined.**

Figure 8. **A)** Vial bottle and **(B)** its placement on the volume head space gas chromatography mass spectrometry (HS-GC/MS) machine. ..... **Error! Bookmark not defined.**

Figure 10. Radar plots depicting **(a)** the variation in sensor responses according to sample type (leaves vs. veins) and health status (ASY vs. HEA), where leaf and ASY samples produced stronger responses, and **(b)** the variation according to heat treatment, where pre-heating increased signal intensity .**Error! Bookmark not defined.**

Figure 11. Principal component analysis (PCA) plot with variations in sample preparation methods, **(a)** PCA of whole leaf sample without heating with λ = 1.60, **(b)** PCA of vein sample without heating with λ = 1.67, **(c)** PCA of leaf sample with heating with λ = 1.58, **(d)** PCA of vein sample with heating with λ = 1.12.**Error! Bookmark not defined.**

Figure 12. Correlation analysis and principle component analysis (PCA) of several parameters using feature maximum (**a**, **b**, **c**), mean (**d**, **e**, **f**), and area under the curve (AUC) (**g**, **h**, **i**), respectively. .. **Error! Bookmark not defined.**

Figure 12. Operating characteristic area under the curve (ROC-AUC) analysis to determine optimal quadratic discriminant analysis (QDA) model threshold. ..... **Error! Bookmark not defined.**

Figure 13. Principal component analysis (PCA) (**a-d**) and linear discriminant analysis (**LDA**) **(e-h)** plots of each feature's max, mean, area under the curve (AUC), and mix in sequence. PCA overlaps significantly with the highest diversity explained using the mix feature **(d)**. **LDA** with the Mix feature best distinguishes between classes, as seen from the largest gap between classes (shown by the total distance between the blue lines) and the number of overlapping plots on the same LDA score (shown by the smallest total length of the black lines **(h)**. **Error! Bookmark not defined.**

Figure 15. Feature importance rankings for ten sensors (1–10) across three feature extraction (FE) methods (max, mean, and AUC) and

UNIVERSITAS
GADJAH MADA
Detection of Asymptomatic Candidatus Liberibacter asiaticus-Infected Citrus using an Electronic Nose
Anthoni Sulthan Harahap, Prof. Dr. Ir. Siti Subandiyah, M.Agr.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

multiple classifiers. Features 0–9, 10–19, and 20–29 correspond to FE_max, FE_mean, and FE_AUC, respectively. Sensor 7 (appearing as feature 6 in FE_max and feature 16 in FE_mean) consistently ranks as a key discriminatory variable across all models. Its ranking per subpanel is: **(a)** FE_max, 3rd; **(b)** FE_mean, 1st; **(c)** FE_max, 2nd; and **(d)** FE_max, 7th – indicating that while sensor 7 is diagnostically relevant, its relative contribution depends on the FE method and classifier used. ..... **Error! Bookmark not defined.**

Figure 16. Chromatogram of headspace gas chromatography mass spectrometry (HS-GC/MS) analysis results and sensor response of e-nose shows peak differences between healthy and asymptomatic samples (shown by red circles in each class). **(a)** HS-GC/MS chromatogram of the healthy sample **(b)** E-nose sensor response of the healthy sample **(c)** GC/MS chromatogram of the asymptomatic CLas-infected sample **(d)** E-nose sensor response of the asymptomatic CLas-infected sample. ..... **Error! Bookmark not defined.**

Figure 17. The percentage stacked bar chart of selected volatile organic compound (VOC) of healthy and asymptomatic CLas-infected groups..... **Error! Bookmark not defined.**

Figure 18. **(a)** The number of volatile organic compounds (VOCs) in the two types of samples (healthy, asymptomatic *C*Las-infected). **(b)** The unit content of VOCs. **(c)** The percentage of VOCs number **(d)** The percentage of VOCs unit content.**Error! Bookmark not defined.**

Figure 19. Boxplot analysis of the headspace-gas chromatography/mass spectrometry (HS-GC/MS)..... **Error! Bookmark not defined.**

Figure 20. Partial least square-discriminant analysis (PLS-DA) of volatile organic compounds (VOCs) (83 compounds) discriminating ASY from HEA. (A) 3D score plot: ASY (green) vs HEA (red) with 95% confidence ellipsoids. (B) 10-fold cross-validation performance: accuracy, explained variance (R^2), and predictive power (Q^2) across compound numbers. (C) VIP scores highlighting top

UNIVERSITAS
GADJAH MADA
Detection of Asymptomatic Candidatus Liberibacter asiaticus-Infected Citrus using an Electronic Nose
Anthoni Sulthan Harahap, Prof. Dr. Ir. Siti Subandiyah, M.Agr.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

discriminatory compounds, including (2-aziridinylethyl)amine, β-myrcene, and (+)-4-carene. ..... **Error! Bookmark not defined.**

Figure 21. Univariate receiver operating characteristic (ROC) curve analysis of potential putative biomarkers alongside their corresponding box-and-whisker plots. ..... **Error! Bookmark not defined.**

Figure 22. Multivariate receiver operating characteristic (ROC) analysis partial least square -discrimination (PLS-DA) with Monte Carlo cross validation (MCCV) and balanced subsampling. **(A)** Mean ROC curves for 2–83 feature models; AUC and 95% CI inset. **(B)** Selection frequency (%) of significant features. **(C)** Accuracy versus feature number, peaking at 10 (red dot); confusion matrix inset. **(D)** Class probabilities per sample for the AUC-best classifier; decision threshold at 0.5 (dashed line) due to balanced training..... **Error! Bookmark not defined.**

LIST OF ATTACHMENTS

Page

UNIVERSITAS
GADJAH MADA
Detection of Asymptomatic Candidatus Liberibacter asiaticus-Infected Citrus using an Electronic Nose
Anthoni Sulthan Harahap, Prof. Dr. Ir. Siti Subandiyah, M.Agr.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Attachment 1. Photograph of e-nose system. E-nose was connected to a sampling con container to place the tested meat floss sample ..... **Error! Bookmark not defined.**

Attachment 2. MOS sensor response specifications. Minimum and maximum responses of the ten MOS sensors of the e-nose to various target gases. ..... **Error! Bookmark not defined.**

Attachment 3. Output sensing signal from gas sensor in the e-nose. The raw and normalised signals generated by 10 semiconductor sensors and temperature humidity sensors (S1-S12) while detecting gases released from citrus leaf (**a, b**) healthy sample and (**c, d**) asymptomatic HLB sample..... **Error! Bookmark not defined.**

Attachment 4. This structured workflow follows these key steps highlighted through the snapshot images below. The analysis begins by selecting the “Biomarker Analysis” module and uploading a pre-processed CSV data set (see table format) **(A-D)**. A data integrity check is performed to ensure data consistency, with correction made via the Data Editor if necessary **(E)**. The analysis was conducted without normalising the data, as the data was already in percentage form. The transformation of the data was executed through the implementation of the cube root transformation method (cube root of data values), owing to the fact that the distribution exhibited pronounced skewness. The Pareto scaling method was employed for data scaling (mean-centred and divided by the square root of the standard deviation of each variable) so that all compounds contributed relatively equally and were not dominated by large peaks **(F)**. The subsequent step involves performing ROC analysis using classical univariate values **(G)**. To enhance statistical reliability a 95%, confidence band is computed, enabling more precise visualization of ROC plots with box-cum-whisker plots (H) The top-ranked metabolic rations and features are further analysed using by AUROC value (**I)**. The combined putative biomarker of several compounds was also analysed using multivariate ROC

UNIVERSITAS
GADJAH MADA
Detection of Asymptomatic Candidatus Liberibacter asiaticus-Infected Citrus using an Electronic Nose
Anthoni Sulthan Harahap, Prof. Dr. Ir. Siti Subandiyah, M.Agr.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

analysis with the built-in PLS-DA classifier engine. (**J-O**). ..... **Error! Bookmark not defined.**

Attachment 5. The performance of the three models with several features. Accuracy, specificity, and sensitivity values based on the 95% confidence interval (CI 95%) is shown in the form of average values, lowest values, and highest values for each accuracy, specificity, and sensitivity..... **Error! Bookmark not defined.**

Attachment 6. Compound emitted by asymptomatic *C*Las-infected citrus leaves and healthy citrus leaves determined by Head Space-Gas Chromatography/Mass Spectrometry (HS-GC/MS). ..... **Error! Bookmark not defined.**

Attachment 7. Variability boxplot analysis..... **Error! Bookmark not defined.**

Attachment 8. Principal component analysis (PCA) was performed to analyse the profiles of volatile organic compounds (VOCs) obtained from ASY samples (N = 4) and HEA samples (N = 4) using the MetaboAnalyst 6.0 platform. **(A)** As illustrated in the two-dimensional biplot of components 1 and 2, the distribution of samples is depicted by red dots, which represent ASY, and green dots, which represent HEA. The compounds displayed are those demonstrating the highest variability. The pink and green ellipses represent the 95% confidence intervals. **(B)** A two-dimensional loadings plot between components 1 and 2 is presented, illustrating the contribution of all compounds to the principal components. .... **Error! Bookmark not defined.**

Attachment 9. Discriminatory analysis was performed using orthogonal partial least squares-discriminant analysis (OPLS-DA) for the citrus leaf's VOC profiles, which were obtained from the ASY and HEA groups. The analysis encompasses a total of 83 compounds. **(A)** The 2D score plot is employed to differentiate between ASY (red dots) and HEA (green dots). **(B)** The OPLS-DA model was subjected to a three-fold cold validation, resulting in the following R2X, R2Y and Q2 cumulative values: 0.617, 0.9925 and 0.6054, respectively**.** **(C)** The fifteen VOCs with the highest variable importance in projection (VIP) scores. ..... **Error! Bookmark not defined.**

UNIVERSITAS
GADJAH MADA
Detection of Asymptomatic Candidatus Liberibacter asiaticus-Infected Citrus using an Electronic Nose
Anthoni Sulthan Harahap, Prof. Dr. Ir. Siti Subandiyah, M.Agr.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Attachment 10. The two-sample t-test and the Wilcoxon rank-sum test indicates that there are no statistically significant alterations in the concentrations of the 83 compounds belonging to the ASY and HEA groups. ..... **Error! Bookmark not defined.**

Attachment 11. The p-value and false discovery rate (FDR) of the altered metabolites obtained from the two-sample t-test, incorporating the ASY and HEA study groups, are presented. All compounds in the ASY group with the HEA group were not significantly different because they had p-values above the threshold of 0.05; therefore, the MetaboAnalyst 6.0 system did not calculate the q-value (FDR).**Error! Bookmark not defined.**

Attachment 12. Classical ROC Curve Analysis for Individual Biomarker..... **Error! Bookmark not defined.**

Attachment 13. Multivariate ROC curves analysis using Monte Carlo cross-validation (MCCV) with balanced subsampling using RF as classification and feature ranking method. **(A)** ROC curves based on the cross validation (CV) performance. The default is the ROC curves from all models averaged from all CV runs including different metabolic panels (2, 3, 5, 10 and 83). The area under the curve (AUC) along with corresponding 95% confidence of interval has been mentioned in the inset. **(B)** Significant features selected for generating multivariate ROC curves lotted as a function of their selected frequency (%). **(C)** Predictive accuracies with different number of compound features. The red dot shows the highest predictive frequencies with 10 compound features. **(D)** the predicted class probabilities (average of the cross-validation) for each sample using the best classifier (based on AUC). You can manually select other models as listed. As a balanced subsampling approach is use for model training, the classification boundary is always at the centre (x= 0.5, the dotted line). ..... **Error! Bookmark not defined.**

Attachment 14. Multivariate receivier operating curve (ROC) curves analysis using Monter Carlo cross-validation (MCCV) with balanced

UNIVERSITAS
GADJAH MADA
Detection of Asymptomatic Candidatus Liberibacter asiaticus-Infected Citrus using an Electronic Nose
Anthoni Sulthan Harahap, Prof. Dr. Ir. Siti Subandiyah, M.Agr.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

subsampling using linear support vector machine (SVM) as classification and feature ranking method. **(A)** ROC curves based on the cross validation (CV) performance. The default is the ROC curves from all models averaged from all CV runs including different metabolic panels (2, 3, 5, 10 and 83). The area under the curve (AUC) along with corresponding 95% confidence of interval has been mentioned in the inset. **(B)** Significant features selected for generating multivariate ROC curves lotted as a function of their selected frequency (%). **(C)** Predictive accuracies with different number of compound features. The red dot shows the highest predictive frequencies with 10 compound features. **(D)** the predicted class probabilities (average of the cross-validation) for each sample using the best classifier (based on AUC). You can manually select other models as listed. As a balanced subsampling approach is use for model training, the classification boundary is always at the centre (x= 0.5, the dotted line). ..... **Error! Bookmark not defined.**

Attachment 15. Multivariate ROC Feature Importance.**Error! Bookmark not defined.**

Attachment 16. Thirty compilation of putative biomarkers from the 15 best putative biomarkers from all biometric approaches. ..... **Error! Bookmark not defined.**