Classification and Gas Concentration Measurements of Human Axillary Odor using Electronic Nose

Shoffi Izza Sabilla
Informatics Department, Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember (ITS)
Surabaya, Indonesia
shoffi12@mhs.if.its.ac.id

Malikhah
Informatics Department, Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember (ITS)
Surabaya, Indonesia
meli2511@gmail.com

Riyanarto Sarno
Informatics Department, Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember (ITS)
Surabaya, Indonesia
riyanarto@if.its.ac.id

***Abstract*— Human axillary odor produces gas from sweat which concentration will change depends on the activities and metabolism in the body. Sweat concentration can be used as information to determine body health. Nowadays, e-nose is widely used in medicine, food industry, agriculture, and biotechnology. An electronic nose (e-nose) is a device that mimics how the human nose works. This paper will build an e-nose with seven sensors from Figaro Taguchi series (TGS) sensors and one sensor from humidity and temperature sensors (SHT-15 series). The e-nose was used to obtain the human axillary odor in the morning, afternoon, and evening. Several classifiers are used in the classification process and the result showed that Random Forest with tuned hyperparameter produced the best result with an accuracy of 87.43%. By using the ANOVA f-test, it is showed that methane and ethanol from sensor TGS 2612 are the most significant gas in the classification process. The experimental result showed that human axillary odor produced different ethanol and methane gas concentration in the morning, afternoon, and evening.**

Keywords—ANOVA, axillary odor, electronic nose, ethanol, methane, random forest.

I. INTRODUCTION

An electronic nose (e-nose) is a device that mimics how the human nose works. Nowadays, e-nose is widely used in various fields. Examples of e-nose applications in various fields include: (1) healthcare, detects respiratory diseases [1], detects diabetes [2], detects bacteria in urine [3]; (2) food industry [4]; (3) environmental monitoring [5]; (4) agriculture, detect the ripeness of tomatoes [6], detect methane in Chile [7], soil sensing [8].

E-nose is also used to identify and analyze human odor, for example: human body odor evaluation for personalized beauty [9], identification authentication based on human odor [10], health status detection based on human odor [11], identification of people based on underarm odor for people identification [12]. This paper tried to classify human axillary odor into three classes based on the time of collection, namely morning, afternoon, and evening obtained from the e-nose system.

Humans produce various kinds of gases through sweat, breath, skin, and urine [13]. Human axillary odors produce various kinds of gases whose numbers will change from time to time, depending on the activity and metabolism of the body [14]. Generally, in the morning, humans only do a little activity; therefore, the concentration of gas that produced by axillary odor is still small, then during the afternoon the activity increases, therefore the concentration of gas produced by axillary odor is increasing. Meanwhile, in the evening, the activity will decrease, therefore the increase in the concentration of gas produced by axillary odor will not be as much as during the afternoon. Normally, people would distinguish human axillary odors based on the intensity of the odor produced, but this method has no basis. By knowing the gas produced by axillary odor we can check the health status of the people. Therefore, the contributions of this research are (i) a model to distinguish or classify human axillary odor based on time of collection, namely morning, afternoon, and evening (ii) find the most significant gas in the classification process, and (iii) estimate the gas concentration produced by axillary odor in the morning, afternoon, and evening from the most significant gas. The remainder of this paper is organized as follows. Section 2 presents the explanation of the experimental methods. Section 3 is described the experimental results and evaluation. Finally, conclusions are generalized in Section 4.

II. EXPERIMENTAL METHODS

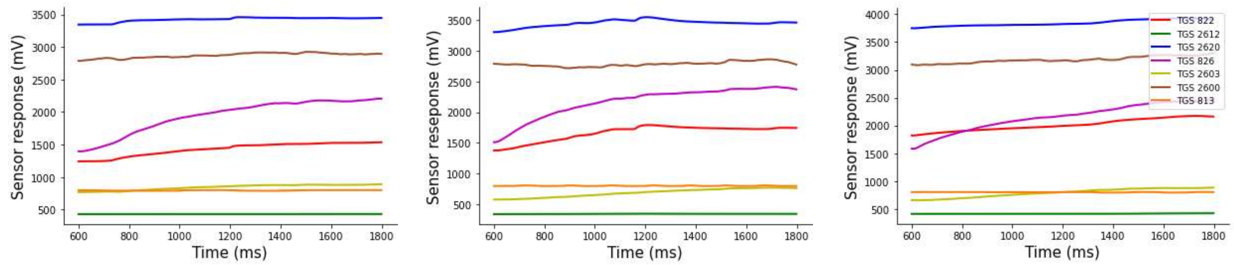
A. Material

The e-nose system was built with an Arduino microcontroller and consisted of eight sensors, seven sensors from Figaro Taguchi series (TGS) sensors, which are TGS 2603, TGS 2612, TGS 2620, TGS 2600, TGS 822, TGS 826, TGS 813, and one sensor from humidity and temperature sensors [15]. In Table 1, we can see that each sensor has an ability to detect various gas. The e-nose system must be

TABLE I. GAS DETECTED BY EACH SENSOR

Sensor	Gas detected
TGS 2603	Ethanol, trimethyl amine
TGS 2612	Methane, ethanol
TGS 2620	Ethanol, isobutane, hydrogen, carbondioxide (CO), methane
TGS 2600	Isobutane, methane
TGS 813	Methane, ethanol, propane, hydrogen
TGS 822	Isobutane, n-hexane, carbondioxide, methane
TGS 826	Ammonia, isobutane, ethanol
SHT 15	Humidity and temperature

978-1-6654-4059-2/21/$31.00 ©2021 IEEE



(a) Sensor response of axillary odor in the morning
(b) Sensor response of axillary odor in the afternoon
(c) Sensor response of axillary odor in the evening

Fig. 1 Axillary odor signal response by using e-nose system (a) morning (b) afternoon (c) evening

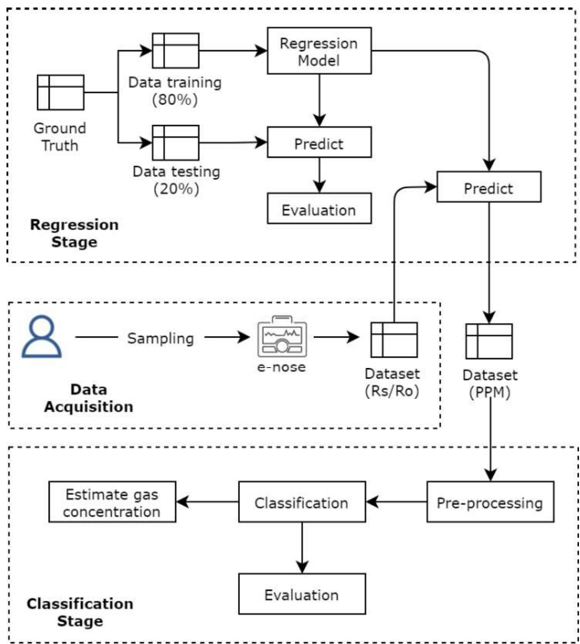


Fig. 2. The scheme of the proposed method

connected to the computer in order to operate the e-nose system and store raw signals obtained from the e-nose system.

This paper used origin Lab to obtain the ground truth of gas detected by each sensor datasheet provided by the Figaro manufacturer and Python 3.8.3 to process the raw signals, classification, regression, and evaluation of the proposed method. Python simulation and the operation of e-nose system were run on a Core i5 9th Gen with 8 RAM, 2.40 GHz frequency, and Windows 10 Home 64-bit edition.

B. Data acquisition

In the recording or sampling process of each sample, the raw signal of the e-nose system was collected using a hose put in the human axillary in the morning, afternoon, and evening. There were three stages in the sampling process, namely flushing, sensing, and purging. At the flushing stage, the e-nose system takes the baseline of surrounding air used as reference air. Meanwhile, in the sensing stage, the sensors in e-nose system absorbs the axillary odor. In the last stage of the sampling process is purging, the e-nose system clean the sensors. A sample was obtained from axillary odor using the e-nose system in 60 seconds of flushing, 120 seconds of sensing, and 100 seconds of purging. Therefore, to take one sample required 280 seconds. The samples were stored in computer every 0.1 seconds or 1 millisecond in Comma Separated Values (CSV) files, thus each sample produced 2800 rows of raw signals.

Axillary odor sampling was carried out for a month, in the morning, afternoon and evening. For class evening axillary odor, there were 48 samples, class morning 58 samples, and class afternoon 59 samples, therefore the total samples obtained from e-nose system were 165. Each sensor's unit value is still expressed in source resistance/output resistance (Rs/Ro) [16]. The results of data acquisition can be seen in Fig. 1, where Fig. 1 (a) is the raw signals of axillary odor obtained in the morning, (b) is the raw signals of axillary odor obtained in the afternoon, while (c) is the raw signals of axillary odor obtained at evening. There is a slight difference in the raw signals obtained by the axillary odor in the morning, afternoon, and evening, as can be seen in Fig. 1.

C. Proposed method

There are two stages in this research, which are regression stage and classification stage. In each phase, there are several steps, as shown in Fig. 2.

1) Regression stage

Each sensor has the ability to detect various gases and has its own graph produced by the manufacturer [17]. In the graph of datasheet, the X-axis shows the parts per million (PPM) value, and the Y-axis shows the ratio of source resistance/output resistance (Rs/Ro). Rs/Ro has an important role in determining the PPM value of each gas based on the datasheet. We can get the ground truth for the Rs/Ro value and the PPM value for each gas from the datasheet. The calibration value of each sensor is the sensor resistance value in clean air (Ro). To calculate the sensor resistance (Rs) can be searched with the ADC value using Equation (1) and Equation (2).

$$Rs = \frac{Vc - VRL}{VRL} \times RL \tag{1}$$

$$VRL = \frac{ADC \times Vc}{1023} \tag{2}$$

Vc is the voltage on the microcontroller, VRL is the sensor voltage in the sample room, RL is the resistance sensor load measured using Ω meter, and ADC is analog to digital value conversion. The ground truth data are obtained from 7 sensors with 22 gasses.

The ground truth obtained will be divided into 80% training data and 20% testing data. The training data will be used to create a regression model and the testing data will be used to evaluate the model. This paper will use K Nearest Neighbor Regression (KNNR) because it is simple, flexible, and can adapt to irregular feature space. KNNR stores all data and predicts targets based on similarity measurements. This paper will use uniform metrics to calculate the similarity measure. In the KNNR, there is also k parameter, which is the number of closest neighbors. Equation (3) is used to make predictions.

$$y' = \frac{1}{K}\sum_{j=1}^{K} y_j \quad (3)$$

where y' is the predicted value, k is the number of nearest neighbors, y_i is the original value.

After the KNNR model is built then we will predict the data testing and evaluate the performance of the model. Two metrics will used to evaluate the performance of the KNNR model, which are Mean Squared Error (MSE) and R Square (R^2) [18]. MSE is the average of the squared difference between the original and the predicted values. Equation (4) is used to calculate MSE.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - y_i')^2 \quad (4)$$

where N is the amount of data, y_i is the original value, and y' is the predicted value. The smaller the MSE value, the better the performance of a regression model. R^2 is used to measure the goodness of fit of the regression model. The value is between 0 and 1, where the closer to 1, the better the regression model. Equation (5) is used to calculate R^2

$$R^2 = 1 - \frac{\sum(y_i - y_i')^2}{\sum(y_i - \bar{y})^2} \quad (5)$$

where y_i, $\bar{y}$, y_i' are the original value, the average value of the original y, the predicted value, respectively. Dataset obtained in the data acquisition which in the form of Rs/Ro will be converted or predicted to PPM value of each gas using the built regression model and we save the prediction result into CSV file. This PPM dataset will be the input for the classification stage.

2) Classification stage

a) Pre-processing

After conversion of Rs/Ro dataset into PPM dataset, pre-processing stage will be carried out. The pre-processing stage began by selecting row obtained from the sensing stage only. After that, statistical feature extraction is performed to summarize each CSV data file into one row using four statistical parameters: average, standard deviation, minimum, and maximum [4]. Average is the average PPM value in each sample, while standard deviation shows the variance distribution of the PPM value in each sample. Maximum is the maximum value of the PPM value in each sample and minimum is the minimum value of the PPM in each sample. Equation (6) to Equation (9) are used to calculate the four statistical parameters above.

$$\sigma = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(r_t - \mu)^2} \quad (6)$$

$$\mu = \frac{1}{T}\sum_{t=1}^{T} r_t \quad (7)$$

$$Maximum = \max(r_t) \quad (8)$$

$$Minimum = \min(r_t) \quad (9)$$

where r_t is the PPM values of each sample, μ is the average PPM of each sample, T is the number of row in each sample, and σ is the standard deviation of each sample.

b) Classification

This paper will compare several classifiers, namely KNN, Support Vector Machine (SVM), and Random Forest (RF). Each classifier will be used to classify the axillary odor sample into three classes, morning, afternoon, and evening, which are the time collection of samples. Each classifier will be evaluated based on accuracy (*Acc*) metrics. In scikit-learn (sklearn), the dataset divided into data training and data testing by using train_test_split to train the classifier and test them. However, train_test_split has a parameter named random_state, which is the accuracy value will change when we change the value of the random_state. Therefore, to compare the accuracy of the classifiers, this paper used stratified k-fold with a number of splits is 10. Stratified k-fold is similar to k-fold, the difference is stratified k-fold will shuffle the data and split it into n_splits. Hyperparameter tuning is performed on each classifier to get the best accuracy from each classifier.

KNN is a classifier based on distance measurement [19]. KNN classifies an object based on the training data that is closest to the object. axillary odor data will be the objects that are classified into three classes. KNN has several parameters to perform classification, namely k as the number of nearest neighbors, weight as weight function, and metrics. Metrics in KNN are used to measure the distance between two data points. There are several metrics that can be used to measure distance, namely, Euclidean, Manhattan, and Minkowski.

SVM is a classifier that plots each data item as a data point in n-dimensional space [20]. SVM works by looking for data points from different classes and creating a boundary between them. The selected data points are called support vectors and their boundaries are called hyperplanes. To perform classification, SVM has several parameters, namely C, kernel, and gamma. C is the parameter used to control error. Kernel is a function that takes two data points as input and calculates its similarity score. Similarity can be defined as a proximity metric. Gamma is a parameter used to weight the curvature of the decision boundary.

RF is a classifier which used bootstrap sample method from the original data sample [21]. Classification tree is created by each data samples. To split the node, each data sample uses a feature subset. Each classification tree produces the prediction on a new data sample. The prediction which has the most votes will be the final prediction of RF. To perform classification, the parameters used by RF are max_depth, criterion, max_features, n_estimators, min_sample_split, and min_samples_leaf. This paper will conduct hyperparameter tuning to find the most optimal model for each classifier.

c) Evaluation

Several metrics are used to evaluate the performance of best classifier, namely accuracy (*Acc*), precision (*P*), recall (*R*), and F1-score (*F1*) [15]. Accuracy is the ratio of the correct number of samples with the entire sample. Precision is the number of positive samples classified as correct divided by the total samples classified as positive samples. Recall is the number of samples classified as positive divided by the total samples in the testing set that are labeled as positive, while the F1-score is the mean of recall and precision. Equation (10) to

(13) are equations to calculate accuracy, precision, recall and F1-score.

$$Acc = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100\% \quad (10)$$

$$P = \frac{TP}{(TP+FP)} \times 100\% \quad (11)$$

$$R = \frac{TP}{(TP+FN)} \times 100\% \quad (12)$$

$$F1-score = \frac{2 \times (R \times P)}{(R+P)} \times 100\% \quad (13)$$

In this study, confusion matrix also used to evaluate the performance of best classifier. The confusion matrix provides an overview of the learning extent of the proposed methods and their ability to produce accurate classifications. In this research, confusion matrix stores classification result information for class morning, class afternoon, and class evening.

d) Estimate gas concentration

Before estimating the gas concentration produced in morning, afternoon, and evening, selection feature is required to select the most significant features in classification by using ANOVA f-test (Analysis of variance). ANOVA f-test can be used to show correlation between features and data target [22]. The higher the ANOVA f-test value, the greater the dependence of the feature on the target variable. After we get the significant features, we will estimate the gas concentration using the Confidence Interval (CI) 95% for selected features. Confidence interval is a parameter used to determine the accuracy of the mean of a sample. By calculating the average value of these measurements, we can estimate the actual value for that measurement. The 95% CI means that if we take 100 samples, it is likely that 95 samples will include the true population mean value. The 95% CI will be used to calculate the average gas produced by human axillary odor in the morning, afternoon, and evening.

III. RESULTS AND DISCUSSION

1) The result of Regression stage

Ground truth of the Rs/Ro value and the PPM value for each gas from each sensor datasheet will be divided into 80% training data and 20% testing data. The ground truth data are obtained from 7 sensors with 22 gasses. The data training obtained from ground truth will be used as the input to build the KNN regression model, and the data testing will be used to check the performance of the KNNR model. We used two parameters to build the KNNR model, namely *k*, which value is 1 and uniform as the parameter to measure the similarity. Performance of the KNNR model, the MSE value is 0 and R^2 is 0.99, these two metrics indicates that KNNR model has a good performance. After successfully building the regression model, we will use this regression model to predict the PPM value of the dataset acquired from data acquisition in the form of Rs/Ro. Rs/Ro dataset has seven columns represent the Rs/Ro value from each sensor. After the prediction, the new dataset in the form of PPM gas value will have 22 columns, each columns represent the PPM gas value. We will save the new dataset into a CSV file. This new dataset will be the input for the classification stage.

2) The result of Classification stage

a) Pre-processing

The first step is to selecting row obtained from sensing stage only, therefore from each CSV file, we only used row 600 to 1800. The next step is to conduct statistical parameter feature extraction. Statistical feature extraction produces four features for each column. Because the number of columns is 22, the total number of features produced in this stage is 88 features.

b) Classification

This paper compares three classifiers, which are KNN, SVM, and RF, then evaluates them based on the accuracy value. The hyperparameter tuning was done automatically using pipeline and Grid Search CV function from sklearn. Grid Search is one of the classification hyperparameter optimization methods that will build several combinations of the given hyperparameters in the pipeline. The list of hyperparameters to be compared for each base model can be seen in Table 2.

The accuracy comparison of classifier can be seen in Table 3. Based on Table 3, by performing hyperparameter tuning, the accuracy of each classifier increased. The RF classifier method with hyperparameters set produces the highest accuracy, which is 87.43%, this result is an increase from before the hyperparameter tuning was carried out, which was 82.02%. The highest accuracy is achieved with the best hyperparameter of RF using the Gini criterion, the number of estimators is 200, the depth level of the decision tree is 8. Followed by SVM with an accuracy of 84.38% and KNN with an accuracy of 83.13%.

c) Evaluation

Detailed performance metrics of experimental methods for the best classifiers (RF) can be seen in Table 4. Precision,

TABLE II. CLASSIFIER AND THEIR HYPERPARAMETERS

Classifier	Hyperparameters
KNN	*k* : range(1, 21, 2), weights : ['uniform', 'distance'], metric' : ['euclidean', 'manhattan', 'minkowski']
SVM	C : [0.001, 0.1, 1, 10, 100, 10e5], Kernel : ['linear', 'rbf'], Gamma : [0.001, 0.01, 0.1, 1, 10, 100]
RF	n_estimators : [100, 200, 300], criterion : ['gini', 'entropy'], max_features : ['auto', 'sqrt', 'log2'], max_depth : [5, 8, 15, 25, 30], min_samples_split : [2, 5, 10, 15, 100], min_samples_leaf : [1, 2, 5, 10]

TABLE III. THE ACCURACY COMPARISON OF CLASSIFIER BEFORE AND AFTER HYPERPARAMETER TUNING

Classifier	Acc before	Best Hyperparameter	Acc after tuning
KNN	67.98%	metric : 'manhattan' n_neighbors : 6 weights : 'distance'	83.13%
SVM	35.77%	C : 10 gamma : 0.1 kernel : rbf	84.38%
RF	82.02%	criterion : 'gini' max_features : auto n_estimators : 200 max_depth : 8 min_samples_split : 2 min_samples_leaf : 1	87.43%

recall and f1-score for class evening are 0.91, 0.85, and 0.88, respectively. Precision, recall and f1-score for the morning are 0.88, 0.86, and 0.87, respectively. Precision, recall and f1-score for class afternoon are 0.84, 0.90, and 0.87, respectively. Fig. 3 is the confusion matrix of RF, which provided the best accuracy in the experiment. It can be seen from the 48 samples belongs to class evening, 41 samples were correctly predicted as class evening, while four samples wrongly predict as class morning, and three samples wrongly predict as class afternoon. For 58 samples that belong to class morning, eight samples were incorrectly predicted, one incorrectly predicted to be class evening, and seven incorrectly predicted to be class afternoon. Then from 59 samples from class afternoon, three samples were wrongly predicted to be class evening and three samples were wrongly predicted to be class morning.

d) Estimate gas concentration

The ANOVA f-test assessment result can be seen in Fig. 4, there are two most significant gas in the classification stages, which are ethanol and methane from sensor TGS 2612 with ANOVA f-test value 20.22 and 19.82, respectively. After obtained the most significant gas, we can estimate the gas concentration produced in the morning, afternoon, and evening using 95% CI. The 95% CI result can be seen in Table 5, that ethanol and methane gas concentration in the morning are 14.80 PPM and 3.23 PPM respectively. The ethanol and methane concentration in the afternoon are 15.03 PPM and 3.45 PPM, respectively; their concentration rise significantly from morning to afternoon. The ethanol and methane concentration in the evening are 15.06 PPM and 3.48 PPM, respectively. Their concentration slightly increases from afternoon to evening. In the morning, generally people only do a little activity, therefore the concentration of gas produced by axillary odor is still small, then during the day the activity increases, therefore the concentration of gas produced by axillary odor is increasing. Meanwhile, at evening, activity will decrease, hence the increase in the concentration of gas produced by axillary odor will not be as much as during the day. This result showed that the gas concentration produced depends on the intensity of the activities.

TABLE IV. THE PERFORMANCE METRICS OF RF

Class	Metrics			
	Precision	*Recall*	*F1-score*	*Accuracy*
Morning	0.88	0.86	0.87	87.43%
Afternoon	0.84	0.90	0.87	
Evening	0.91	0.85	0.88	

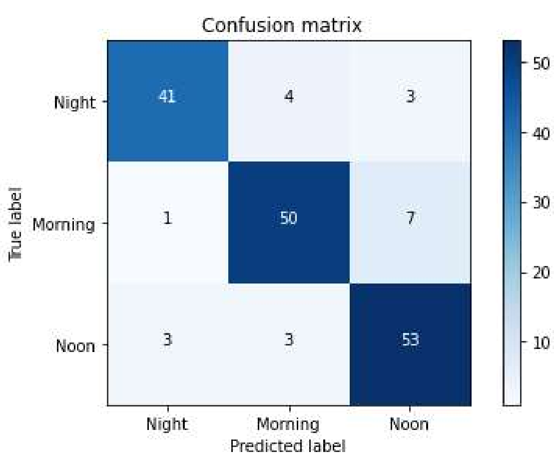


Fig.3. Confusion matrix of RF

IV. CONCLUSION

In this paper, e-nose can distinguish human axillary odor in the morning, afternoon, and evening. Three classifiers were chosen to classify three classes, which are RF, SVM, and KNN, which produced the highest accuracy is RF with an accuracy of 87.43%, followed by SVM with an accuracy of 84.38%, and KNN with an accuracy of 83.13%. Based on ANOVA f-test assessment, it is known that methane and ethanol detected from TGS 2612 were the most significant features in the classification stage. Using CI 95%, it has been shown that human axillary odor produced different amount of methane and ethanol gas concentration in the morning, afternoon, and evening. Ethanol and methane gas

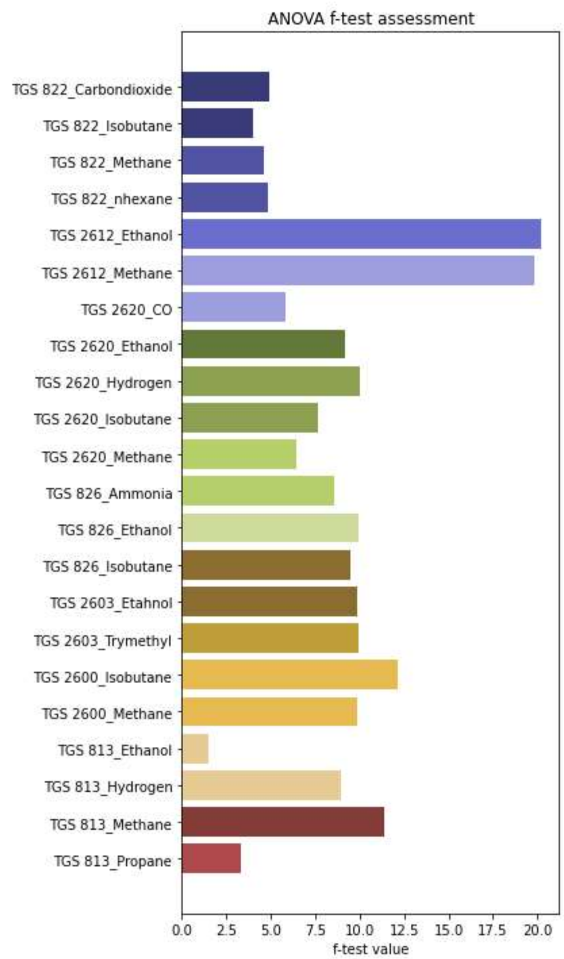


Fig. 4. ANOVA f-test value for each feature

TABLE V. ETHANOL AND METHANE GAS CONCENTRATION IN EACH CLASS

Class	Ethanol	Methane
Morning	14.80 PPM	3.23 PPM
Afternoon	15.03 PPM	3.45 PPM
Evening	15.06 PPM	3.48 PPM

concentration will rise significantly from morning to afternoon and then slightly increase from afternoon to evening. The gas concentration of ethanol and methane in the morning are 14.80 PPM and 3.23 PPM, 15.03 PPM and 3.45 PPM at afternoon, and 15.06 PPM and 3.48 PPM in the evening. This result has shown that the gas concentration produced depends on the intensity of activity.

For future works, we will try to optimize the classification of the axillary odor using ensemble learning or deep learning. Also, in the future e-nose system can be used to detect COVID-19 disease through axilla odor.

ACKNOWLEDGMENT

This research was funded by Special Program for Research Against COVID-19 (SPRAC) managed by AUN/SEED-Net and the Indonesian Ministry of Education and Culture under Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT) Program, under Riset Inovatif-Produktif (RISPRO) Invitation Program managed by Lembaga Pengelola Dana Pendidikan (LPDP), and under PMDSU Program managed by Institut Teknologi Sepuluh Nopember (ITS).

REFERENCES

[1] S. Dragonieri, G. Pennazza, P. Carratu, and O. Resta, "Electronic Nose Technology in Respiratory Diseases," *Lung*, vol. 195, no. 2. Springer New York LLC, pp. 157–165, Apr. 01, 2017, doi: 10.1007/s00408-017-9987-3.

[2] Hariyanto, R. Sarno, and D. R. Wijaya, "2017 International Conference on Information & Communication Technology and System (ICTS)," *Detect. Diabetes from Gas Anal. Hum. Breath using E-Nose*, vol. 0, pp. 241–246, 2017.

[3] L. Capelli *et al.*, "Application and uses of electronic noses for clinical diagnosis on urine samples: A review," *Sensors (Switzerland)*, vol. 16, no. 10. MDPI AG, Oct. 14, 2016, doi: 10.3390/s16101708.

[4] S. Wakhid, R. Sarno, S. I. Sabilla, and D. B. 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, 2020, doi: 10.22266/ijies2020.0831.06.

[5] J. Nicolas, A. C. Romain, V. Wiertz, J. Maternova, and P. André, "Using the classification model of an electronic nose to assign unknown malodours to environmental sources and to monitor them continuously," *Sensors Actuators, B Chem.*, vol. 69, no. 3, pp. 366–371, Oct. 2000, doi: 10.1016/S0925-4005(00)00487-1.

[6] A. H. Gómez, G. Hu, J. Wang, and A. G. Pereira, "Evaluation of tomato maturity by electronic nose," *Comput. Electron. Agric.*, vol. 54, no. 1, pp. 44–52, Oct. 2006, doi: 10.1016/j.compag.2006.07.002.

[7] S. I. Sabilla and R. Sarno, "Development of wavelet transforms to predict methane in chili using the electronic nose," in *Proceeding - ICAMIMIA 2017: International Conference on Advanced Mechatronics, Intelligent Manufacture, and Industrial Automation*, Jun. 2018, pp. 271–276, doi: 10.1109/ICAMIMIA.2017.8387600.

[8] T. Pobkrut and T. Kerdcharoen, "Soil sensing survey robots based on electronic nose," in *International Conference on Control, Automation and Systems*, Dec. 2014, pp. 1604–1609, doi: 10.1109/ICCAS.2014.6987829.

[9] T. Eamsa-Ard, M. M. Swe, T. Seesaard, and T. Kerdcharoen, "Development of Electronic Nose for Evaluation of Fragrance and Human Body Odor in the Cosmetic Industry," *2018 IEEE 7th Glob. Conf. Consum. Electron. GCCE 2018*, pp. 592–593, 2018, doi: 10.1109/GCCE.2018.8574831.

[10] M. Shu, Y. Liu, and H. Fang, "Identification authentication scheme using human body odour," *Proc. - 2014 IEEE Int. Conf. Control Sci. Syst. Eng. CCSSE 2014*, pp. 171–174, 2015, doi: 10.1109/CCSSE.2014.7224531.

[11] T. Eamsa-Ard, T. Seesaard, and T. Kerdcharoen, "Human Odor Sensing for Health Status Detection and Tracking by Using Electronic Nose," *2019 IEEE Int. Conf. Consum. Electron. - Taiwan, ICCE-TW 2019*, vol. 100, no. 1, pp. 2019–2020, 2019, doi: 10.1109/ICCE-TW46550.2019.8991859.

[12] C. Wongchoosuk, T. Youngrod, H. Phetmung, M. Lutz, T. Puntheeranurak, and T. Kerdcharoen, "Identification of people from armpit odor region using networked electronic nose," *2011 Def. Sci. Res. Conf. Expo, DSR 2011*, pp. 3–6, 2011, doi: 10.1109/DSR.2011.6026826.

[13] I. Mogilnicka, P. Bogucki, and M. Ufnal, "Microbiota and malodor—etiology and management," *Int. J. Mol. Sci.*, vol. 21, no. 8, pp. 1–21, 2020, doi: 10.3390/ijms21082886.

[14] Y. Y. Broza, L. Zuri, and H. Haick, "Combined volatolomics for monitoring of human body chemistry," *Sci. Rep.*, vol. 4, pp. 2–7, 2014, doi: 10.1038/srep04611.

[15] S. I. Sabilla, R. Sarno, K. Triyana, and K. Hayashi, "Deep learning in a sensor array system based on the distribution of volatile compounds from meat cuts using GC–MS analysis," *Sens. Bio-Sensing Res.*, vol. 29, no. July, p. 100371, 2020, doi: 10.1016/j.sbsr.2020.100371.

[16] Hariyanto, R. Sarno, and D. R. Wijaya, "Detection of diabetes from gas analysis of human breath using e-Nose," in *Proceedings of the 11th International Conference on Information and Communication Technology and System, ICTS 2017*, Jan. 2018, vol. 2018-January, pp. 241–246, doi: 10.1109/ICTS.2017.8265677.

[17] "TGS 2600-for the detection of Air Contaminants." Accessed: Jun. 12, 2021. [Online]. Available: www.figaro.co.jp.

[18] A. Navot, L. Shpigelman, N. Tishby, and E. Vaadia, "Nearest Neighbor Based Feature Selection for Regression and Its Application to Neural Activity," in *Proceedings of the 18th International Conference on Neural Information Processing Systems*, 2005, pp. 996–1002.

[19] M. H. Dunham, "Data mining introductory and advanced topics," *Prentice Hall*, 2006.

[20] N. Abramson, D. Braverman, and G. Sebestyen, "Pattern recognition and machine learning," *IEEE Trans. Inf. Theory*, vol. 9, no. 4, pp. 257–261, 1963, doi: 10.1109/TIT.1963.1057854.

[21] M. Ijaz, A. ur Rehman, M. Hamdi, and A. Bermak, "Recursive Feature Elimination with Random Forest Classifier for Compensation of Small Scale Drift in Gas Sensors," Sep. 2020, pp. 1–5, doi: 10.1109/iscas45731.2020.9181246.

[22] S. M. Scariano and J. M. Davenport, "The Effects of Violations of Independence Assumptions in the One-Way ANOVA," *Am. Stat.*, vol. 41, no. 2, pp. 123–129, 1987.