Men and Women Classification at Night Through the Armpit Sweat Odor using Electronic Nose

Irzal Ahmad Sabilla
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
irzal.ahmad.s@gmail.com

Doni Putra Purbawa
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
doniputrapurbawa@gmail.com

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

Asra Al Fauzi
Department of Neurosurgery
Airlangga University
Surabaya, Indonesia
asra.al@fk.unair.ac.id

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

Rudy Gunawan
Department of Neurosurgery
Airlangga University
Surabaya, Indonesia
rudygunawan10@gmail.com

2021 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob) | 978-1-7281-9475-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/APWIMOB51111.2021.9435273

***Abstract*— Sweating at night can be an indication that there is a disturbance in the human metabolic system. Sweat itself is a substance that is unused in the body or the result of human excretion. The sweat glands are scattered in all parts of the body, but mostly in three locations: armpits, palms, and feet. Several kinds of research related to sweat and Electronic Nose (E-Nose) have also been studied. The study used a patch to absorb sweat and proved the presence of nicotine content from a smoker. However, the previous research has not focused on human sweat at night for potential disease. This paper aims to propose a system to distinguish men and women at night through the armpit sweat odor using Taguchi Gas Sensors (TGS) and SHT15. Researchers found four significant sensors for further investigation: TGS 822, TGS 826, TGS 833, and TGS 2620. This study obtained a total of 165 armpit sweat data, which have been processed and adjusted for this case into 25 data, 12 men (ME) and 13 women (WO). Several classification models are implemented, such as Support Vector Machine (SVM), Naïve Bayes (NB), and Decision Tree (DT) with accuracy 92.30%, 96.15%, and 84.62%, respectively. Based on the highest accuracy and the Volatile Organic Compounds (VOC) measurement, women are more likely to suffer from several diseases than men, such as leukemia.**

Keywords—ANNOVA, Electronic Nose, Gender, Machine Learning

I. INTRODUCTION

In general, men and women can be distinguished easily by their anatomy and physique. The study [1] tested men and women's physical performance with several workouts, such as repeated reach-up task, forward reach, 50-ft walk, and distance walked on 6 minutes. Women feel tired more quickly than men, but the difference is not too significant. Another study [2] states that women have stronger immunity than men due to biological factors such as genetic factors and hormones: androgens, progesterone, prolactin, and estrogens.

Sweat is a result of human secretions that are produced by the apocrine sweat glands. These glands will active if there are hormonal changes at puberty. Apocrine glands are found on the head, axilla, anogenital, eyelids, face, stomach, and armpits. Primary disorders of the apocrine glands are Bromhidrosis and Chromhidrosis. The smell of a person's sweat is excessive than usual and usually lies in the armpits, scalp, between the fingers, soles of the feet, and genitals [3].

A researcher in Switzerland revealed that men's sweat smells like cheese, while women have a similar odor like onions [4]. Another study [5] stated that the resulting sweat odor depended on what they consumed, by experimenting with eating meat for 15 days. The results of the samples taken in the armpits showed the sweat odor is similar to meat. That case can be detected because there are several chemical compounds or Volatile Organic Compounds (VOC) in the sweat content.

VOC from sweat, respiration, and skin have been used as biomarkers for humans to identify and diagnose disease [6]. The VOC characteristics of human respiration are also commonly used to detect several diseases such as cancer, lung, diabetes, and hepatitis. In the study [6], analyzes were performed using gas chromatograph and spectrograph, and 100 compounds were identified, some of which varied with age. Another study proved the nicotine content by placing a patch on the respondent's upper arm for 72 hours [7].

Leukemia is a cancer of white blood cells (WBC) that can be indicated through breathing and sweating [8]. Red blood cells (RBC), platelets, and WBC are components of blood. In general, leukemia has two categories based on its growth process, namely acute and chronic. Infected WBC overgrows in acute leukemia, whereas chronic leukemia expands slowly [9]. Chronic Lymphocytic Leukemia (CLL) is a type of chronic leukemia that grows deliberately, and many people don't know the symptoms. However, in some cases, night sweats can present as a symptom even before a diagnosis of CLL is made. When consulting a doctor, it is usually reinforced by other supporting symptoms such as weight loss, fatigue, bruising, enlarged spleen, and enlarged lymph nodes.

In this developing world, there are a lot of researches related to machine learning. Machine learning itself can be applied or implemented in various scientific fields. One example is in the health sector, such as detecting a disease from a person's body through substances contained in breath, sweat, or based on heart rate, oxygen saturation, and others.

In Swapna G's study [10], a machine learning classification was applied to early detection of a diabetic patient based on Heart Rate Variability (HRV) signal data.

Currently, E-Nose is widely used to research various scientific problem because it is cheap, fast, and of course, accurate. In recent year, E-Nose has been developed as a mul-

TABLE I. GAS SENSOR PACKAGES ON E-NOSE

Sensor	Volatile Compound Target
SHT 15	Humidity, temperature
TGS 2600	CO, isobutane, ethanol, hydrogen, methane
TGS 2603	Ethanol, hydrogen, methyl mercaptan, trimethyl amine
TGS 2612	Methane, isobutane, ethanol, propane
TGS 2620	Ethanol, isobutane, hydrogen, CO, methane
TGS 813	CO, methane, ethanol, propane, isobutane, hydrogen
TGS 822	Methane, acetone, CO, ethanol, isobutane, n-hexane, benzene
TGS 826	Ammonia, isobutane, ethanol, hydrogen
TGS 832	Chlorodifluoromethane (R-22), dichlorodifluoromethane (R-12), 1,1,1,2-tetrafluoroethane (R-134a), ethanol

tifunctional instrument for detecting the authenticity of beef with pork mix [11], [12], detection of food quality [13], [14], detection of borax on meatballs [15], detection of diabetic patients [16] and lung cancer from inhalation [17]. So, it is possible to distinguish men and women at night based on their armpit sweat odor's VOC using E-Nose.

In this study, we propose a model to distinguish men and women at night based on the analysis of several biomarkers in sweat. We use an electronic nose with a package of TGS and SHT sensors. The resulting output is signal data represented with a CSV file, and it is sent to the computer. That data will be processed and used for the classification using machine learning. After classification, we also measure the VOC composition to know which women and men are more likely to suffer from several diseases, such as leukemia. To better understand how the system works, Section II will explain the sampling process, material, and method details. Section III describes the results of the research and discussion. Section IV describes the conclusions of the study.

II. MATERIALS AND METHODS

This research's main objective is the classification of men and women at night through the analysis of substances in sweat captured by the E-Nose sensors.

A. Materials

This research uses hardware from Google Collaboratory (Google Colab) with the following specifications: AMD EPYC 7B12 processor with 2CPU, 13GB RAM, and 114GB hard drive size. This device is accessed via an internet connection at this link colab.research.google.com. E-Nose is a device that designed to resemble the human sense of smell. E-nose can be used to identify an object based on the smell and chemical compounds. One E-nose kit consists of the SHT 15 series sensor packages and eight Taguchi Sensor Series (TGS) packages. We can see in Table I that each sensor has the expertise to detect various volatile compounds. The TGS 832 sensor is sensitive to chlorofluorocarbons (CFCs), such as R-12, R-134a, and Ethanol. However, TGS 2612 is more susceptible to methane and LPG gases such as Ethanol, Methane, Iso-Butane, and Propane. E-Nose generally consists of several components, namely sample chamber, sensor array chamber, control circuit, power supply, fan, voltage divider, headspace system, and converter to convert analog signals to digital [18]. The base component of E-Nose used in this research is Arduino Microcontroller.

The sensor output on the E-Nose in the form of a digital signal will be processed in this study using the Python programming language. Then it will be classified, evaluated, and visualized by using several libraries such as pandas, glob, NumPy, seaborn, time, pywt, filecmp, and skelarn.

B. Data Acquisition

In this study, data was taken from the armpit connected to the vacuum pump on the e-nose system. The underarm aroma is sucked in by the pump, released into the sensor tube. Each sensor detects the aroma and sends signal data to microcontroller in the form of analog data. The microcontroller receives a signal from the sensor which is converted into signal data to be sent to the computer as the role of machine learning in the form of a Comma Separated Values (CSV) file represented in Figure 1. The sample data were obtained by taking a sample of the respondent's armpit for about 280 seconds. The first 60 seconds is the flushing phase, the next 120 seconds is the sampling phase and the next 100 seconds is the purging phase. In the sampling data process, there are several stages and limitations in this research.

First, turn on the power on the E-Nose and wait for about 1 minute. Then plug the end of the polyurethane hose at the input port, and the other end is narrowed at the respondent's armpit. Next, press the start button to start sampling the data. In this sampling phase, or at 61 - 180 seconds, the respondent has to reduce their move because it starts working and absorbs odors in their armpits.

After the sampling process is complete, the E-nose will automatically generate sensitivity data for each sensor into a CSV file. In this study, the amount of data used was 165 Armpit Sweat Odor Data (ASOD), which consisted of ages four to 70 years, morning, afternoon, night time, and male and female sex. This dataset will be filtered and processed according to this study's needs in the preprocessing stage, precisely the sampling time at night, the gender of men and women in adulthood, or 18 years over [19].

The results of plotting the ME and WO samples in Figure 2 show a significant difference in the gas sensor's sensitivity at TGS 832, TGS 822, TGS 826, and TGS 2620. The other sensors show only slight sensitivity, even almost invisible. Each sensor's unit value is still expressed in source resistance/output resistance (Rs / Ro).

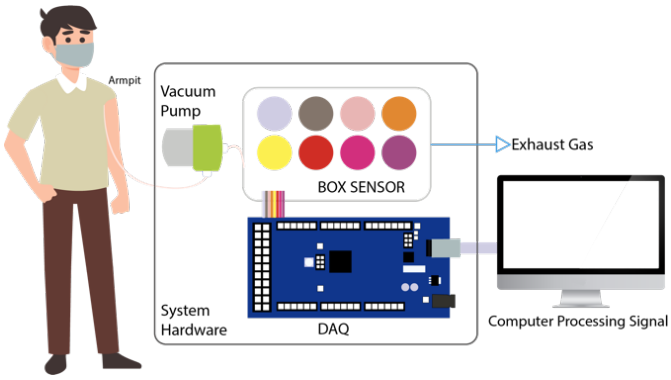
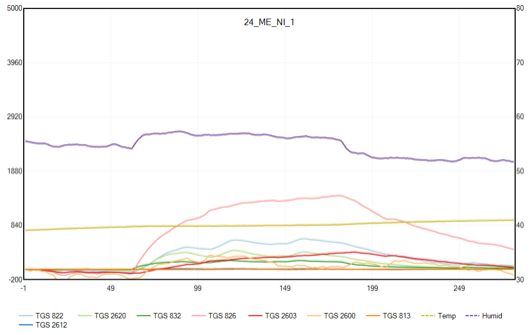
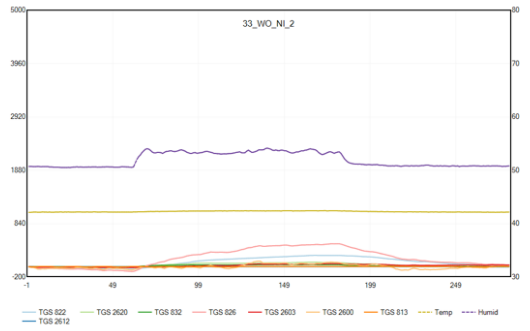


Fig. 1. Electronic Nose Working Flow



(a)



(b)

Fig. 2. Men (a) and Women (b) Armpit Signal Responses

This value will later be regressed with existing ppm data using K Nearest Neighbor (KNN) [20]. So that the value of each substance in a sensor will be obtained in units per million (PPM), or the equivalent of 1 mg / L.

C. Proposed Method

The stages in this study are preprocessing, classification, and evaluation, as shown in Figure 3.

1. Preprocessing:

There are four phases in data preprocessing. The first is processing and cleaning the original signal or raw data to ensure the data is ready to process. Next is the statistical feature extraction stage, which is the phase to find sensor relationships and summarize all CSV files' data into one representation value. After that, there will be a feature reduction based on the p-value index and the score of each feature in the feature selection stage. In a previous study [21], a feature selection stage confirmed that there are a sensor reduction and accurate improvement.

Feature Correlation is used to see the level of correlation between one feature and another. This technique requires a threshold for selecting which features to ignore and which ones to use. In chapter 3, a heatmap diagram representing each correlation value between features will be shown to describe more. Statistical Feature Extraction uses to summarize all data in a CSV file using aggregation techniques or applying mathematical formulas. In this study, the methods we use are maximum (amax), minimum (amin), mean (mean), and standard deviation (std). Equation (1), (2), (3), (4) is used to calculate the four statistical parameters above.

$$V = \max(y(t)) \tag{1}$$

$$S = \min(y(t)) \tag{2}$$

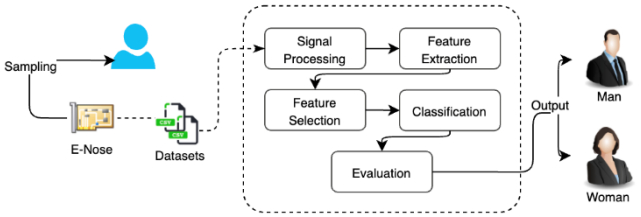


Fig. 3. Research Methodology Flow

Where V and S are smallest and largest values in a signal on the $y(t)$.

$$\mu = \frac{1}{n}\, x \sum_{t=0}^{n} y(\mathrm{t}) \tag{3}$$

Where μ is the average of all signals, and n is the number of signals, while $y(t)$ is the time-related signal.

$$\sigma = \sqrt{\frac{1}{n}\, x \sum_{t=0}^{n} (y(\mathrm{t}) - \mu)^2} \tag{4}$$

Where σ is the standard deviation of signal y, n is the number of, $y(t)$ is the signal against (t), and μ is the average of the whole signals.

Feature Selection is also required for this stage, because there are four times features after the aggregation process. Analysis of Variance (ANOVA) can help to show the correlation between each feature and its target. The null hypothesis, features that unrelated to the target, can be ignored if the score is higher than the p-value.

Principal Component Analysis (PCA) is a feature extraction process that creates new parallel / linear features based on the initial feature combination. PCA maps each instance of a given dataset in dimensional space d to subspace k so that k<d. The resulting new set of k dimensions is called a principal component (PC), and each major part is directed to the maximum variance, excluding variants that have been taken into account in all the previous elements [22]. Principal Components are represented using Equation 5.

$$PC_i = a_1 X_1 + a_2 X_2 + \cdots + a_j X_j \tag{5}$$

Where PC_i is the principal component, X_j is the original feature of j; and a_j is the coefficient of X_j. PCA is the most popular feature extraction method and a dimensional reduction in machine learning, pattern recognition, signal and image processing, and exploratory data analysis.

Due to the small amount of data, this study tries to find a correlation between the number of datasets using oversampling in the dominant class. It generates synthetic examples by operating in "feature space" rather than "data space". The minority class is over-sampled by taking each minority class sample, joining all nearest neighbors and generating new data on it vectors.

When we want to train our ML model in machine learning, we split our entire dataset into *training_set* and *test_set* using *train_test_split()* class present in sklearn. But, when we change *random_state*, the accuracy we get is also different, so that we cannot establish an exact accuracy. Hence, we used StratifiedKFold for this research also to prevent overfitting of a dataset. StratifiedKFold is the same as KFold cross-validation but is carried out stratified instead of random

sampling. In this study, we further evaluated the models using n_split 5 or we could say data training 80% and testing 20%.

2. Classification:

This study compares several classification methods and evaluates them based on an accuracy matrix. However, this research will focus on SVM, NB, and DT as classifiers.

The SVM method has been widely used in machine learning for classification, regression, clustering, and anomaly detection applications. In this study, SVM was used to classify men and women at night based on the armpit sweat odor sample. In the study, PCA and SVM were used to differentiate normal and odorous bodies with an accuracy of 86%. In some cases, the data dimensions must be transformed until they can be separated into a hyperplane. The hyperplane of the linear SVM is defined in Equation 6.

$$w^T x + b = 0 \tag{6}$$

Where w^T denotes the normal line to the hyperplane, and x is the p-dimensional vector point of the input or sensor data coming from the E-Nose. Whereas b represents a bias term.

Naïve Bayes can do well with small numbers of training data to solve classification problem. We supposed to have a sample tuple T of eight sensors as follows (Equation 7):

$$T = \{S_1, S_2, \ldots, S_8\} \tag{7}$$

Where S_1 to S_8 represent the eight features or sensors on the E-Nose. Each instance refers to the Men (C_{men}) and Women (C_{women}) classes. Naïve bayes classifier can predict T as ME class, if it has a higher posterior score than WO class ($P(C_{men}|T) > (P(C_{women}|T)$. The posterior probability of naïve bayes can be seen Equation 8.

$$P(C|T) = \frac{P(T|C)\,P(C)}{P(T)} \tag{8}$$

Where $P(T|C)$, $P(C)$ are the likelihoods and each probability prior. Since $P(T)$ is a constant value, it takes a maximum value of $P(T|C)\,P(C)$ to build a classifier from a probability model. Thus, the naïve bayes classification can be expressed as a function that assigns the label $y = c$ to as many classes as follow:

$$\hat{y} = \underset{i\in\{1\ldots I\}}{\arg\max}\left(P(C_i)\prod_{j=1}^{n} P(T_j \mid C_i)\right) \tag{9}$$

Where n is the total feature of data that in this study were analyzed using python and arg max is a function for finding the class with the largest predicted probability.

The Decision Tree is also a powerful and popular classification and prediction model. DT represents rules that are easy for humans to understand and use in knowledge systems such as databases. DT can solve both numerical and categorical data problems. Decision node to determine the test on a single attribute. The leaf node shows the target attribute value—edge as a separation of one feature. The path is a disjunction of the test to make the final decision. This study implements the decision tree classifier by considering several grid search parameters.

3. Evaluation:

The evaluation technique used in this research is a confusion matrix. Data sampling or ground truth is divided into two classes, namely the original and predictive classes. In measuring the performance of a model, four values can represent the classification results. False Positive (FP) is negative data that is predicted to be positive. False Negative (FN) is data positive but is predicted to be negative. True Positive (TP) is positive data that is detected as positive, and True Negative (TN) is negative data that is detected as negative. Based on the confusion matrix, we can determine accuracy, precision, recall, and specificity, respectively Equation (10), (11), (12), and (13).

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{10}$$

$$Precision = \frac{TP}{(TP+FP)}\ x\ 100\% \tag{11}$$

$$Recall = \frac{TP}{P} = \frac{TP}{(TP+FN)} = 1 - FNR \tag{12}$$

$$Specificity = \frac{TN}{N} = \frac{TN}{(TN+FP)} = 1 - FPR \tag{13}$$

III. Results and Discussion

A. Preprocessing

The first stage is signal processing to ensure all signals have no empty values and are not duplicated. Each CSV file also needs to have 2800 rows of value because we will only take the data on the sampling phase, located at rows 600 to 1800. Its column average subtracts each value to ensure that the value is an odor value representation, not the sweat intensity. As we can see on the Figure 4, the dataset used in this study can be displayed into two clusters. The data well grouped by its features. However, one man (ME) as 36_ME_NI_1 data overlap into the woman class (WO). Total data after processed are 25 samples, 13 men and 12 women. The red dots represent the centroid of clusters, while the black and blue represent the data distribution.

Feature correlation is also used in this study to reduce sensors that are not correlated with sensitive sensors. By giving a threshold of 0.8 and selecting the TGS 2620, and TGS 832 sensors as keys, uncorrelated sensors are found, namely TGS 813, TGS 2600, TGS 2603, and TGS 2612. As we can see in Figure 5, the darker color, the more it uncorrelated or having a value far from one.

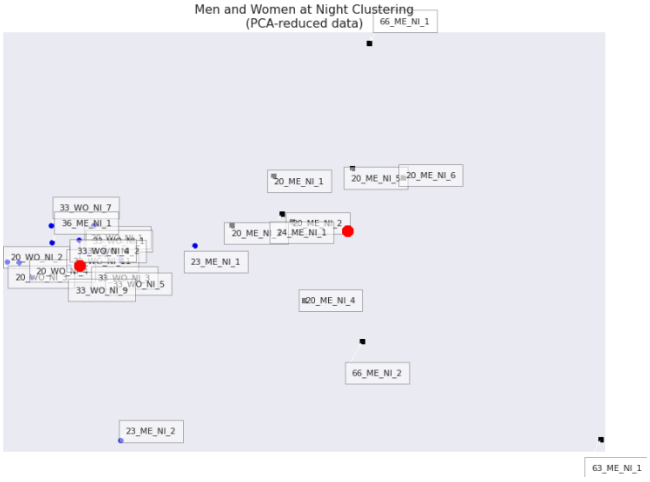


Fig. 4. Men and Women's Armpit Sweat Clustering using K-Means

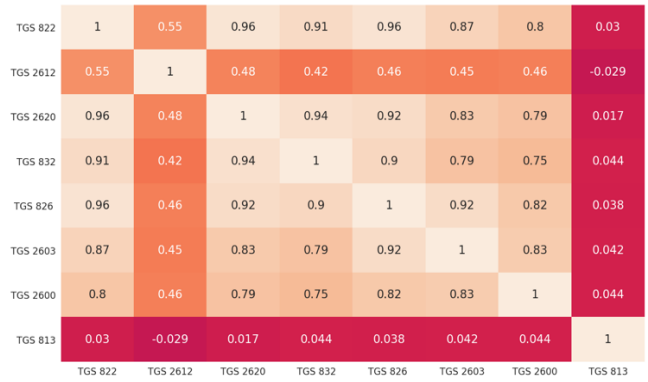


Fig. 5. Feature Correlation using Heatmap

The next phase is summarizing all data in a CSV file using aggregation techniques or applying mathematical formulas. In this study, we use amax, amin, mean, and std. After applying this statistical extraction, each sensor becomes four times. So, the total current feature is about 16 features.

After both data summarized and features extracted, we implemented Feature Selection using selectkbest, or we often say Analysis of Variance (ANOVA). This process helps to show the correlation between each feature and its target. The null hypothesis attributes unrelated to the target can be ignored if the score is higher than the p-value. Table II shows the entire sequence of features based on the score and p-value.

Because this research has small datasets, we use the cross-validation with 80% training and 20% testing, also PCA as a control before entering the classification phase. Table III will show the result differences. Implementing PCA can improve the accuracy for each model except the NB, which remains unchanged.

TABLE II. SELECTED FEATURES BY SCORE AND P-VALUE

Feature	Score	P-Value
TGS 2660, std	44.79	9.97e-07
TGS 826, std	41.90	1.64e-06
TGS 2620, amin	38.00	3.32e-06
TGS 826, amin	33.96	7.29e-06
TGS 832, amin	32.47	9.90e-06
TGS 826, amax	31.03	1.34e-05
TGS 822, amin	30.56	1.48e-05
TGS 822, std	24.54	5.88e-05
TGS 832, std	24.46	6.02e-05
TGS 2620, amax	22.60	9.59e-05
TGS 822, amax	15.21	7.70e-04
TGS 832, amax	11.63	2.51e-03
TGS 832, mean	2.60	1.21e-01
TGS 826, mean	2.28	1.45e-01
TGS 822, mean	1.19	2.87e-01
TGS 2620, mean	0.06	8.14e-01

B. Classification

TABLE III. CLASSIFICATION RESULT WITH AND WITHOUT PCA

Reduction	Model	Parameter	Accuracy
-	SVM	C=1, gamma = 0.1	84.00%
-	NB	Smoothing 0.001	96.00%
-	DT	Gini activation	84.00%
PCA	SVM	C=1, gamma = 0.1	92.00%
PCA	NB	Smoothing 0.001	96.00%
PCA	DT	Gini activation	88.00%

TABLE IV. FINAL CLASSIFICATION RESULT AFTER SMOTE

Model		Class		Accuracy
		MEN	*WOM*	
SVM	Precision	1	0.87	92.30%
	Recall	0.85	1	
	F-1 Score	0.92	0.93	
NB	Precision	1	0.93	96.15%
	Recall	0.92	1	
	F-1 Score	0.96	0.96	
DT	Precision	0.85	0.85	84.62%
	Recall	0.85	0.85	
	F-1 Score	0.85	0.85	

This study compares several classification methods and evaluates them based on an accuracy matrix. However, this research will focus on SVM, NB, and DT as classifiers. Based on the classification results in Table III, the most stable dan higher accuracy without PCA is NB in 96.00%. Next, we try to implement PCA before classification to see the difference and NB still gets better accuracy after implementing PCA than other models, which is 96.00%. DT increase 4% from the previous. SVM significantly increase by 12% of the prior experiment (without PCA).

C. Synthetic Minority Oversampling Technique (SMOTE)

Due to the small amount of data, this study tries to find a correlation between the number of datasets using oversampling in the dominant class [23]. After oversampling, the data from 12 for ME class and 13 for WO class each amounted to 13 (degenerate one for ME class). After that, the classification process is carried out again to find out the results. The accuracy results showed NB slightly increased to 96.15%, SVM increased to 92.30%, but DT seems slightly decreased to 84.62% as described in Table IV.

D. Evaluation

The evaluation results show that NB recall or sensitivity and specificity values are close to 1. Recall and specificity are 0.9615, which means that the proposed model is susceptible and specific. As we can see in Figure 6, there is one data that

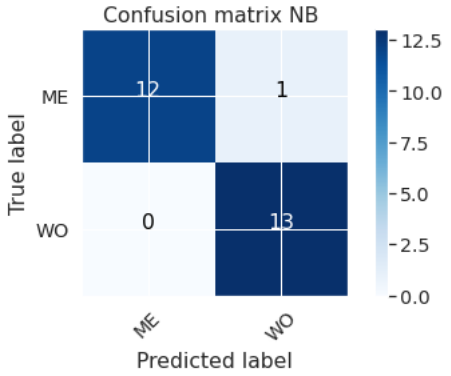


Fig. 6. Confusion Matrix of Optimized NB

was wrongly predicted. The original class or ground truth of the sample is ME but it predicted to be WO.

E. Measurement Gas Composition

To better determine what substances in sweat could differentiate between ME and WO at night, we further investigated the sample data results. The sample data taken is expressed in units of Rs / Ro. Therefore, we use the K-NN Regression to create a model from the ppm datasets, then create a new output with the PPM unit. The steps are as follows:

TABLE V. CHEMICAL COMPOUND ON THE ARMPIT SWEAT SAMPLE

Sensor	Chemical Compounds	ME (mg/L)	WO (mg/L)
TGS 822	Methane	10.30	10.47
	CO	4.94	5.21
	N-hexane	2.61	2.98
TGS 826	Ammonia	0.88	1.07
	Isobutane	2.14	2.50
	Ethanol	0.32	0.38
TGS 2620	Ethanol	3.29	3.64
	Isobutane	4.30	4.76
	Hydrogen	2.65	2.85
	CO	6.23	6.83
	Methane	16.85	18.03

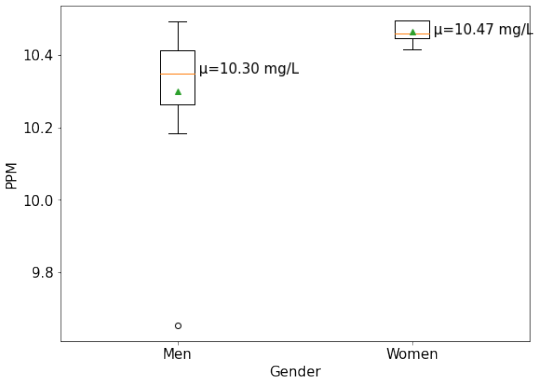


Fig. 7. Methane Gas Concentration on TGS 822

a) Make a model from data that already has ppm units with the appropriate sensor array and save it in a .sav file;

b) Import the model and regress the Rs / Ro data and get the ppm output value of each sensor;

c) After the ppm data is obtained, we summarize each data using an aggregation technique such as the Statistical Feature Extraction step;

Based on Table V, each sensor, which initially only had one value in Rs / Ro units, now has several ppm values based on its sensitivity

VOC in human armpit sweat odor has been measured using regression from existing results. The four TGS 822, TGS 826, TGS 832, and TGS 2620 sensors in Table 6 have different sensitivities. At night, the average WO contains more methane, CO, N-hexane, Ammonia, Isobutane, Ethanol, and Hydrogen than ME. In terms of methane gas concentration, the distribution of data can be seen in Figure 6, some men have high methane, but when taken the average for all classes, the results are below the average for women.

IV. CONCLUSION

In this study, the ME and WO classification processes were successfully carried out using E-Nose with a reasonably good accuracy level. The three SVM, NB, and DT models have an accuracy value of 92.30%, 96.15%, and 84.62%. This research also states a feature reduction or sensor array for this case, based on the correlation and feature selection results. By ignoring several sensors: SHT 15, TGS 813, TGS 2600, TGS 2603, and TGS 2612, the E-Nose array sensor becomes only four sensors.

The maximum accuracy achieved by the entire model in this study is 96.15% because it is also due to the lack of datasets and one overlap of data. The oversampling technique can prove it. The results of measuring gas concentrations with the boxplot match the clusters of both WO and ME classes. WO is more assembled in one place while ME is more spread out. At night, the armpit sweat odor in women is higher than in men, which is indicated by the volatile compound's intensity in Table V. Based on the chemical compound measurement, women are more likely to be affected by several diseases, likely leukemia. Therefore, in the subsequent research, more samples should be obtained for better performance and to measure women's possibility of getting cancer.

ACKNOWLEDGMENT

This research was funded by Special Program for Research Against COVID-19 (SPRAC) managed by AUN/SEED-Net and the Indonesian Ministry of Research and Technology or National Agency for Research and Innovation and the Indonesian Ministry of Education and Culture under Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT) Program, and under PMDSU Program managed by Institut Teknologi Sepuluh Nopember (ITS).

REFERENCES

[1] J. Q. Lee, M. J. Simmonds, X. S. Wang, and D. M. Novy, "Differences in Physical Performance between Men and Women with and Without Lymphoma," *Arch. Phys. Med. Rehabil.*, vol. 84, no. 12, pp. 1747–1752, 2003, doi: 10.1016/S0003-9993(03)00437-4.

[2] E. Ortona, M. Pierdominici, and V. Rider, "Editorial: Sex hormones and gender differences in immune responses," *Frontiers in Immunology*, vol. 10, no. MAY. Frontiers Media S.A., 2019, doi: 10.3389/fimmu.2019.01076.

[3] R. Chen *et al.*, "Sweat gland regeneration: Current strategies and future opportunities," *Biomaterials*, vol. 255. Elsevier Ltd, p. 120201, Oct. 2020, doi: 10.1016/j.biomaterials.2020.120201.

[4] "The Nose Knows: Men's Sweat Smells Like Cheese, Women's Like Onions | Discover Magazine." .

[5] J. Havlicek and P. Lenochova, "The Effect of Meat Consumption on Body Odor Attractiveness," *Chem. Senses*, vol. 31, no. 8, pp. 747–752, Oct. 2006, doi: 10.1093/chemse/bjl017.

[6] M. Gallagher, C. J. Wysocki, J. J. Leyden, A. I. Spielman, X. Sun, and G. Preti, "Analyses of volatile organic compounds from human skin," *Br. J. Dermatol.*, vol. 159, no. 4, pp. 780–791, Oct. 2008, doi: 10.1111/j.1365-2133.2008.08748.x.

[7] P. Kintz, A. Henrich, V. Cirimele, and B. Ludes, "Nicotine monitoring in sweat with a sweat patch," *J. Chromatogr. B Biomed. Appl.*, vol. 705, no. 2, pp. 357–361, Feb. 1998, doi: 10.1016/S0378-4347(97)00551-3.

[8] M. Shirasu and K. Touhara, "The scent of disease: Volatile organic compounds of the human body related to disease and disorder," *Journal of Biochemistry*, vol. 150, no. 3. pp. 257–266, Sep. 01, 2011, doi: 10.1093/jb/mvr090.

[9] N. Bibi, M. Sikandar, I. U. Din, A. Almogren, and S. Ali, "IOMT-based automated detection and classification of leukemia using deep learning," *J. Healthc. Eng.*, vol. 2020, pp. 1–12, Dec. 2020, doi: 10.1155/2020/6648574.

[10] G. Swapna, R. Vinayakumar, and K. P. Soman, "Diabetes detection using deep learning algorithms," *ICT Express*, vol. 4, no. 4, pp. 243–246, Dec. 2018, doi: 10.1016/j.icte.2018.10.005.

[11] B. Kuswandi, A. A. Gani, and M. Ahmad, "Immuno strip test for detection of pork adulteration in cooked meatballs," *Food Biosci.*, vol. 19, pp. 1–6, Sep. 2017, doi: 10.1016/j.fbio.2017.05.001.

[12] L. Yang *et al.*, "Rapid Identification of Pork Adulterated in the Beef and Mutton by Infrared Spectroscopy," *J. Spectrosc.*, vol. 2018, 2018, doi: 10.1155/2018/2413874.

[13] S. I. Sabilla, R. Sarno, and J. Siswantoro, "Estimating Gas Concentration using Artificial Neural Network for Electronic Nose," in *Procedia Computer Science*, Jan. 2017, vol. 124, pp. 181–188, doi: 10.1016/j.procs.2017.12.145.

[14] S. Qiu, L. Gao, and J. Wang, "Classification and regression of ELM, LVQ and SVM for E-nose data of strawberry juice," *J. Food Eng.*, vol. 144, pp. 77–85, Aug. 2014, doi: 10.1016/j.jfoodeng.2014.07.015.

[15] D. Sunaryono, S. Rochimah, R. Sarno, S. I. Sabilla, I. Ahmad Sabilla, and D. Sekarini, "Geotagging for mapping distribution of meatballs containing borax based on electronic nose detection," in *Proceedings - 2020 International Seminar on Application for Technology of Information and Communication: IT Challenges for Sustainability, Scalability, and Security in the Age of Digital Disruption, iSemantic 2020*, Sep. 2020, pp. 353–358, doi: 10.1109/iSemantic50169.2020.9234279.

[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-Janua, pp. 241–246, doi: 10.1109/ICTS.2017.8265677.

[17] D. M. Wong *et al.*, "Development of a breath detection method based E-nose system for lung cancer identification," in *Proceedings of 4th IEEE International Conference on Applied System Innovation 2018, ICASI 2018*, Jun. 2018, pp. 1119–1120, doi: 10.1109/ICASI.2018.8394477.

[18] S. N. Hidayat, A. Rusman, T. Julian, K. Triyana, A. C. A. Veloso, and A. M. Peres, "Electronic nose coupled with linear and nonlinear supervised learning methods for rapid discriminating quality grades of superior java cocoa beans," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 6, pp. 167–176, Dec. 2019, doi: 10.22266/ijies2019.1231.16.

[19] "Indonesia - Law on Child Protection (No. 23/2002)." .

[20] Z. Yao and W. L. Ruzzo, "A regression-based K nearest neighbor algorithm for gene function prediction from heterogenous data," *BMC Bioinformatics*, vol. 7, no. SUPPL.1, Mar. 2006, doi: 10.1186/1471-2105-7-s1-s11.

[21] 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, Aug. 2020, doi: 10.1016/j.sbsr.2020.100371.

[22] H. Roopa and T. Asha, "A Linear Model Based on Principal Component Analysis for Disease Prediction," *IEEE Access*, vol. 7, pp. 105314–105318, Jul. 2019, doi: 10.1109/access.2019.2931956.

[23] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.