Original Article
Obstet Gynecol Sci 2025;68(4):293-303
https://doi.org/10.5468/ogs.25038
eISSN 2287-8580

Obstetrics & Gynecology Science

Identification of volatile organic compound as a novel modality for cervical cancer detection

Muhammad Ary Zucha, MD, OBGYN, PhD[1], Dian Kesumapramudya Nurputra, MD, PhD[2], Ardhanu Kusumanto, MD, OBGYN[1], Shidiq Nur Hidayat, MSi[3], Kuwat Triyana, MSi[3], Widha Apriani Priadi, MD[1], Nadia Yasmine, SGz[4], Ova Emilia, MD, OBGYN, PhD[1]

Departments of [1]Obstetrics and Gynecology, [2]Child Health, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, [3]Department of Physics, Biomedical Science, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, [4]Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia

Objective
In this study, we developed a novel detection strategy based on volatile organic compound (VOC) sensing. Persistent infection by human papillomavirus (HPV) may cause biological changes in the cervical epithelium, leading to VOC production.

Methods
This study included 200 urine samples from cervical cancer patients and controls that were HPV-negative. Urine samples were collected and measured using a gas sensor array composed of a matrix of 10 sensors. For each analyzed sample, the instrument produced a vector signal encoding the VOC emitted from the urine (urine prints). The urine prints of cervical cancer patients were differentiated from those of healthy controls.

Results
Identification of VOC for cervical cancer detection showed reliable accuracy (91% sensitivity, 85% specificity, and 89% accuracy).

Conclusion
Our results demonstrated the applicability of VOC sensing for cervical cancer detection and its potential application in treatment monitoring.

Keywords: Artificial intelligence; Cancer screening test; Cervical cancer; Volatile organic compounds

Introduction

Cervical cancer is the second most common malignancy in women and one of the leading causes of death, especially in developing countries. Its estimated incidence is 348,809 cases, with a mortality rate of approximately 60% (207,210 deaths) [1]. Cervical cancer-related deaths are expected to reach 12 million by 2030. Approximately 18,000 new cases of cervical cancer are reported annually in Indonesia, with a first-year mortality rate of 75%. The majority of newly diagnosed patients at an advanced stage at the time of diagnosis (70% of cases) and those with terminal cancer account for this death rate [2].

Received: 2025.01.16. Revised: 2025.04.29. Accepted: 2025.05.12.
Corresponding author: Ova Emilia, MD, OBGYN, PhD
Department of Obstetrics and Gynecology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia
E-mail: ovaemilia@ugm.ac.id
https://orcid.org/0000-0002-2184-796X

Cervical cancer can be prevented and controlled by screening and human papillomavirus (HPV) vaccination. Effective screening programs, such as Papanicolaou smear, have significantly reduced cervical cancer rates in developed countries [3]. In South Korea, the incidence of cervical cancer is decreasing annually owing to screening and prophylactic vaccination [4]. Cervical cancer remains an extremely important disease in developing countries [5]. However, such programs are not widely available in developing nations. The World Health Organization (WHO) advocates for national programs that use visual inspection with acetic acid (VIA) examinations to detect and manage precancerous lesions, particularly in low-resource countries. Therefore, sexually active women with a positive VIA test result should undergo cryoablation [6]. Meta-analysis showed an VIA sensitivity of 49-98% and specificity of 75-91% [7].

Several undeveloped and developing countries have failed to meet the WHO recommendations for screening 80% of the targeted population for cervical cancer. There is an urgent need for a reliable and convenient detection method for volatile biomarkers of cervical cancer. Volatile organic compounds (VOCs) are present in biological matrices such as urine and blood under various conditions, including cancer [8]. Detection methods such as gas chromatography-mass spectrometry (GC-MS) and electronic noses have been developed. An electronic nose, which uses arrays of chemical sensors, employs artificial intelligence to differentiate and classify chemical vapor samples using multivariate statistical techniques [9].

Persistent HPV infection may cause biological changes in the cervical epithelium, leading to VOC production. In this study, we developed a novel strategy for detecting cervical cancer using VOC sensing. Subsequently, we aimed to quantitatively study volatile biomarkers in the urine samples of women with cervical cancer and healthy controls using an electronic nose.

Materials and methods

This quantitative and experimental case-control study included 200 urine samples from HPV-negative patients. This study was conducted in Indonesia between May and July 2022. The eligibility criteria for the cervical cancer group were as follows: 1) patients with a histopathological diagnosis of cervical cancer; 2) patients who had not undergone or were not undergoing any standard treatment, such as radiation therapy, concurrent chemoradiotherapy, or surgery; 3) patients who had not been cytologically diagnosed according to lesion type (Bethesda classification: low grade squamous intraepithelial lesion or high grade squamous intraepithelial lesion); and 4) patients who signed an informed consent form.

The eligibility criteria for the healthy control group were as follows: 1) women negative for HPV infection and 2) patients who signed an informed consent form. The exclusion criteria were as follows: 1) insufficient sample collection; 2) death during follow-up; and 3) comorbidities (liver disease, diabetes mellitus, hypercholesterolemia, and other malignancies). Our study included cervical cancer patients admitted to a tertiary university hospital. Control participants were recruited through cervical cancer screening programs at primary healthcare centers and underwent a thorough clinical assessment to evaluate potential risk factors, followed by a physical examination using visual inspection with acetic acid. This was complemented by an HPV DNA test to confirm the absence of HPV infection both clinically and at the molecular level. The collected samples reflected the clinical settings and study design. As inpatients, multiple urine samples were collected from cervical cancer participants at various time points during their hospital stay. This strategy was intended to capture temporal variations in urinary VOCs, potentially associated with disease progression, treatment response, and physiological changes. Healthy control participants were outpatients who provided only a single urine sample per individual, as their VOC profiles were presumed to be more stable over time. Samples from both groups were collected using standardized protocols to ensure consistency and reliability of VOC analysis.

Informed consent was obtained from all participants and the study was approved by the Medical and Health Research Ethics Committee. In accordance with the journal's guidelines, we will provide our data for independent analysis by a team selected by the editorial team for the purpose of additional data analysis or for the reproducibility of this study in other centers, if requested.

1. Sample collection

The patients' first-void and midstream urine samples were collected after an 8-hour fast at home. The samples were collected in sterile polypropylene cups (60 mL) and transported to the laboratory at 4°C. The samples were analyzed in the same week as collection, and VOC fingerprints were measured using an electronic nose composed of 10 sensors. A cervical smear was obtained from healthy controls for HPV genotyping.

2. VOC determination in the urine samples

Urine (25 mL) was heated to 70°C for 5 minutes using a thermomagnetic stirrer. This produced a vector signal encoding the VOCs emitted from the urine. The VOC components in the urine samples from healthy controls were compared with those in the urine samples from cervical cancer patients, as determined by GC-MS. The VOCs in urine were detected using GeNose (Python, Amsterdam, Netherlands), which contains 10 chemical nanosensors that specify certain functional VOC groups present in the sample. Resistance changes across the array were captured as a digital pattern representative of the test smell. The overall response to a particular sample ("smell print") was specific to a stimulus. In a study mentioned by Díaz de León-Martínez et al. [10], analysis for the detection of the VOC fingerprint in urine was performed using an electronic nose. The Shapley additive explanations (SHAP) summary plot visually ranks the sensors (S1-S10) based on their individual contributions to the model's predictive power.

The configuration of urine prints for cervical cancer is shown in Fig. 1. The sensor conductivity changes because of redox reactions between the active metal-oxide-semiconductor material and absorbed gas molecules. Real-time signal monitoring of VOC exposure on the sensor surface was performed using a data logging software (Python). The data were analyzed using the ExtraTreesClassifier (Python). A hybrid learning algorithm combining hierarchical agglomerative clustering and permutation feature importance enhances the GeNose performance (Python) and simultaneously reduces the required sensor number.

3. Statistical analysis

Descriptive statistics for the study groups were calculated us-

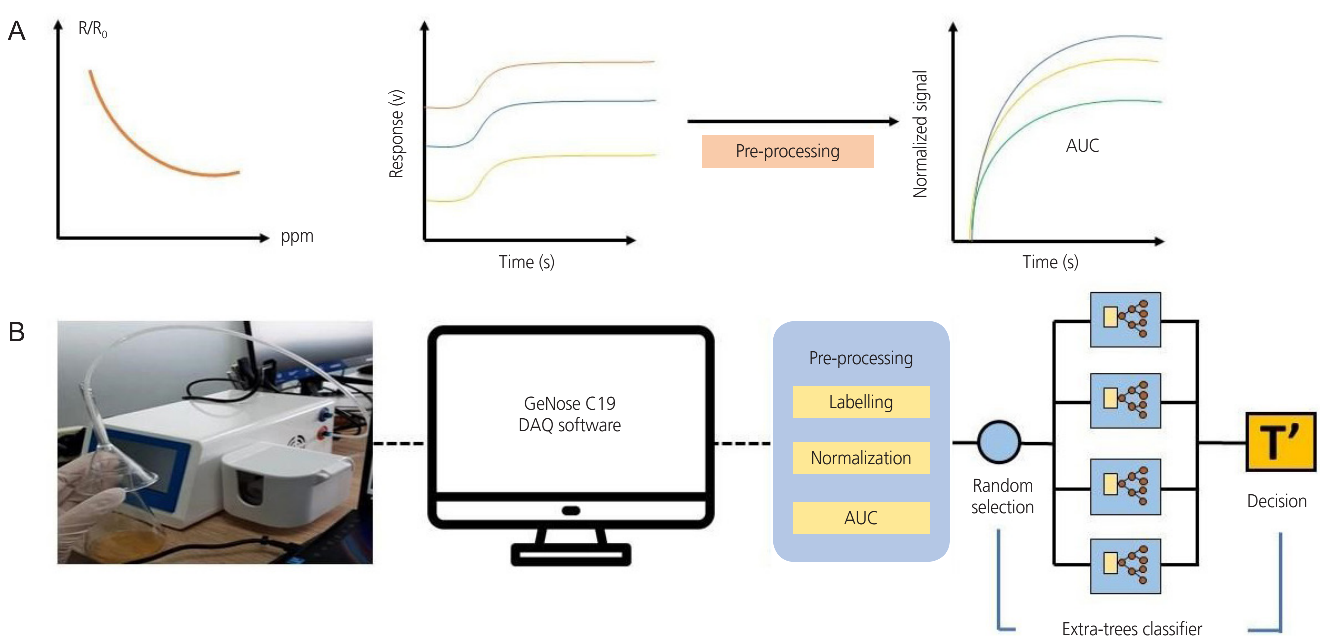


Fig. 1. Configuration of a portable e-nose for cervical cancer detection. (A) Output signal characteristics of chemoresistive metal-oxide-semiconductor (MOS) gas sensors. The sensor conductivity changes because of redox reactions between the active MOS material and adsorbed gas molecules. (B) The real-time signal monitoring regarding VOC exposure to the sensor surface is performed using data logging software (DAQ software [Python, Amsterdam, Netherlands]). Procedure to collect urine samples and process the data utilizing an extra-tress classifier. A hybrid learning algorithm combining hierarchical agglomerative clustering analysis and permutation feature importance method enhances the GeNose performance (Python) and simultaneously reduces the required sensor number. ppm, parts per million; AUC, area under the curve; C19, covid-19; VOC, volatile organic compounds.

ing SPSS version 24 (IBM, Armonk, NY, USA), and the means and standard deviations of the anthropometric parameters were recorded. Principal component analysis (PCA) was performed using a multivariate data cloud to differentiate the predefined groups. Different models were applied to overcome statistical bias. The two methods for preprocessing the data were feature extraction for maximum different values and standardization for the scale data. We used model selection data via train-test splits and classification using XGBoost (Python). This analysis identified correlations among the included variables.

Linear discriminant analysis (LDA) is a supervised method that leverages class labels to identify the linear combination of features that best differentiates between classes. To verify the findings and conduct a more complex classification analysis, an XGBoost model (Python) was employed to differentiate between the negative and positive classes. By leveraging the insights gained from PCA and LDA, the XGBoost model (Python) effectively learned intricate patterns and optimized the classification accuracy. This approach ensured a comprehensive and rigorous analysis, leading to a more robust and reliable classification model for distinguishing between the two classes. The performance of the XGBoost model (Python) was further evaluated using a separate testing dataset to assess its generalizability.

Table 1. Clinical parameters of the study groups

Parameter	Healthy controls	CC patients	*P*-value
Samples	20	20	
Age (yr)	48.5±1.6	37.0±1.6	0.28
Weight (kg)	60.2±13.3	59.95±12.9	0.306
Height (cm)	1.55±0.05	1.55±0.05	0.71
BMI (kg/m^2)	23.19±2.7	25.3±3.25	0.265
Urine samples	95	235	
Parity	2	3	0.447
Comorbid			1
Yes	0 (0.0)	0 (0.0)	
No	20 (100.0)	20 (100.0)	
Histopathology			
SCC		14 (70.0)	
Non-SCC		6 (30.0)	
Grade			
I		2 (10.0)	
II		6 (30.0)	
III		12 (60.0)	
Stage			
IA-IIA		11 (55.0)	
IIB-IV		9 (45.0)	

Values are presented as mean±standard deviation or number (%).
CC, cervical cancer; BMI, body mass index; SCC, squamous cell cancer.

Result

We obtained 95 urine samples from healthy controls and 235 urine samples from cervical cancer patients. The age of the study participants ranged from 20 years to 58 years, as shown in Table 1.

The data were projected onto the first two principal components (PC1 and PC2) to capture the differences among the groups from the PCA plot and the most salient variance within the dataset (Fig. 2). PC1 accounted for 76.83% of the observed variance, while PC2 accounted for 9.22% of the observed variance. This revealed a degree of overlap, suggesting that PCA may not optimally differentiate between these classes. However, a clear separation between the "negative" and "positive" classes was evident along linear discriminant 1 (LD1).

The performance of the XGBoost model (Python) on the training data describes key metrics such as precision, recall, and F1-score for each class, as well as the overall accuracy and weighted averages. The model achieved high precision and recall for both classes, as listed in Table 2.

The model maintained a high level of performance and achieved an overall accuracy of 0.89. The results indicated that the model is effectively generalizable, demonstrating its robustness and reliability in real-world scenarios. Although the precision and recall for the "negative" class using the test data were slightly lower than those using the training data, they remained respectable, indicating that the model can still effectively identify the majority of negative instances. The continued strong performance for the "positive" class further emphasizes the model's ability to accurately classify positive instances.

Fig. 3A, B show the GC-MS analysis of two samples from the healthy group, which revealed that carbon monoxide and nitrogen compounds were the predominant components detected at a retention time of two seconds. Fig. 3C, D show

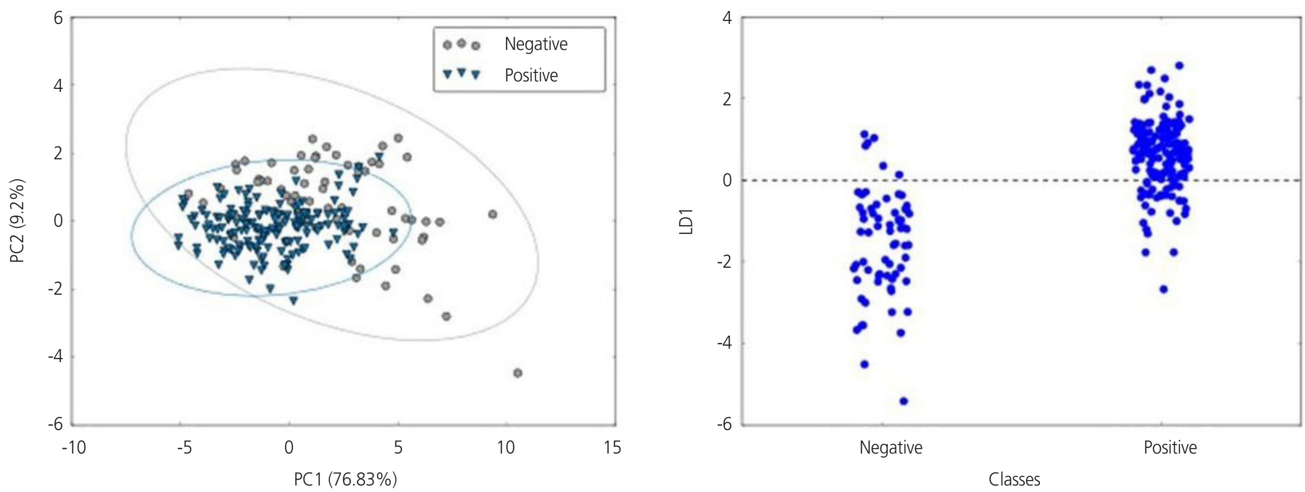


Fig. 2. Principal component analysis and linear discriminant analysis. PC, principal component; LD1, near discriminant 1.

Table 2. Performance of the XGBoost model (Python, Amsterdam, Netherlands) was further evaluated on a separate testing dataset, providing a crucial assessment of its generalization ability

	Precision	Recall	F1-score	Support
Negative	0.79	0.85	0.81	13
Positive	0.94	0.91	0.92	33
Accuracy			0.89	46
Macro avg	0.86	0.88	0.87	46
Weighted avg	0.89	0.89	0.89	46

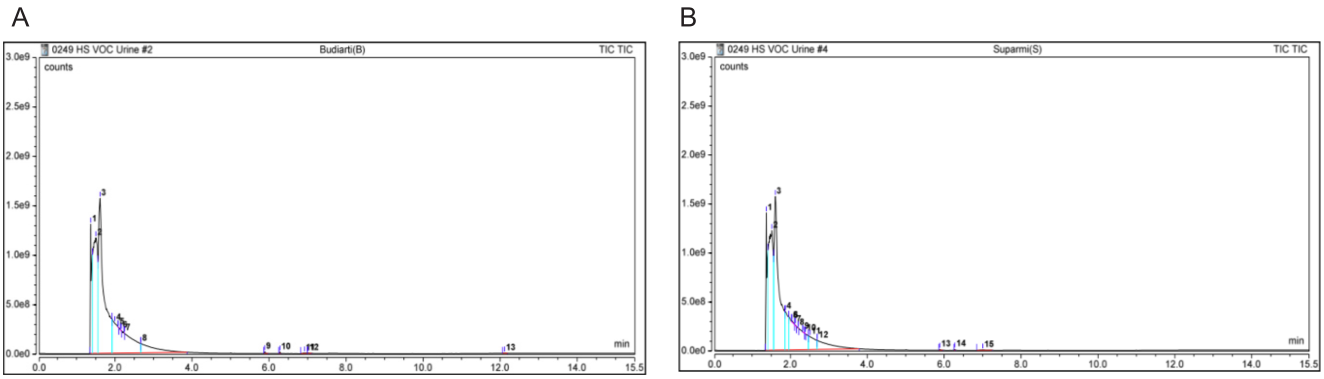
the components identified in two samples obtained from the asymptomatic individuals. GC-MS identified carbon monoxide, nitrogen, and nickel tetracarbonyl in the first, second, and third hits, respectively.

The GC-MS analysis of two samples from cervical cancer patients revealed the presence of carbon monoxide, nitrogen, and nickel tetracarbonyl over a retention period of two seconds (Fig. 4A, B). Table 3 summarises identical and varying VOCs between groups. The GC-MS analysis of urine samples from two other patients diagnosed with cervical cancer revealed the presence of carbon monoxide, nitrogen, and trisilaxane (Fig. 4C, D). Comprehensive overview of global feature importance in the XGBoost model (Python) to validate the LDA results. The SHAP plot revealed the dominance of sensors S3, S2, and S10. Conversely, sensors S4 and S6 demonstrated relatively low SHAP values, implying a less significant role in the decision-making process of the model.

Fig. 5A, which consists of 5X, shows the force plot of the impact of individual sensors on the XGBoost model's (Python) prediction and Fig. 5 presents a compelling visualization of the complex interplay between individual sensor responses and the XGBoost model's (Python) prediction for both negative and positive classes. Fig. 5 shows that S1 exerted a strong negative influence, pushing the prediction to be significantly lower. Conversely, sensors S3 and S2 counteracted this effect, driving the prediction toward the positive class. Despite their opposing influences, the combined effect of all the sensors resulted in a final prediction ($f(x)=-1.56$), which remained within the negative range.

The bottom plot, representing a positive sample, reveals a different pattern of sensor contributions. Sensor S10 was the primary driver of positive classification and exerted a strong positive influence. S2 also contributed positively, albeit to a lesser extent. In addition, S3 exerted a slightly positive influence. This contrasting behavior highlights the complex relationships between the sensor responses and sample classification.

Additionally, Fig. 5B, using boxplot analysis, further highlights these findings by providing the distribution and variability of the sensor responses, aiding in the identification of sensors with higher discriminatory power. The boxplot distribution of the sensor responses for the positive and negative samples illustrates the maximum response increase from the baseline for each sensor (S1-S10) in the electronic nose array. This analysis displays the range and variability of the maximum response increase for each sensor. By comparing the positive and negative samples, distinct patterns emerged across the different sensors. Notably, sensors S2, S3, and S10



C

No	Ret.Time (min)	Hit#1	Chemical Formula	Mol Weight	SI Hit#1	Hit#2	Chemical Formula	Mol Weight	SI Hit#2 min	Hit#3 counts*min	Chemical Formula	Mol Weight	SI Hit#3 counts*min
1	1.37	Nitrogen	N_2	28	801	Carbon monoxide	CO	28	761	Nichel tetacarbonyl	C_4NiO_4	170	602
2	1.50	Carbon monoxide	CO	28	718	Nitrogen	N_2	28	660	Nichel tetacarbonyl	C_4NiO_4	170	635
3	1.61	Carbon monoxide	CO	28	758	Nitrogen	N_2	28	732	Nichel tetacarbonyl	C_4NiO_4	170	649
4	2.00	Carbon monoxide	CO	28	818	Nitrogen	N_2	28	739	Nichel tetacarbonyl	C_4NiO_4	170	671
5	2.08	Nitrogen	N_2	28	795	Carbon monoxide	CO	28	755	Nichel tetacarbonyl	C_4NiO_4	170	653
6	2.17	Nitrogen	N_2	28	797	Carbon monoxide	CO	28	757	Nichel tetacarbonyl	C_4NiO_4	170	657
7	2.24	Nitrogen	N_2	28	792	Carbon monoxide	CO	28	753	Nichel tetacarbonyl	C_4NiO_4	170	633
8	2.67	Nitrogen	N_2	28	791	Carbon monoxide	CO	28	751	Nichel tetacarbonyl	C_4NiO_4	170	677
9	5.89	Fenchone	$C_{10}H_{16}O$	152	865	L-Fenchone	$C_{10}H_{16}O$	152	848	D-fenchone	$C_{10}H_{16}O$	152	804
10	6.28	Cyclopentasiloxane, decamethyl-	$C_{10}H_{30}O_5Si_5$	370	859	2,6-Dihydroxybenzoic	[illegible]	370	735	3,4-Dihydroxymandelic	[illegible]	428	690
11	6.92	Estragole	$C_{10}H_{12}O$	148	772	Benzene, 1-methoxy-4	$C_{10}H_{12}O$	148	753	Anethole	$C_{10}H_{12}O$	148	746
12	7.01	Benzene, 1-methoxy-4-(1-propenyl)-,(Z)-	$C_{10}H_{12}O$	148	668	Estragole	$C_{10}H_{12}O$	148	667	2H-Pyran, 2-(7-heptade	[illegible]	336	661
13	12.11	9,10-Secocholesta-5,7,10(19)-triene-3,24,25-triol,	$C_{27}H_{44}O_3$	416	670	Ethyl	[illegible]	436	661	Ergosta-5,22-dien-3-ol	[illegible]	440	652

D

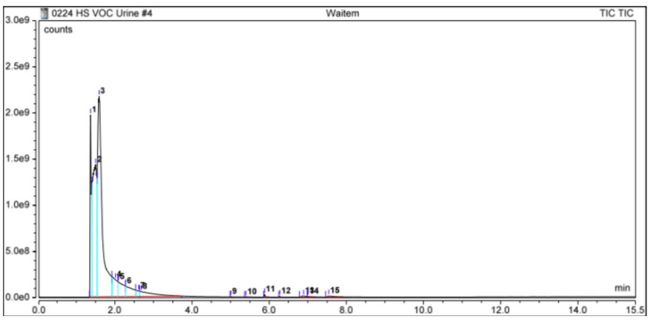
No	Ret.Time (min)	Hit#1	Chemical Formula	Mol Weight	SI Hit#1	Hit#2	Chemical Formula	Mol Weight	SI Hit#2 min	Hit#3 counts*min	Chemical Formula	Mol Weight	SI Hit#3 counts*min
1	1.37	Nitrogen	N_2	28	803	Carbon monoxide	CO	28	763	Nichel tetacarbonyl	C_4NiO_4	170	594
2	1.51	Carbon monoxide	CO	28	713	Nitrogen	N_2	28	657	Nichel tetacarbonyl	C_4NiO_4	170	628
3	1.61	Carbon monoxide	CO	28	773	Nitrogen	N_2	28	719	Nichel tetacarbonyl	C_4NiO_4	170	649
4	1.86	Carbon monoxide	CO	28	744	Nitrogen	N_2	28	682	Nichel tetacarbonyl	C_4NiO_4	170	647
5	2.03	Nitrogen	N_2	28	320	Carbon monoxide	CO	28	741	Nichel tetacarbonyl	C_4NiO_4	170	650
6	2.04	Nitrogen	N_2	28	325	Carbon monoxide	CO	28	746	Nichel tetacarbonyl	C_4NiO_4	170	650
7	2.13	Nitrogen	N_2	28	801	Carbon monoxide	CO	28	761	Nichel tetacarbonyl	C_4NiO_4	170	655
8	2.20	Nitrogen	N_2	28	323	Carbon monoxide	CO	28	744	Nichel tetacarbonyl	C_4NiO_4	170	654
9	2.33	Carbon monoxide	CO	28	320	Nitrogen	N_2	28	741	Nichel tetacarbonyl	C_4NiO_4	170	678
10	2.38	Carbon monoxide	CO	28	319	Nitrogen	N_2	28	740	Nichel tetacarbonyl	C_4NiO_4	170	680
11	2.51	Nitrogen	N_2	28	800	Carbon monoxide	CO	28	760	Nichel tetacarbonyl	C_4NiO_4	170	669
12	2.70	Nitrogen	N_2	28	798	Carbon monoxide	CO	28	759	Nichel tetacarbonyl	C_4NiO_4	170	682
13	5.89	Fenchone	$C_{10}H_{16}O$	152	803	L-Fenchone	$C_{10}H_{16}O$	152	789	D-fenchone	$C_{10}H_{16}O$	152	751
14	6.28	Cyclopentasiloxane, decamethyl	[illegible]	370	869	2,6-Dihyroxybenzoic	[illegible]	370	743	3,4-Dihydroxymandelic	[illegible]	428	703
15	7.01	Benzene, 1-methoxy-4-(1-propenyl)-,(Z)-	$C_{10}H_{12}O$	148	680	Estragole	$C_{10}H_{12}O$	148	673	1,2,3,4-Tetrahyroisoqui	$C_{10}H_{11}NO_3$	193	670

Fig. 3. GC-MS analysis comparing two samples from healthy patients. (A, B) The gas chromatography-mass spectrometry (GCMS) analysis of two samples from healthy patients revealed that carbon monoxide and nitrogen compounds were the predominant components detected at a retention duration of two seconds. (C, D) Components were identified in two urine samples obtained from asymptomatic individuals. The GCMS analysis detected the following compounds: carbon monoxide, nitrogen, trisilaxane, 3 acetocy-p-methan-1-ol, nickel tetracarbonyl, fenchone, cyclopentosiloxane, and dematehyl. The chemical compound is (Z)-1-methoxy-4-(1-propenyl) benzene, also known as estragole, with the molecular formula $C_{10}H_{12}O$. The compound-triene-3,24,25-triol was identified first, followed by the compounds carbon monoxide, nitrogen, ethyl iso-alocholate, 2,6-dihydroxibenzoic acid, benzene, 1-methoxy-4-(1-propenyl)-, L-fenchone, and estragole in the second hit. Finally, nickel tetracarbonyl, D-fenchone, 3,4-dhihydroxymandelic acid, 1,2,3,4-tetrahydroisoquinone, anethole, 2H-pyran, and ergosta were identified in the third hit. GC-MS analysis comparing two samples from healthy patients. HS, head space gas chromatography; VOC, volatile organic compound; TIC, total ion chromatogram; GC-MS, gas chromatography-mass spectrometry.

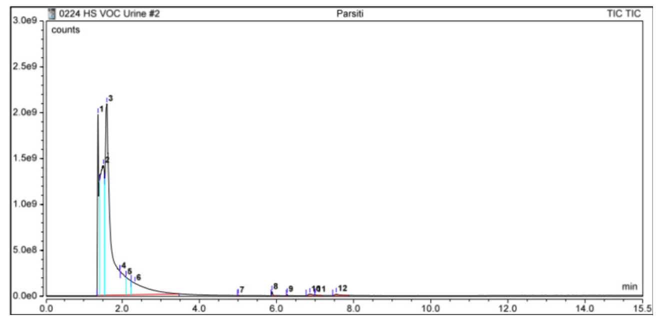
exhibited a clear separation between the median responses of the positive and negative samples. Conversely, sensors S4, S5, and S6 showed substantial overlap in their interquartile ranges, indicating a lower discriminatory power.

The presence of outliers, depicted as individual points beyond whiskers, highlighted the variability within each sample group. Furthermore, the varying lengths of the boxes and whiskers across different sensors emphasized the differences in the dynamic ranges and sensitivities of the sensors.

A



B



C

No	Ret. Time (min)	Hit#1	Chemical Formula	Mol Weight	SI Hit#1	Hit#2	Chemical Formula	Mol Weight	SI Hit#2 min	Hit#3 counts*min	Chemical Formula	Mol Weight	SI Hit#3 counts*min
1	1.37	Carbon monoxide	CO	28	756	Nitrogen	CO	28	679	Nichel tetacarbonyl	C_4NiO_4	170	565
2	1.51	Carbon monoxide	CO	28	645	Nitrogen	N_2	28	597	Nichel tetacarbonyl	C_4NiO_4	170	595
3	1.60	Nickel tetracarbonyl	C_4NiO_4	17	612	Carbon monoxide	N_2	28	571	Nitrogen	N_2	170	526
4	1.94	Nitrogen	N_2	28	782	Carbon monoxide	N_2	28	742	Nichel tetacarbonyl	C_4NiO_4	170	687
5	2.30	Nitrogen	N_2	28	784	Carbon monoxide	CO	28	745	Nichel tetacarbonyl	C_4NiO_4	170	655
6	2.33	Nitrogen	N_2	28	787	Carbon monoxide	CO	28	747	Nichel tetacarbonyl	C_4NiO_4	170	743
7	5.01	trisiloxane, 1,1,1,5,5,5-hexamethyl-3-[(trimethylsilyl)oxy]	C_6H_{28}O35i 4	296	834	Cyclotetrasiloxane	$C_8H_{24}O_4Si_4$	296	761	Nichel 1-[2,4-Bistrimethylsilox	$C_{24}H_{18}$O4 Si3	170	720
8	5.88	Fenchone	$C_{10}H_{16}O$	152	917	L-Fenchone	$C_{10}H_{16}O$	152	915	D-Fenchone	$C_{10}H_{16}O$	170	871
9	6.28	Cyclopentasiloxane, decamethyl-	$C_{10}H_{30}$O5 5i5	370	865	2,6-Dihydroxy beenxoic acid	$C_{14}H_{22}O_4Si_3$	370	737	Isoproterenol	$C_{18}H_{41N}$O3 Si3	170	700
10	6.86	Estragole	$C_{10}H_{12}O$	148	874	Benzen-1-methoxy-4	$C_{10}H_{12}O$	148	864	Anethole	$C_{10}H_{12}O$	170	853
11	7.01	Benzene, 1-methoxy-4-(1-propenyl)-,(Z)-	$C_{10}H_{12}O$	148	670	Bicyclo[4,5,6]nonane	$C_{15}H_{24}$	204	660	Spiro[adamantane-2,5']	$C_{13}H_{16}$O3	170	656
12	7.55	Anethole	$C_{10}H_{12}O$	148	682	Benzen-1-methoxy-4	$C_{10}H_{12}O$	148	674	Estragole	$C_{10}H_{12}O$	170	570

D

No	Ret.Time (min)	Hit#1	Chemical Formula	Mol Weight	SI Hit#1	Hit#2	Chemical Formula	Mol Weight	SI Hit#2 min	Hit#3 counts*min	Chemical Formula	Mol Weight	SI Hit#3 counts*min
1	1.37	Carbon monoxide	CO	28	730	Nitrogen	N_2	28	654	Nichel tetacarbonyl	C_4NiO_4	170	587
2	1.50	Carbon monoxide	CO	28	677	Nitrogen	N_2	28	628	Nichel tetacarbonyl	C_4NiO_4	170	612
3	1.59	Carbon monoxide	CO	28	629	Nichel tetacarbonyl	$C_4N_1O_4$	170	586	Nitrogen	C_4NiO_4	28	570
4	2.01	Nitrogen	N_2	28	782	Carbon monoxide	CO	28	742	Nichel tetacarbonyl	C_4NiO_4	170	650
5	2.09	Nitrogen	N_2	28	782	Carbon monoxide	CO	28	742	Nichel tetacarbonyl	C_4NiO_4	170	661
6	2.28	Nitrogen	N_2	28	776	Carbon monoxide	CO	28	736	Nichel tetacarbonyl	C_4NiO_4	170	663
7	2.62	Nitrogen	N_2	28	782	Carbon monoxide	CO	28	742	Nichel tetacarbonyl	C_4NiO_4	170	677
8	2.68	Nitrogen	N_2	28	765	Carbon monoxide	CO	28	760	Phenetylamine	$C_{10}H_{15}N$	149	707
9	5.01	trisiloxane, 1,1,1,5,5,5-hexamethyl-3-[(trimethylsilyl)oxy]	C_6H_{28}O3 5i4	296	853	Cyclotetrasiloxane	$C_8H_{24}O_4Si_4$	296	769	3-Ethoxy-1,1,5,5,5	$C_{11}H_{32}$O4 Si4	340	734
10	5.40	3-acetoxy-p-menthan-1-ol	$C_{12}H_{122}$O 3	214	622	Bicyclo[4,5,6]nonane	$C_{10}H_{18}O$	154	613	Bicyclo[2,2,1]heptane-2	$C_{11}H_{19}$N3 O	209	613
11	5.88	Fenthone	$C_{10}H_{16}O$	152	910	L-Fenchone	$C_{10}H_{16}O$	152	888	D-Fenchone	$C_{10}H_{16}O$	152	848
12	6.28	Cyclopentasiloxane, decamethyl-	$C_{10}H_{30}$O3 5Si5	370	831	2,6-Dihydroxybenzoic acid	$C_{10}H_{16}O$	370	712	3,4-Dihydroxymandelic	$C_{18}H_{36}$O5 Si3	428	681
13	6.89	Estragole	$C_{10}H_{12}O$	148	845	Benzene, 1-methoxy-4	$C_{16}H_{18}$O4 Si3	148	832	Anethole	$C_{10}H_{12}O$	148	827
14	7.00	Benzene, 1-methoxy-4-(1-propenyl)-,(Z)-	$C_{10}H_{12}O$	148	692	Estragole	$C_{10}H_{12}O$	148	690	9-Methoxy bycido[6,1,0]	$C_{10}H_{12}O$	148	671
15	7.54	Anethole-	$C_{10}H_{12}O$	148	682	Estragole	$C_{10}H_{12}O$	148	679	Benzene, 1-methoxy-4	$C_{10}H_{12}O$	148	673

Fig. 4. GC-MS analysis of cervical cancer patients. (A, B) Analysis of two samples from cervical cancer patients revealed a prominent presence of carbon monoxide, nitrogen, and nickel tetracarbonyl components over a retention time of two seconds, as indicated by the GCMS pattern. (C, D) The GCMS analysis detected the following compounds: carbon monoxide, nitrogen, trisilaxane, 3 acetocy-p-methan-1-ol, nickel tetracarbonyl, fenchone, cyclopentosiloxane, and dematehyl-. The following compounds were identified in the first hit: estragole, benzene, 1-methoxy-4-(1-propenyl)-, (Z)-, and anethole. Additionally, carbon monoxide compounds, nitrogen, cyclo tetrasiloxane, 3 acetocy-p-methan-1-ol, nickel tetracarbonyl, bicydo (2,2,1) heptan-2-OH, 2,6-dihydroxibenzoic acid, benzene, 1-methoxy-4-(1-propenyl)-, and L-fenchone were also detected. The following compounds were identified in the second hit: estragole, nickel tetracarbonyl, nitrogen, phenethylamine, 3-ethoxy-1,1,1,5,5,5-hexane, bicyclo (2,2,1) heptan-2-OH, D-fenchone, 3,4-dihydroxymandelic acid, anethole, 9-methoxybicyclo, benzene, 1-methoxy-4-(1-propenyl), and 1-(2,4,-bis(trimethylsiloxane). Additionally, isoproterenol and spiro were identified in the third hit. HS, head space gas chromatography; VOC, volatile organic compound; GC-MS, gas chromatography-mass spectrometry; GCMS, gas chromatography-mass spectrometry.

Table 3. VOC urine components in healthy patients vs. cervical cancer patients

No	Identical elements	Varying elements
1	Carbon monoxide	9-methoxybicydo
2	Nitrogen	Phenetylamine
3	Fenchone	9,10-secocholesta-5,7,10(19)-triene-3,24,25-triol
4	Estragole	Ethyl iso-alocholate
5	Benzene, 1-methoxy-4-(1-propenyl;)-, (Z)	3-ethoxy-1,1,1,5,5,5-hexane
6	Nickel tetracarbonyl	1-(2,4,-bis (trimethylsiloxane)
7	Anethole	1,2,3,4-tetrahydroisoquinone
8	Trisilaxane, 3 asetocy-p-methan-1-ol	2H-pyran
9	Cyclopentosiloxane, dematehyl	Ergosta
10	2,6-dihydroxibenzoic acid	Isoproterenol
11	Bicydo (2,2,1) heptan-2-OH	Spiro
12	3,4-dhihydroxymandelic acid	

VOC, volatile organic compound.

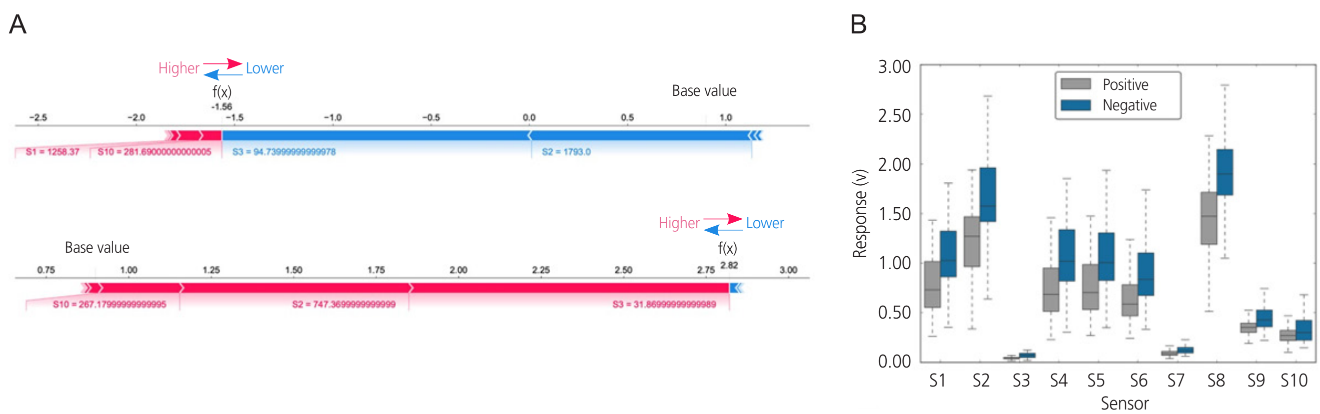


Fig. 5. (A) SHAP force plot showing the impact of individual sensors on the XGBoost model's (Python, Amsterdam, Netherlands) prediction. Red/blue features push the prediction higher/lower (towards positive/negative classification). The top plot shows a negative sample, the bottom a positive sample. The base value is the average prediction. f(x) is the model's prediction for that sample. (B) Boxplot distribution of sensor responses for positive and negative samples. The plot displays the maximum response increase from the baseline for each sensor (S1-S10) in the electronic nose array. The grey boxes represent positive samples, while the blue boxes represent negative samples. The central line within each box indicates the median response, the box edges represent the interquartile range (IQR), and the whiskers extend to 1.5 times the IQR. Outliers beyond this range are plotted individually. SHAP, shapley additive explanations.

Discussion

1. Summary of the main results

Our results indicate the applicability of VOC sensing for cervical cancer detection, as well as its potential application in treatment monitoring. In this study, an electronic nose was used to evaluate the impact of 10 sensors on the separation of patients into groups and the differences in VOCs between the control and cervical cancer groups. Urinary metabolites are the end products of a wide variety of metabolic activities and are used to diagnose various illnesses.

We highlighted the performance of PC1 and LD1 to discriminate positive and negative instances. Hereby, the XGBOOST system (Python) finalized all the results and validate the results coming from LD1 and PC1. This robust performance demonstrated the model's generalizability and accurate differentiation of the negative and positive classes in real-world applications. The XGBoost model (Python) effectively differentiated between the negative and positive classes, with the highest mean absolute SHAP values obtained from the sensors. These sensors capture critical chemical information and provide valuable insights into the overall performance of the electronic nose system. For instance, it may inform sensor selection and optimization strategies to enhance the sensitivity and specificity of the electronic nose for target applications. By identifying intricate relationships, the SHAP force plots provide valuable insights for interpreting model predictions and understanding the underlying mechanisms driving sample classification. This analysis can guide future studies and development efforts by informing sensor selection and optimization strategies to enhance the sensitivity and specificity of the electronic nose for target applications.

The non-invasive nature of the test also contributes to the high acceptability of HPV testing among females. The WHO proposed that the development of screening methodologies for HPV-associated cervical cancer should focus on these aspects because cost-effective screening tests are crucial for boosting programs. Thus, our exploratory study is relevant because of its ease of analysis, accessibility to cervical cancer screening, and non-invasive method of sample collection.

2. Results in the context of the published literature

In a study by Rajasekaran et al. [11], in which steady-state and time-resolved fluorescence spectroscopy was used, cervical cancer patients (n=60) could not be differentiated from healthy controls (n=60) by evaluating the biomarkers indoxyl sulfate, neopterin, and riboflavin. The rate of correct classification was 96.4% [11]. Furthermore, they demonstrated low specificity of the discovered compounds, which have been associated with various diseases. For example, indoxyl sulfate is a biomarker of kidney disease. Changes in urine metabolic patterns may be associated with changes in the vaginal microbiota due to HPV infection, which releases VOCs into the urine. This explains the differences in urine samples between women with HPV and women who are controls.

In the present study, the sensitivity and specificity of the tools for detecting cervical cancer using the VOC assay were 91% and 85%, respectively. A previous study from Portugal demonstrated that VOCs exhibit a high sensitivity (89%) and specificity (83%) for prostate cancer detection [12]. A study from the USA Gao et al. [13] demonstrated that urinary VOCs exhibit a sensitivity and specificity of 87% and 77%, respectively, for prostate cancer detection. Díaz de León-Martínez et al. [10] evaluated the global chemical pattern of VOCs in urine for the detection of cervical cancer. The sensitivity and specificity of the test in their study were 91.6% (range, 61.5-99.7%) and 100% (range, 73.5-100%), respectively [10].

Elia et al. [14] conducted a study using a GC-MS approach to identify certain VOCs in urine for the detection of HPV and cervical cancer. The study included 17 patients with cervical intraepithelial neoplasia one and nine healthy women. The detected VOCs differed between the two groups. These substances were consistent with those identified in the present study (ethanol, hexane, isobutane, and methane). Analysis of the 10 sensors in our study revealed distinct patterns between the healthy and cervical cancer groups. Specifically, sensors S3, S2, and S10 exhibited a tendency to have a positive predictive impact. Furthermore, S5 had the greatest impact on the prediction of the post-therapy data. However, sensor S3 had the greatest impact on predicting pre-therapy data. The metabolites shared between the two groups were ethanol, hydrogen, isobutane, and methane.

3. Strengths and weaknesses

This study represents a preliminary investigation of the utilization of electronic nose technology in Indonesia to demonstrate the potential of urinary VOCs as biomarkers for the detection of cervical cancer. The sensitivity of this method can be used to differentiate VOC patterns between cervical cancer patients and healthy controls. This analysis is rapid, cost-effective, and non-invasive. Furthermore, urine samples exhibit high potential for identifying biomarkers indicative of physiological conditions and generating high concentrations of VOCs that can serve as biomarkers for other cancers.

A limitation of this study was the small sample size and inability to compare the sensitivity and specificity of the test against standard benchmarks. Thus, future studies should include an expanded study population and post-therapy patients should be validated via routine cytological examinations or HPV testing.

4. Implications for practice and future studies

The identification of VOCs as a novel modality for cervical cancer detection holds significant promise in clinical practice and future studies. The integration of VOC analysis could enhance the sensitivity and specificity of cervical cancer screening, potentially leading to earlier detection and improved patient outcomes. This non-invasive approach presents significant potential advantages, especially in resource-limited settings where screening methods are less accessible.

Conflicts of interest

The authors declare no conflicts of interest.

Ethical approval

This study was approved by the Medical and Health Research Ethics Committee of the Faculty of Medicine, Public Health, and Nursing (reference number: KE/FK/0509/EC/2022).

Patient consent

Prior to participation, the patient received a detailed explanation of the study's objectives, procedures, potential risks, and benefits. The patient was given the opportunity to ask questions and subsequently provided written informed consent. By signing the consent form, the patient confirmed voluntary participation as a subject in this research.

Funding information

This work was supported by Universitas Gadjah Mada with grant number (10529/UN1/P4/PT.01.03).

References

1. World Health Organization. WHO indonesian cancer data 2018 [Internet]. Geneva: World Health Organization; c2019 [cited 2024 Dec 12]. Available from: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://cdn.who.int/media/docs/default-source/country-profiles/cancer/idn-2020.pdf?sfvrsn=46ea6569_2%26download=true.
2. Center for data and information of the Ministry of Health of the Republic of Indonesia [Internet]. Jakarta: The National Institute of Health Research and Development; c2018 [cited 2024 Dec 12]. Available from: https://repository.badankebijakan.kemkes.go.id/id/eprint/3514/1/Laporan%20Riskesdas%202018%20Nasional.pdf.
3. Silkensen SL, Schiffman M, Sahasrabuddhe V, Flanigan JS. Is it time to move beyond visual inspection with acetic acid for cervical cancer screening? Glob Health Sci Pract 2018;6:242-6.
4. Park SM, Lee JW, Lee Y, Myong JP, Chang HK, Min KJ, et al. Factors associated with cervical cancer screening behaviors among young married female (aged 20-29) immigrants in South Korea. Obstet Gynecol Sci 2023;66:26-33.
5. Wongpratate M, Bumrungthai S. Cervical cancer in Thailand: 2023 update. Obstet Gynecol Sci 2024;67:261-9.
6. Bedell SL, Goldstein LS, Goldstein AR, Goldstein AT. Cervical cancer screening: past, present, and future. Sex Med Rev 2020;8:28-37.
7. Crisp WE, Craine BL, Craine EA. The computerized digital imaging colposcope: future directions. Am J Obstet Gynecol 1990;162:1491-7; discussion 1497-8.
8. Rodríguez-Aguilar M, Díaz de León-Martínez L, Gorocica-Rosete P, Pérez-Padilla R, Domínguez-Reyes CA, Tenorio-Torres JA, et al. Application of chemoresistive gas sensors and chemometric analysis to differentiate the fingerprints of global volatile organic compounds from diseases. Preliminary results of COPD, lung cancer and breast cancer. Clin Chim Acta 2021;518:83-92.
9. Johnson G, Patel D, Alluhayb A, Li N, Shen C, Webster T. Preliminary studies of honey queen bee conditions using Cyranose 320 nose technology. ACM 2018;30:1-6.
10. Díaz de León-Martínez L, Flores-Ramírez R, López-Mendoza CM, Rodríguez-Aguilar M, Metha G, Zúñiga-Martínez L, et al. Identification of volatile organic compounds in the urine of patients with cervical cancer. Test concept for timely screening. Clin Chim Acta 2021;522:132-40.
11. Rajasekaran R, Aruna PR, Koteeswaran D, Bharanidharan G, Baludavid M, Ganesan S. Steady-state and timeresolved fluorescence spectroscopic characterization of urine of healthy subjects and cervical cancer patients. J Biomed Opt 2014;19:37003.
12. Lima AR, Pinto J, Azevedo AI, Barros-Silva D, Jerónimo C, Henrique R, et al. Identification of a biomarker panel for improvement of prostate cancer diagnosis by volatile metabolic profiling of urine. Br J Cancer 2019;121:857-68.

13. Gao Q, Su X, Annabi MH, Schreiter BR, Prince T, Ackerman A, et al. Application of urinary volatile organic compounds (VOCs) for the diagnosis of prostate cancer. Clin Genitourin Cancer 2019;17:183-90.
14. Elia P, Raizelman S, Katorza E, Matana Y, Zeiri O, Boger Z, et al. Biomarkers for the detection of pre-cancerous stage of cervical dysplasia. J Mol Biomark Diagn 2015; 6:255.