Utilization of GC-MS in Determination of Electronic Nose Sensor Array for Classification of Gambung Green Tea Quality

Rini Handayani
Dept. of Informatics Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
7025232001@student.its.ac.id;

Riyanarto Sarno
Dept. of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
riyanarto@its.ac.id

Dedy Rahman Wijaya
School of Applied Science
Telkom University
Bandung, Indonesia
dedyrw@telkomuniversity.ac.id

Daniel Yeri Kristiyanto
Dept. of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
7025232003@student.its.ac.id

Kelly Sungkono
Dept. of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
kelly@its.ac.id

2024 2nd International Conference on Technology Innovation and Its Applications (ICTIIA) | 979-8-3503-5161-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICTIIA61827.2024.10761564

***Abstract*—Green tea is recognized for its numerous advantages. Several studies have been conducted employing diverse approaches to categorize the quality of green tea, one of which involves the development of an electronic nose system (e-nose). The selection of the gas sensor array is a critical factor in the construction of an effective e-nose system. Thus, in the present study, the identification of gas sensors array was executed based on the result of gas chromatography-mass spectrometry (GC-MS) on two specimens, categorized as good and quality defects. The sensory outputs from each sensor are gathered into a dataset, and subsequently analyzed for predictive classification. The dataset generated by the deployment of sensors, namely MQ3, TGS822, TGS2602, MQ5, MQ138, and TGS2620, exhibits optimal performance, specifically 0.989 in random forest modeling with regards to metrics such as accuracy, precision, recall, and F1 score. In summary, it is proven that employing a combination of these six sensors and random forest modeling yields a performance above 98%.**

Keywords*—*green tea, GC-MS, e-nose, gas sensors, Random Forest

I. INTRODUCTION (*HEADING 1*)

Consuming green tea is acknowledged not only for its delightful flavor and pleasant aroma but also for its numerous health benefits[1]. One of the most notable factors impacting the character of tea is its aroma. Various classifications of green tea display unique olfactory profiles, wherein premium grade frequently demonstrates an increased level of favorable aroma intensity[2]. Gas chromatography-mass spectrometry (GC-MS) is a frequently occupied method in the determination of specific volatile aroma compounds (VOC)[3]. Nevertheless, when this approach is utilized across all manufacturing processes, it is costly, time-consuming, challenging for green tea cultivators to execute, and unsuitable for immediate quality assessment[4]. Therefore, this method is more suitable to be used as ground truth and then compared with the organoleptic score from the tea tester so that the gas sensor array can be determined more effectively and efficiently.

In previous research, e-noses have also been developed for quality classification of green tea, such as West Lake Longjing using 8 MOS sensors and random forest modeling producing a performance of 99.42% and Chaoqing using eight MOS sensors, made by Figaro Engineering inc, delivering a performance of 95% with SVM analysis[5][6].

Here, the classification model comparing different quality levels for 78 chops of Gambung green tea with six MOS sensors was explored using logistic regression, SVM, and random forest models according to the result of GC-MS analysis.

II. LITERATURE REVIEW

A. *Green Tea Aroma*

Numerous studies have employed various approaches to evaluate green tea's volatile organic compounds (VOCs). There exists a correlation between health benefits and volatile compounds which contribute to cultivating a positive perception; these compounds demonstrate antioxidant activities, anticarcinogenic, and antimicrobial[8].

Usually, to get details of the aroma constituent substances, green tea samples were acquired through headspace solid-phase microextraction (HS-SPME)[9] then analyzed via gas chromatography, subsequently quantifying the mass through mass spectrometry[10]. Researchers find this method highly beneficial in establishing a reliable ground truth, enabling the precise identification of volatile organic compounds.

Various evaluation standards for green tea, such as ISO11287-2011 defining the essential requirements of green tea; (GB/T 23776-2018) outlining the sensory evaluation methodology for tea in the Chinese national guideline; and the Indonesian national regulation (SNI3945:2016) concerning green tea, aroma emerges as a crucial parameter consistently considered in all sensory evaluations of various types of green tea (dry, infused, and infused residues). The assessment involved multiple assessors (tea testers) of different genders and age groups.[5]. International Tea Masters Association (ITMA) classified the typical aromas of green tea into the tea aroma wheel into several characters such as earthy, herbaceous, floral, nutty/milky, sweet, fire/animal, spicy, fresh/candied fruit, marine, and mineral[11].

B. *Electronic Nose*

The electronic nose (E-Nose) is an artificial nose that imitates the nose function of humans. E-nose, enhanced with

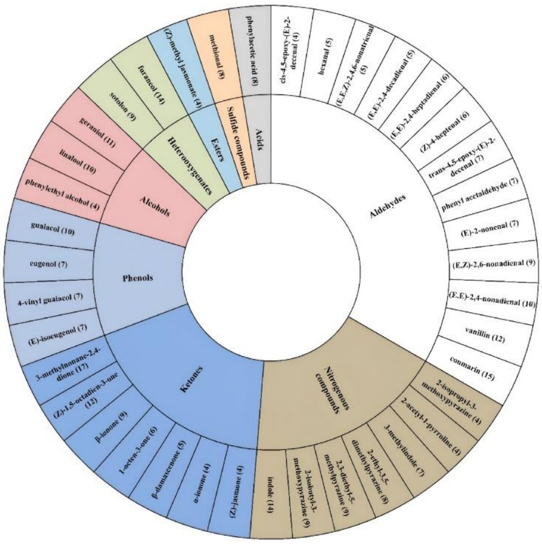

Fig. 1. Example of Key Aroma identified in Chinese Green Tea[10]

artificial intelligence, has the capability to replicate the olfactory perception of humans. This system is made up of three components, illustrated in Fig. 2: (1) a block for handling the aroma of the sample, (2) a block for detection, which consists of gas sensors array, and (3) a block for predicting the sample through data processing and pattern recognition.

E-nose systems have consistently been outfitted with sensors that employ gas sensors that exhibit sensitivity towards aromatic compounds[12]. A recent study highlighted the efficiency of an Electronic Nose (E-nose) machine, which uses Metal Oxide Semiconductor (MOS) sensors, in categorizing intricate scents emitted by volatile substances found in food items such as tea. Subsequently, this tool may be employed for differentiation and assessing alterations in quality[13][14][15].

C. *Predicting Analytics*

Seeing that the data representing the quality of green tea consists of two classes, it was decided that the predictive analysis model used was binary classification models such as logistic regression, support vector machine, decision trees, and random forest.

Logistic regression is a statistical technique commonly used to categorize data observations into distinct classes, typically focusing on a binary outcome variable[16]. While support vector machine (SVM) finds the optimal hyperplane to separate classes in feature. With a kernel, SVM can handle imbalanced data classification[17]. A decision tree splits the dataset into subsets based on input attributes and forms a tree structure for classification. This model is easy to interpret and is able to handle categorical and numerical data[18]. Whereas random forest is an ensemble method that uses multiple decision trees to increase accuracy and reduce overfitting. The final decision is made based on majority voting from all trees[19].

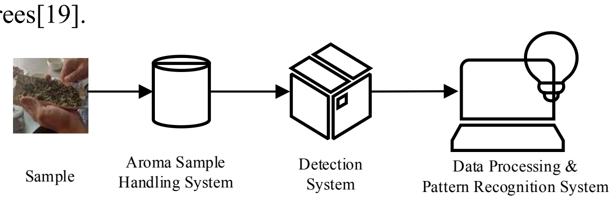

Fig. 2. General diagram of electronic nose system

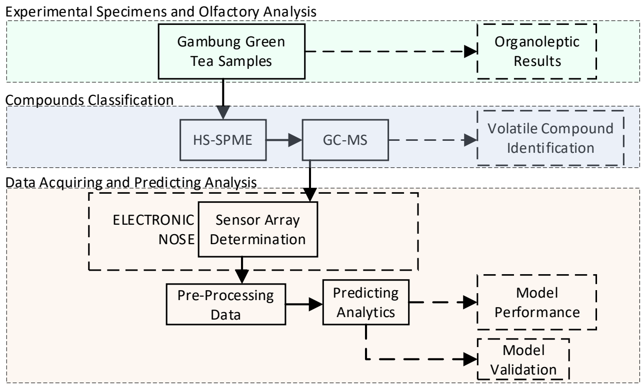

Fig. 3. Research Methods

III. Materials and Methods

In order to construct the electronic nose equipment, an initial assessment of the volatile organic compound (VOC) was conducted. Following identifying the organic compounds, a classification of these compounds was performed, leading to the determination of the suitable gas sensors to be utilized. Through the utilization of the data obtained from the sensor readings, a predictive analysis was conducted to derive the model demonstrating optimal performance as illustrated in Fig. 3.

A. *Experimental Specimens and Olfactory Analysis*

A total of 78 chops, about 50 grams each, of Gambung green tea were subjected to analyse. The cultivation of this variety takes place within a tea plantation located on the slopes of Mount Tilu Ciwidey, West Java situated at an elevation of 1350 meters above sea level. Each chop corresponds to a specific quantity resulting from a single batch processing of green tea.

Each chop was organoleptic tested by tea testers from the Gambung tea research center based on SNI3945:2016, which also refers to ISO11287-2011. Tests were conducted on dried, steeped, and steeped dregs forms of tea. Testing included aroma, taste, and appearance, which were then categorized into good quality tea and quality defects.

The tea testers exhibited scores according to the standards outlined in SNI 3945-2016, with ratings ranging from 2.8 for both appearance and aroma, followed by a threshold of 33 for taste to formulate categories of good quality and quality defects. If focused on aroma, the organoleptic score can only provide a typical or non-typical category of green tea aroma.

B. *Compounds Classification*

Gambung green tea samples are examined for their volatile organic compounds using Headspace-Solid Phase Microextraction (HS-SPME) in Gas Chromatography-Mass Spectrometry (GC-MS) as the gas extraction technique to acquire the ground truth. According to the organoleptic score tested by the tea testers, the samples reviewed consist of one sample with the highest and one with the lowest, as categorized as organoleptic.

The extraction process of these samples was heated in a water bath to a temperature of 50 degrees Celsius for 30 minutes then absorbed by fiber made of Divinylbenzene /Carboxen/Polydimethylsiloxane (DVB/CAR/PDMS) up to 2cm then GC-MS analysis was carried out for 60 minutes.

TABLE I. ORGANOLEPTIC SCORE FOR AROMA

SNI 3945-2016	Gambung Tea Research Center Representation
A = normal typical green tea; B = normal typical green tea; C = normal typical green tea; D = less typical green tea; E = abnormal.	1 ← Quality Defect 2.8 Good Quality → 5

The volatile compounds were determined through a comparison of the acquired mass spectra with established standards from the NIST Mass Spectral library and by the computation of linear retention indices (LRI) for individual peaks concerning the standard normal alkanes.

C. *Data Acquiring and Predicting Analysis*

The utilization of predictive analysis for categorizing discrete data generated by the six sensors within the electronic nose (e-nose) system can be accomplished through different classification modeling techniques such as logistic regression, decision tree, random forest, and support vector machine.

Examining the performance of these models considered a range of points of view, like precision, accuracy, recall, and F1 score. Subsequently, the model exhibiting the best performance was reviewed extensively through the analysis of a confusion matrix.

This study utilized Python for analysis and all programming tasks were conducted within the Scikit-learn framework. In scikit-learn (sklearn), the dataset is divided into training data and testing data utilizing train_test_split to facilitate classifier training and performance evaluation[20].

IV. RESULT AND ANALYSIS

The process of obtaining data starts from an organoleptic testing approach, and then representatives of the two categories are studied using GCMS to define VOCs so that the most relevant sensors can be determined.

A. *Organoleptic Score*

The organoleptic tests carried out in this test include dried appearance, steeped color, taste, aroma, and steeped dregs. Out of the 78 chops that were accessible, 51 chops exhibited good quality, whereas 27 chops had quality defects.

B. *Volatile Organic Compounds Identification*

Through Chromatography-Mass Spectrometry (GC-MS), it was found that there were 143 VOCs in the good-quality sample and 118 VOCs in the quality-defect sample. Then it also obtained that there were 102 similar compounds present

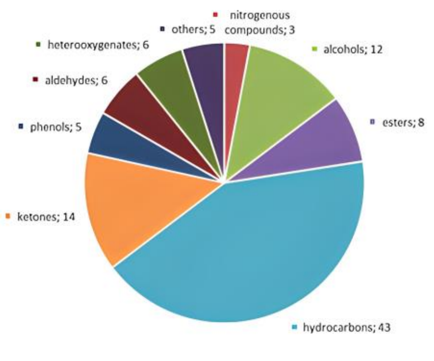

Fig. 4. Number of Compounds in Each Classes

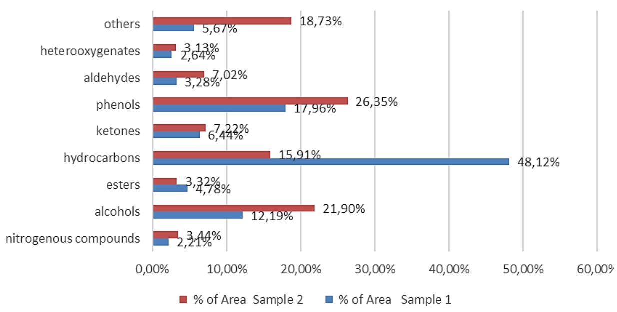

Fig. 5. Percentage of Area of Compounds in Each Classes

in both sample 1 and sample 2, which are detailed in TABLE II.

The results of grouping chemical substances into several groups, including nitrogenous compounds, alcohols, esters, hydrocarbons, ketones, phenols, aldehydes, hetero-oxygenates, and others. Substance content in each class of compounds means that different concentrations are shown as Area as shown in TABLE II.

In GC-MS, in order to obtain the percentage peak area, the approach involves utilizing the target component peak area in relation to the total area encompassing all identified peaks for quantitative analysis. This technique is commonly employed for assessing variations in the concentration levels of a specified sample mixture, or for approximating the concentration of a given sample mixture.

In sample 1, hydrocarbons as much as 48.12% of the total emerge as the predominant compound class among all others, followed by phenols, alcohols, ketones, esters, aldehydes, and nitrogenous compounds. On the other side in sample 2, phenols have a portion almost twice as large as hydrocarbons. Significant differences in class of compound between samples are illustrated on the graphic in Fig. 5.

There are several similarities in the aroma compounds contained in Gambung green tea and Long Jin green tea. These include benzaldehyde, (E, E)-2,4-heptadienal, benzeneacetaldehyde, (E, E)-3,5-octadien-2-one, linalool, nonanal, methyl salicylate, geraniol, and β-ionone[21]. Meanwhile, Hangzhou Green tea and Ya'an green tea samples exhibit a diverse array of volatile compounds, such as 27 hydrocarbons, 2 furans, 9 alcohols, 11 ketones, 7 esters, 17 aldehydes, and 9 nitrogen compounds[22].

On the contrary, Gambung green tea is distinguished by many higher concentrations of volatile compounds, comprising 43 hydrocarbons, 12 alcohols, 14 ketones, 6 aldehydes, and 3 nitrogen compounds. This makes it very possible that Gambung green tea has its characteristics compared to green tea from other countries. One example of its uniqueness is the presence of the compound (-)-Norephedrine, which is usually used for the treatment of nasal congestion substance, in both samples of Gambung green tea.

C. *Sensor Array Determination for Electronic Nose*

The sensor array is an important and complex component for detecting gases or odors in an electronic nose, consisting of multiple sensors that exhibit extensive response typical and significant cross-sensitivity to various gases. This sensor array is occupied to convert gas concentration into voltage level, which is then used for quantitative or qualitative analysis of

TABLE II. COMPARISON OF 102 COMPOUNDS IN EACH SAMPLE

Class of Compounds	Number of Compounds	Sum of Area		% of Area		difference of sample 1 and sample 2	% difference between sample 1 and sample 2
		Sample 1	*Sample 2*	*Sample 1*	*Sample 2*		
nitrogenous compounds	3	1.270.119.694	513.330.846	2,21%	3,44%	756.788.848	59,58%
alcohols	12	7.019.137.150	3.271.937.004	12,19%	21,90%	3.747.200.146	53,39%
esters	8	2.751.264.230	496.023.535	4,78%	3,32%	2.255.240.695	81,97%
hydrocarbons	43	27.706.350.041	2.377.299.475	48,12%	15,91%	25.329.050.566	91,42%
ketones	14	3.709.765.794	1.077.871.081	6,44%	7,22%	2.631.894.713	70,95%
phenols	5	10.340.746.772	3.936.394.869	17,96%	26,35%	6.404.351.903	61,93%
aldehydes	6	1.517.805.615	467.773.440	3,28%	7,02%	838.929.945	44,45%
hetero-oxygenates	6	3.261.843.902	2.797.134.495	2,64%	3,13%	1.050.032.175	69,18%
others	5	1.887.336.128	1.048.406.183	5,67%	18,73%	464.709.407	14,25%

TABLE III. COMPARISON SENSOR ARRAY SENSITIVITY TO CLASS OF COMPOUND

Class of Compounds	Sensor Array					
	MQ3	*MQ5*	*MQ 138*	*TGS 822*	*TGS 2602*	*TGS 2620*
nitrogenous compounds					v	
alcohols	v	v	v	v		v
esters						
hydrocarbons		v	v	v		v
ketones			v	v		
phenols						
aldehydes			v			
hetero-oxygenates						
others						

the observed samples. To improve the electronic nose system monitoring of selected samples, the selection of proper gas sensors is one of the main design requirements [3][4]. The number of compounds contained in each class is a reference in determining which gas sensors to use as shown in .

It is observable, as shown on Fig. 4, that the dominant compounds in both samples of Gambung green tea are hydrocarbons, alcohols, ketones, nitrogenous compounds, and aldehydes. So the following sensors were selected, including MQ3, TGS822, TGS2602, MQ5, MQ138, and TGS2620.

Despite the existence of a hydroxyl group, phenol is not considered a direct derivative of alcohol. This is due to significant differences in their chemical structure and properties. Alcohols exhibit -OH groups linked to aliphatic carbons, whereas phenols display -OH groups attached to aromatic rings[23]. It was rather difficult to determine which gas sensor to use for esters and phenols, so this research focused on optimizing existing sensors.

The sampling process carried out on this e-nose device uses 15 grams per dry sample from Gambung green tea in the sample container. The air in the sample container then flows to the sensing area to detect the gas levels that contribute to the aroma of green tea by the gas sensor array. The process of examining the gas is conducted every second,

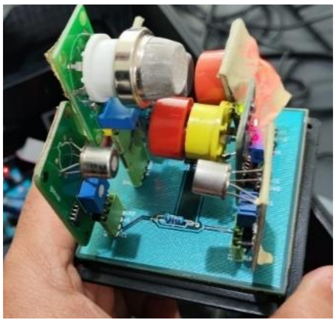

Fig. 6. Gas sensors used in this E-Nose

lasting for one minute producing 9180 data for good quality and 4860 data quality defects.

Data is acquired from the sensor array within the electronic nose device. This data manifested as digital signals, necessitating extraction for statistical parameter derivation. Subsequently, the dataset was inputted into various classification algorithms for predictive outcomes. Assessment of the classification outcomes ensued to ascertain the efficacy and performance of the employed methodologies and parameters[24].

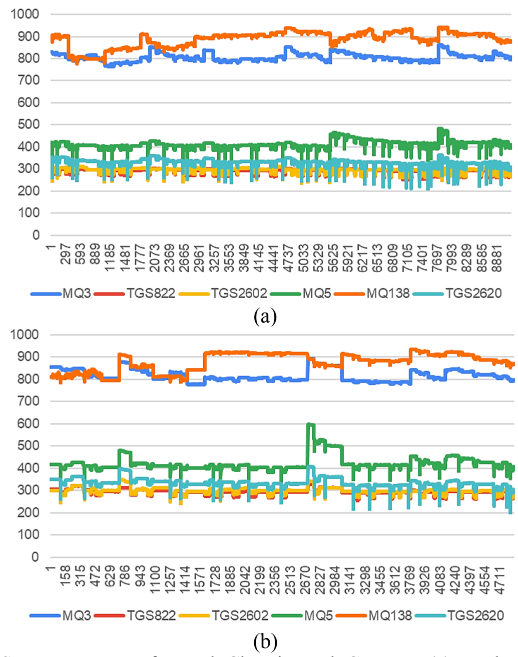

Fig. 7. Sensor Response for Each Chop in Each Category (a) good quality; (b) quality defect

TABLE IV. COMPARISON PERFORMANCE OF EACH MODEL

Model	Performance			
	Accuracy	*Precision*	*Recall*	*F1 Score*
Random Forest	0.989	0.989	0.989	0.989
Decision Tree	0.979	0.979	0.979	0.979
Logistic Registrion	0.700	0.688	0.700	0.663
SVM	0.695	0.779	0.695	0.607

D. Performance Comparison of Predictive Analysis

The prediction analysis used in this research compares the performance of three models, including random forest, logistic regression, decision tree, and SVM, to handle the classification of the quality of Gambung green tea. The experiment used 2 classes by comparing two types of quality, namely good quality and quality defects of Gambung green tea.

Before going to the classification stage, the dataset was split into two separate parts using the cross-validation method. These subsets constituted 80% of the total designed as training data, with testing data making up the remaining 20%. The examination of the data was executed utilizing a cross-sectional technique with a k-fold parameter set at 10[25]. In this study, the training set has 11,232 samples, and the testing set has 2,808 samples, each with 7 features.

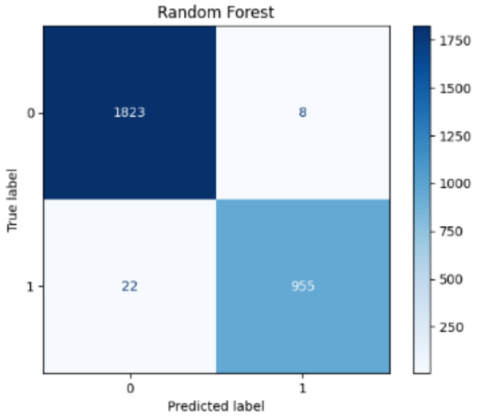

Fig. 8. Confusion Matrix for Random Forest

The dataset comprising sensor responses related to quality is deemed imbalanced due to the disproportion between samples categorized as good quality and those labeled as quality defects. Consequently, the model performance evaluation was reviewed in accuracy, precision, recall, and the F1 score. It can be seen in TABLE IV that Random Forest has the best accuracy, precision, recall, and F1 Score of the three other models. Meanwhile, if explored more deeply through the confusion matrix as shown in Fig. 8, Random Forest is better for predicting which ones are true and false. It can predict 2778 events correctly according to 2808 actual conditions.

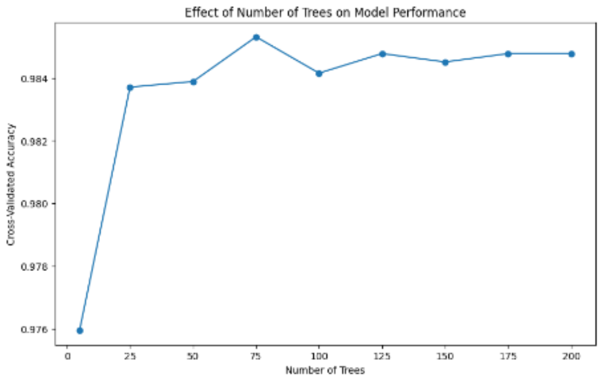

Fig. 9. Number of Trees on Performance of Random Forest

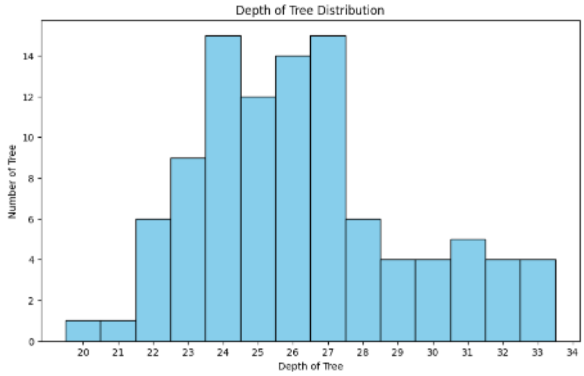

Fig. 10. Depth of Trees Distribution on Each Number of Tree

Random forest involves the utilization of numerous decision trees. The depth of tree in the Random Forest model refers to the number of layers or levels from the root of the tree to the deepest leaf. Tree depth is a measure of how many divisions or branches occur from the root of the tree to the leaves, which represent the final decision or classification[26].

Analysis of the number of trees, ranging from 5 to 200, that is most likely for random forest in this study was also carried out as depicted in the curve in Fig. 9. From this, the best number of trees and the most likely to get the highest performance in this study is 75 trees with the depth of tree distribution shown in Fig. 10.

V. CONCLUSION

The main objective of this research is to determine which prediction model has the best performance for the electronic nose system in detecting the quality of Gambung green tea. Before obtaining this model, researchers analyzed the volatile organic compounds. The aim is to specify the gas sensors used so that the predictive analysis modeling does not require too complex computations.

The results of GCMS analysis of two samples of Gambung green tea showed that the group of aromatic hydrocarbon compounds was the most dominant in quantity. Meanwhile, in terms of the content portion of the class of compounds, in sample 1 the most dominant ones were hydrocarbons around 48% and phenols around 26% in sample 2.

The limited availability of gas sensors to detect esters, phenols, and hetero-oxides makes this system focus on what was available. Sequentially, the sensors used were dominated by those capable of detecting hydrocarbons, alcohols, ketones, aldehydes, and nitrogenous compounds. Those are MQ3, TGS822, TGS2602, MQ5, MQ138, and TGS2620 which produces a dataset of 11,232 samples as the training set and 2,808 as the testing set.

The prediction analysis carried out on this system used a classification model consisting of logistic regression, SVM, decision tree, and random forest. Among the various models considered, the random forest demonstrates superior performance in terms of accuracy, precision, recall, and F1 score of approximately 0.989. This is possible because random forests belong to the ensemble learning categories which builds many decision trees during the training phase and combines the results to increase accuracy and control overfitting. From the experiment, it can be seen that a larger number of trees does not always result in better performance because this is very likely for overfitting to occur.

Compared to the previous e-nose, this system, even though it used 6 metal oxide semiconductor sensors, can produce a performance above 0.98. Despite that, it is not certain that this system will be able to detect the antioxidant catechin contained in green tea that is also belong to the group of phenolic compounds. So that further investigations can be pursued in the development of an electronic tongue, for example.

ACKNOWLEDGMENT

All of green tea sample was provided by Research Institute of Tea and Cinchona, Gambung, West Java. This research was funded by Institut Teknologi Sepuluh Nopember (ITS) under Penelitian Dana Departemen Program, Penelitian Keilmuan, Penelitian Flagship, Penelitian Kolaborasi Pusat and under the Project Scheme of the Publication Writing and Intellectual Property Rights (IPR) Incentive (Penulisan Publikasi dan Hak Kekayaan Intelektual/PPHKI) Program 2024, and Telkom University under Research and Community Service Directorate.

REFERENCES

[1] L. Zhang, Q.-Q. Cao, D. Granato, Y.-Q. Xu, and C.-T. Ho, ‘Association between chemistry and taste of tea: A review’, Trends Food Sci. Technol., vol. 101, pp. 139–149, Jul. 2020, doi: 10.1016/j.tifs.2020.05.015.

[2] H. R. Tan et al., ‘Characterisation of key odourants in Japanese green tea using gas chromatography-olfactometry and gas chromatography-mass spectrometry’, LWT, vol. 108, pp. 221–232, Jul. 2019, doi: 10.1016/j.lwt.2019.03.054.

[3] Q. Peng, S. Li, H. Zheng, K. Meng, X. jiang, R. Shen, J. Xue, and G. Xie, ‘Characterization of different grades of Jiuqu hongmei tea based on flavor profiles using HS-SPME-GC-MS combined with E-nose and E-tongue’, Food Res. Int., vol. 172, p. 113198, 2023, doi: 10.1016/j.foodres.2023.113198.

[4] S. N. Hidayat, K. Triyana, I. Fauzan, T. Julian, D. Lelono, Y. Yusuf, N, Ngadiman, A.C.A. Veloso, and A.M. Peres, ‘The Electronic Nose Coupled with Chemometric Tools for Discriminating the Quality of Black Tea Samples In Situ’, Chemosensors, vol. 7, no. 3, p. 29, Jul. 2019, doi: 10.3390/chemosensors7030029.

[5] X. Lu, J. Wang, G. Lu, B. Lin, M. Chang, and W. He, ‘Quality level identification of West Lake Longjing green tea using electronic nose’, Sens. Actuators B Chem., vol. 301, p. 127056, 2019, doi: 10.1016/j.snb.2019.127056.

[6] D. Huo, Y. Wu, M. Yang, H. Fa, X. Luo, and C. Hou, ‘Discrimination of Chinese green tea according to varieties and grade levels using artificial nose and tongue based on colorimetric sensor arrays’, Food Chem., vol. 145, pp. 639–645, 2014, doi: 10.1016/j.foodchem.2013.07.142.

[7] H. Yu, J. Wang, H. Zhang, Y. Yu, and C. Yao, ‘Identification of green tea grade using different feature of response signal from E-nose sensors’, Sens. Actuators B Chem., vol. 128, no. 2, pp. 455–461, 2008, doi: 10.1016/j.snb.2007.07.048.

[8] Y. Guo, X. Yang, H. Wang, X. Yang, X. Su, J. Kong, Y. Cheng, W. Yao, and H. Qian, ‘Study on the establishment of quality discrimination model of Longjing 43 green tea (Camellia sinensis(L.) Kuntze)’, J. Appl. Res. Med. Aromat. Plants, vol. 31, p. 100389, 2022, doi: 10.1016/j.jarmap.2022.100389.

[9] S. Kaushal, P. Nayi, D. Rahadian, and H.-H. Chen, ‘Applications of Electronic Nose Coupled with Statistical and Intelligent Pattern Recognition Techniques for Monitoring Tea Quality: A Review’, Agriculture, vol. 12, no. 9, p. 1359, Sep. 2022, doi: 10.3390/agriculture12091359.

[10] P. Yin, Y.-S. Kong, P.-P. Liu, J.-J. Wang, Y. Zhu, G.-M. Wang, M.-F. Sun, Y. Chen, G.-Y. Guo, and Z.-H. Liu, ‘A critical review of key odorants in green tea: Identification and biochemical formation pathway’, Trends Food Sci. Technol., vol. 129, pp. 221–232, Nov. 2022, doi: 10.1016/j.tifs.2022.09.013.

[11] ITMA Tea Aroma Wheel. [Online]. Available: https://teamasters.org/itma-tea-aroma-wheel/

[12] M.A. Hasan, R. Sarno, and S.I. Sabilla, ‘Optimizing Machine Learning Parameters for Classifying the Sweetness of Pineapple Aroma Using Electronic Nose’, Int. J. Intell. Eng. Syst., vol. 13, no. 5, pp. 122–132, Oct. 2020, doi: 10.22266/ijies2020.1031.12.

[13] S. M. T. Gharibzahedi, F. J. Barba, J. Zhou, M. Wang, and Z. Altintas, ‘Electronic Sensor Technologies in Monitoring Quality of Tea: A Review’, Biosensors, vol. 12, no. 5, p. 356, May 2022, doi: 10.3390/bios12050356.

[14] S. Ahmed and S. K. Sinha, ‘Studies on nanomaterial-based p-type semiconductor gas sensors’, Environ. Sci. Pollut. Res., vol. 30, no. 10, pp. 24975–24986, Jun. 2022, doi: 10.1007/s11356-022-21218-6.

[15] A. Khorramifar, H. Karami, L. Lvova, A. Kolouri, E. Lazuka, M.P. - Rozek, G. Lagod, J. Ramos, J. Lozano, M. Kaveh, and Y. Darvishi, ‘Environmental Engineering Applications of Electronic Nose Systems Based on MOX Gas Sensors’, Sensors, vol. 23, no. 12, p. 5716, Jun. 2023, doi: 10.3390/s23125716.

[16] C. Zhu, C. U. Idemudia, and W. Feng, ‘Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques’, Inform. Med. Unlocked, vol. 17, p. 100179, 2019, doi: 10.1016/j.imu.2019.100179.

[17] H. Shamsudin, U. K. Yusof, Y. Haijie, and I. S. Isa, ‘AN OPTIMIZED SUPPORT VECTOR MACHINE WITH GENETIC ALGORITHM FOR IMBALANCED DATA CLASSIFICATION’, J. Teknol., vol. 85, no. 4, pp. 67–74, Jun. 2023, doi: 10.11113/jurnalteknologi.v85.19695.

[18] L. I. Santos, M.O. Camargos, M.F.S.V. D'Angelo, J.B. Mendes, E.E.C.d. Medeiros. A.L.S. Guimaraes, R.M. Palhares ‘Decision tree and artificial immune systems for stroke prediction in imbalanced data’, Expert Syst. Appl., vol. 191, p. 116221, Apr. 2022, doi: 10.1016/j.eswa.2021.116221.

[19] H. Hairani, A. Anggrawan, and D. Priyanto, ‘Improvement Performance of the Random Forest Method on Unbalanced Diabetes Data Classification Using Smote-Tomek Link’, JOIV Int. J. Inform. Vis., vol. 7, no. 1, p. 258, Feb. 2023, doi: 10.30630/joiv.7.1.1069.

[20] S. I. Sabilla, Malikhah, and R. Sarno, ‘Classification and Gas Concentration Measurements of Human Axillary Odor using Electronic Nose’, in 2021 13th International Conference on Information & Communication Technology and System (ICTS), Surabaya, Indonesia: IEEE, Oct. 2021, pp. 161–166. doi: 10.1109/ICTS52701.2021.9608597.

[21] M.-Q. Wang, W.-J. Ma, J. Shi, Y. Zhu, Z. Lin, and H.-P. Lv, ‘Characterization of the key aroma compounds in Longjing tea using stir bar sorptive extraction (SBSE) combined with gas chromatography-mass spectrometry (GC–MS), gas chromatography-olfactometry (GC-O), odor activity value (OAV), and aroma recombination’, Food Res. Int., vol. 130, p. 108908, Apr. 2020, doi: 10.1016/j.foodres.2019.108908.

[22] Y.-Q. Yang, H.-X. Yin, H.-B. Yuan, Y.-W. Jiang, C.-W. Dong, and Y.-L. Deng, ‘Characterization of the volatile components in green tea by IRAE-HS-SPME/GC-MS combined with multivariate analysis’, PLOS ONE, vol. 13, no. 3, p. e0193393, Mar. 2018, doi: 10.1371/journal.pone.0193393.

[23] L. Kiss, D. Bősz, F. Kovács, H. Li, and S. Kunsági-Máté, ‘Electrooxidation of phenol in alcohols and establishment of the permeability of the electrodeposited films’, Polym. Bull., vol. 76, no. 1, pp. 215–226, Jan. 2019, doi: 10.1007/s00289-018-2372-4.

[24] S. Wakhid, R. Sarno, S. Sabilla, and D. Maghfira, ‘Detection and Classification of Indonesian Civet and Non-Civet Coffee Based on Statistical Analysis Comparison Using E-Nose’, Int. J. Intell. Eng. Syst., vol. 13, no. 4, pp. 56–65, Aug. 2020, doi: 10.22266/ijies2020.0831.06.

[25] G. Jiang and W. Wang, ‘Error estimation based on variance analysis of k -fold cross-validation’, Pattern Recognit., vol. 69, pp. 94–106, Sep. 2017, doi: 10.1016/j.patcog.2017.03.025.

[26] A. Nadi and H. Moradi, ‘Increasing the views and reducing the depth in random forest’, Expert Syst. Appl., vol. 138, p. 112801, Dec. 2019, doi: 10.1016/j.eswa.2019.07.018.