RESEARCH ARTICLE

Editorial Process: Submission:02/01/2026 Acceptance:18/08/2026 Published:21/08/2026

Clinical Evaluation of Urinary VOCs as Non-Invasive Biomarkers in Prostate Cancer

Chatiwat Piyarom[1]*, Settawut Chalermwat[2], Sekdusit Aekgawong[1], Dusit Kongnawakun[3], Watcharapong Anakkamatee[4], Anawin Pechbooranin[5]

Abstract

Objective: Prostate cancer is one of the most common malignancies in men, and current diagnostic tools, such as prostate-specific antigen (PSA) and biopsy remain limited by low specificity and invasiveness. Urinary volatile organic compounds (VOCs) are increasingly recognized as promising non-invasive biomarkers reflecting tumor-associated metabolic changes. This study aimed to evaluate the diagnostic potential of urinary VOCs using a semiconductor-based gas-sensor platform. **Materials:** Urine samples from patients with prostate cancer and non-cancer controls were analyzed using an array of five metal oxide semiconductor sensors (S1–S5), each tuned to specific VOC groups. Measurements were performed under six thermal cycles with sequential heater voltages (2000–5000 mV). Sensor signals were converted into characteristic parameters following a validated extraction protocol. Logistic regression was used to determine discriminatory performance. **Results:** Ten candidate VOC parameters were initially identified with significant intergroup separation. In the multivariate model, the resistance gap at 2500 mV targeting trimethylamine and methyl mercaptan remained independently associated with prostate cancer. When combined with PSA in the logistic regression model, this VOC-derived signal yielded an AUC of 0.88, with a 77% sensitivity and 88% specificity for distinguishing cancer from controls. **Conclusion:** Urinary VOCs demonstrate strong potential as non-invasive biomarkers for prostate cancer detection. The diagnostic performance of the combined VOC–PSA model (AUC 0.88) exceeded that of PSA alone, underscoring the feasibility of gas-sensor platforms in urologic cancer diagnostics and supporting validation in larger, prospective cohorts.

Keywords: Prostate cancer, Volatile organic compounds (VOCs), Metal oxide biosensor, Urinary biomarkers

Asian Pac J Cancer Prev, **27 (8)**, 3045-3053

Introduction

Prostate cancer (PCa) is one of the most common malignancies in men and a leading contributor to cancer-related morbidity and mortality worldwide [1]. Incidence continues to rise with aging populations, and PCa accounts for a substantial clinical and socioeconomic burden. Early detection remains critical, as curative interventions are most effective when disease is confined to the prostate, whereas advanced disease has limited therapeutic options and is associated with poor prognosis.

Current diagnostic strategies rely on a combination of clinical assessment, prostate-specific antigen (PSA) testing, magnetic resonance imaging (MRI), and histopathological confirmation by prostate biopsy. While PSA has been the cornerstone of early detection for decades, it lacks specificity: benign conditions such as prostatitis and benign prostatic hyperplasia (BPH) can cause elevated PSA levels, leading to unnecessary biopsies [2-4]. Conversely, clinically significant cancers can occur at PSA levels traditionally considered "normal," resulting in missed diagnoses. Biopsy remains the diagnostic gold standard but has limitations, including sampling error, with false-negative rates reported up to 20%, as well as risks of bleeding, infection, and patient reluctance [5]. Multiparametric MRI has improved localization and risk stratification, but it remains resource-intensive and is not universally accessible. These challenges highlight the unmet need for a noninvasive, cost-effective, and accurate adjunct or alternative to current pathways.

In recent years, attention has turned toward noninvasive biomarkers derived from biofluids. Advances in metabolomics have highlighted volatile organic compounds (VOCs) low-molecular-weight metabolites released in urine, breath, and other excretions—as potential indicators of underlying pathophysiology [6, 7]. VOCs

*[1]School of Surgery, Institute of Medicine, Suranaree University of Technology, Thailand. [2]School of Internal Medicine, Institute of Medicine, Suranaree University of Technology, Thailand. [3]School of Pathology, Institute of Medicine, Suranaree University of Technology, Thailand. [4]Department of Mathematics, Faculty of Science, Naresuan University, Phitsanulok, Thailand. [5]School of Mechatronics Engineering, Institute of Engineering, Suranaree University of Technology, Thailand. *For Correspondence: chatiwat@sut.ac.th*

arise from altered cellular metabolism, oxidative stress, microbial activity, and host–tumor interactions, thereby capturing a biochemical 'odor fingerprint' characteristic of disease [8, 9]. Previous work has reported specific urinary VOCs associated with urologic cancers, including methane, iso-butane, hydrogen, ethanol, ammonia, and methyl mercaptan [10, 11]. Importantly, VOCs can be sampled noninvasively, offering a promising avenue for population-level screening or triage.

Biosensing platforms capable of capturing these VOC signatures are evolving rapidly. Semiconductor-based chemiresistive sensors, in particular, offer advantages of low cost, portability, and rapid measurement, distinguishing them from complex laboratory methods such as gas chromatography–mass spectrometry (GC–MS). While GC–MS provides precise compound-level identification, it is impractical for large-scale clinical deployment. Conversely, semiconductor sensors can be deployed at the point of care and generate complex dynamic response patterns when exposed to VOC mixtures [12-14]. However, many prior studies have applied dimensionality-reduction approaches, such as principal component analysis, which emphasize overall clustering patterns but limit the interpretability of individual signal features. This can obscure mechanistic insights and hinder clinical translation.

The present study builds on these foundations by focusing on engineered sensor features derived from thermal cycle responses, including minimum resistance values, latency times, slope dynamics, and inter-cycle resistance gaps. These parameters capture both the intensity and temporal characteristics of VOC interactions with sensor surfaces and may more directly reflect pathophysiological alterations in prostate cancer [10, 11]. By interrogating interpretable VOC-derived features rather than relying solely on pattern recognition, we aim to bridge the gap between statistical discrimination and clinically meaningful signals.

To address this gap, we focused on direct analysis of signal-derived features from VOC sensor outputs [15, 16]. Rather than applying dimensionality reduction, we extracted engineered characteristics including minimum values, latency, slope dynamics, and inter-cycle variance that may reflect disease-related biochemical alterations and tissue remodeling processes, as also described in transplant pathology [17]. In subsequent modeling, we emphasized overall discriminative performance. It should be noted that probability cutoffs derived from logistic regression are model-specific and not equivalent to clinical thresholds; therefore, the present analysis was intended to assess discriminative capacity rather than to establish actionable diagnostic thresholds.

Materials and Methods

Clinical Data Collection

Demographic Data

This exploratory cohort study aimed to investigate alterations in electrical resistance derived from VOCs in urine samples using n-type metal oxide semiconductor sensors. Participants were male patients with histopathologically confirmed prostate adenocarcinoma, who had undergone PSA level monitoring both before and after surgery. Eligible subjects were also required to have been available for follow-up and able to provide urine samples for VOC analysis. Informed consent was obtained from all participants.

Demographic data included age, medications, and comorbidities or chronic conditions that could potentially affect VOC detection. Histopathological data were reviewed and recorded from medical records. Subjects were recruited from nephrology and urology outpatient clinics between August 2024 and July 2025. All participants were male, aged over 20 years, and had been mentally competent to independently consent to the use of their clinical and pathological information.

Urine collection for target VOCs

Participants provided ~20 mL urine samples, which were collected in sterile containers and analyzed within 4 hours using our established biosensor protocol [10, 11].

Laboratory Evaluation

Histopathological evaluation of all prostate specimens, including biopsies and surgical samples, was initially performed by an experienced primary pathologist. Reports included the site and number of cancer-positive cores, histologic classification, highest Gleason score, and additional features such as pattern 4 proportion, intraductal carcinoma, and cribriform morphology. Tumor extent was described based on the percentage of core involvement and the presence of periprostatic, seminal vesicle, lymphovascular, or perineural invasion [18]. To ensure diagnostic accuracy and consistency of group allocation, all malignant biopsy specimens were independently reviewed by a second experienced pathologist, who was blinded to clinical information and group categorization.

Serum creatinine levels were obtained via the enzymatic method, performed using the cobas c 702 clinical chemistry analyzer (Roche Diagnostics, Mannheim, Germany) [19]. Serum PSA concentrations were measured using the electrochemiluminescence immunoassay (ECLIA) method, performed on the cobas e 801 immunoassay analyzer (Roche Diagnostics, Mannheim, Germany) [20]. The estimated glomerular filtration rate (eGFR) was determined using the CKD-EPI Creatinine Equation (2021), based on serum creatinine values and relevant clinical parameters [21].

Gas Sensor Analysis

The gas sensing platform comprised five n-type metal oxide semiconductor sensors (S1–S5) integrated into a custom-designed chamber. Each measurement lasted six minutes and was divided into six thermal cycles (i1–i6). During each cycle, the sensor heater voltage was sequentially modulated from 2000 to 5000 mV, allowing repeated exposure to a controlled heating–cooling profile. Electrical resistance was continuously recorded at high temporal resolution across all sensors, producing dynamic response curves for subsequent feature extraction. Specifically, S1 was primarily responsive to alcohol and solvent vapors; S2 to hydrogen and ethanol;

S3 to ammonia and hydrogen sulfide; S4 to butane and liquefied petroleum gas; and S5 to trimethylamine and methyl mercaptan. A schematic of the sensor platform and heat cycle sequence is shown in Supplementary Figure 1.

Feature Extraction

Raw resistance signals obtained from each sensor across six thermal cycles were transformed into quantitative features to capture characteristic response dynamics. For every cycle, the minimum resistance value (Min) was calculated to reflect the nadir of sensor activation; the latency to minimum (Time) was derived as an indicator of kinetic delay; the slope of resistance change (Slope) was estimated from the steepest descent during heating to capture dynamic sensitivity; and the recovery gap (Gap) between successive cycles was measured to assess inter-cycle variance. This feature engineering procedure yielded 120 structured variables per subject (5 sensors × 6 cycles × 4 parameters), thereby enhancing interpretability beyond raw waveform inspection and providing a systematic basis for signal analysis and biomarker prediction [10, 11].

Clinical Allocation and Assessment

Participants were initially categorized into three groups for demographic characterization: (1) normal male subjects, (2) patients with benign prostatic hyperplasia (BPH), and (3) patients with histopathologically confirmed prostate adenocarcinoma. For the purpose of signal analysis, groups (1) and (2) were subsequently combined into a non-PCa cohort, allowing a binary comparison between prostate cancer and non-cancer cases. Subjects with an eGFR below 15 mL/min/1.73 m² or those with high-grade prostatic intraepithelial neoplasia (PIN) were excluded. Smoking status was classified into three categories based on standardized definitions from the National Health Interview Survey: never, former, and current smokers [22].

For the cancer group, the diagnosis of prostate adenocarcinoma was confirmed by pathological examination of biopsy or surgical specimens. In the BPH group, inclusion required histopathological confirmation and either lower urinary tract symptoms or a prostate volume ≥25 mL as determined by transabdominal ultrasonography [23].

For the normal control group, inclusion required a PSA <4 ng/mL, a normal DRE, and prostate volume ≤20 mL. Biopsy was not performed in this group; therefore, strict PSA and ultrasound criteria were applied to minimize misclassification. To ensure clear separation, individuals with prostate volumes of 21–24 mL were excluded.

Statistical analysis

All analyses were conducted using SPSS version 18.0, R version 2.9.1, and Python 3.10 [24-26]. Continuous variables were summarized as mean ± standard deviation (SD), categorical variables as counts and percentages, and a two-sided $p < 0.05$ was considered statistically significant.

To identify discriminative volatile organic compound (VOC) features from sensor signals, an analysis of variance (ANOVA) F-test was applied. The top 10 features ranked by F-score (with $p < 0.001$) were shortlisted for further evaluation. Scatter plots were generated to illustrate separation trends, and receiver operating characteristic (ROC) curves were constructed for univariate performance evaluation. For each candidate feature, the area under the ROC curve (AUC), one-way ANOVA p-value, and Spearman's rank correlation (ρ, p-value) were reported.

Results

Clinical Demographic Data

A total of 117 participants were included, comprising 21 normal controls, 67 patients with BPH, and 29 patients with histologically confirmed prostate cancer. Patients with BPH and prostate cancer were significantly older than healthy controls (22.9 ± 4.9 vs. 67.5 ± 10.5 vs. 74.0 ± 10.2 years; $p < 0.01$). Body mass index was higher in BPH patients (25.2 ± 3.4 kg/m²) compared with controls (20.0 ± 3.0 kg/m²; $p < 0.01$) and cancer patients (22.1 ± 3.7 kg/m²; $p < 0.01$), while no significant difference was observed between controls and cancer patients ($p = 0.12$). (Table 1).

The eGFR was significantly lower in BPH (78.2 ± 17.6 mL/min/1.73 m²) and prostate cancer patients (72.5 ± 19.3 mL/min/1.73 m²) compared with controls (116.4 ± 12.3 mL/min/1.73 m²; $p < 0.01$), with no significant difference between BPH and cancer groups.

Prostatic volume was higher in BPH (39.1 ± 21.0 mL) and prostate cancer patients (59.5 ± 28.2 mL) compared with controls (15.8 ± 2.1 mL; $p < 0.01$). Serum PSA levels increased across groups (controls: 0.16 ± 0.1 ng/mL; BPH: 4.3 ± 7.0 ng/mL; prostate cancer: 328.6 ± 571.9 ng/mL; $p < 0.01$). Smoking status distribution differed significantly ($p < 0.01$): most controls were never-smokers (76.2%), whereas current smokers were prevalent in BPH (55.2%) and prostate cancer (51.7%) groups. Among prostate cancer cases, 62.1% had high-grade Gleason patterns, 62.0% were classified as grade 4–5, and 55.2% had radiologically evident bone metastases at diagnosis.

Sensor-Derived Resistance Dynamics Across Thermal Cycles

Reproducible resistance dynamics were observed across all five sensors (Figure 1), arranged from right to left (Sensor 1 to Sensor 5), over six thermal cycles within the 6-minute measurement period. For each cycle, application of heater voltage produced a drop in electrical resistance, indicating increased conductivity during heating. After cessation of heating, resistance values gradually returned toward baseline; however, recovery was incomplete, and baseline resistance levels progressively decreased with successive cycles. This cumulative decline was consistently observed across all sensors.

When data were color-coded by clinical group (green = normal, red = prostate cancer, blue = BPH), partial group-wise separation was evident, particularly in Sensors 2 through 4. Normal samples demonstrated clustering at higher resistance levels, whereas prostate cancer and BPH groups showed overlapping but lower

Table 1. Baseline Demographic and Clinical Characteristics of Study Participants

Clinical Characteristics	Normal (n= 21)	BPH (n=67)	CA prostate (n=29)	P value		
				Normal vs. BPH	Normal vs. CA prostate	BPH vs. CA prostate
Age (years: mean ± SD)	22.9±4.9	67.5±10.5	74.0±10.2	<0.001	<0.001	0.02
BMI (Kg/m^2: mean ± SD)	20.0±3.0	25.2±3.4	22.1±3.7	<0.001	NS	<0.001
eGFR (mL/min: mean ± SD)	116.4 ± 12.3	78.2 ± 17.6	72.5 ± 19.3	<0.001	<0.001	NS
Prostatic Volume (mL: mean ± SD)	15.8±2.1	39.1±21.0	59.5±28.2	<0.001	<0.001	0.007
PSA (ng/mL: mean ± SD)	0.16±0.1	4.3±7.0	328.6±571.9	<0.001	0.013	0.015
Smoking Status (%: Never/Former/Current)	76.2/9.5/14.3	22.4/22.4/55.2	13.8/34.5/51.7	< 0.001		

Abbreviations: BPH, benign prostatic hyperplasia; CA prostate, prostate cancer; BMI, body mass index; eGFR, estimated glomerular filtration rate; PSA, prostate-specific antigen; SD, standard deviation; NS, not significant; Chi-sq, chi-square test; n, number of participants.

Table 2. Top 10 VOC-derived features with highest univariate discriminative performance (AUC, Spearman's ρ, p-value)

Feature	AUC	Spearman (ρ)	Spearman p-value
Gap_S5_i2	0.77	0.4	p<0.001
Slope_S5_i4	0.76	0.38	p<0.001
Gap_S2_i6	0.75	0.38	p<0.001
Gap_S3_i5	0.75	0.37	p<0.001
Slope_S2_i1	0.75	0.37	p<0.001
Gap_S2_i1	0.75	0.37	p<0.001
Slope_S2_i6	0.75	0.37	p<0.001
Min_S3_i5*	0.26	-0.37	p<0.001
Slope_S3_i5	0.74	0.37	p<0.001
Min_S2_i6*	0.26	-0.37	p<0.001

Note: Features with AUC < 0.5 and negative Spearman's ρ indicate inverse discrimination, meaning lower values are associated with prostate cancer cases.

resistance trajectories across cycles.

Potential VOC signals identification

For subsequent signal analysis, participants were regrouped into two categories: prostate cancer (PCa) versus non-PCa (normal controls and BPH combined). This binary classification was consistent with the study aim of identifying VOC features predictive of cancer presence.

Univariate screening identified several features with strong discriminative ability for separating PCa from non-PCa cases. As shown in Table 2, the top-ranked feature, Gap_S5_i2, achieved an AUC of 0.77 with a positive Spearman correlation ($\rho \approx 0.40$, $p < 0.001$) and a significant group difference by ANOVA ($p < 0.001$). Larger values were observed in the PCa group. Other leading features included Slope_S5_i4 (AUC 0.76, $\rho \approx 0.39$, $p < 0.001$) and Gap_S2_i6 (AUC 0.75, $\rho \approx 0.38$, $p < 0.001$). The top-10 list was dominated by gap and slope

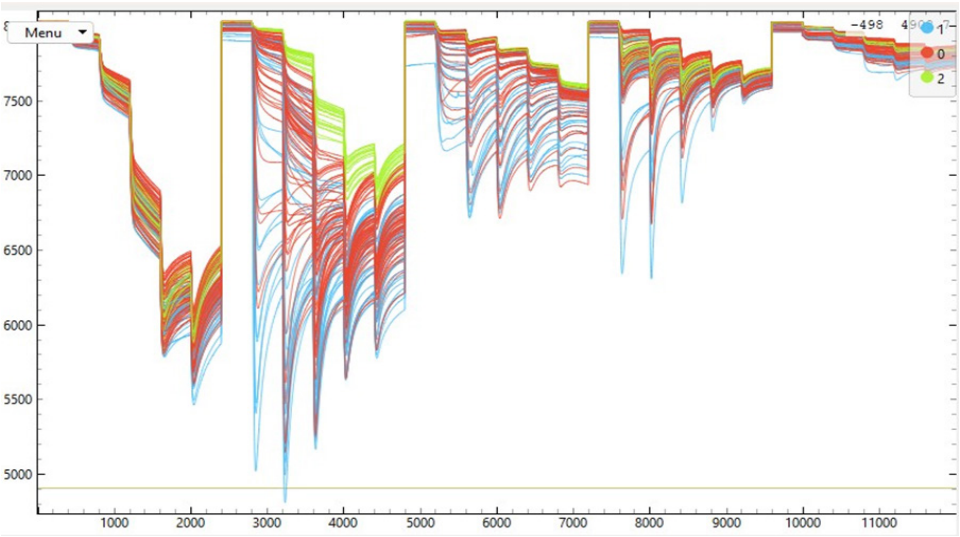

Figure 1. Raw sensor output waveform from five metal oxide sensors across six thermal cycles. Urinary VOC responses measured by five metal oxide semiconductor (MOS) gas sensors over a 6-minute session consisting of six consecutive thermal cycles. Sensor channels are displayed from right to left (Sensor 1 to Sensor 5). Electrical resistance was recorded every 0.15 seconds, yielding 2,400 data points per sensor. Thermal stimulation was applied during cycles 2–6 by modulating heater voltages at five discrete levels (2000, 2500, 3500, 4500, and 5000 mV), while cycle 1 served as the non-heated baseline. Each heating cycle induced a marked decrease in electrical resistance followed by partial recovery, a pattern reproducible across all sensors. Curves represent three diagnostic groups: healthy controls (green), benign prostatic hyperplasia (blue), and prostate cancer (red). The groups can also be distinguished by differences in signal trajectories and amplitudes across heating cycles.

Table 3. Apparent Performance of the Three Logistic Regression Models for Prostate Cancer Detection

Model	AUC	Sensitivity	Specificity	Youden's J
Signal only	0.77	0.63	0.69	0.44
PSA only	0.83	0.63	0.95	0.65
Signal + PSA	0.88	0.77	0.88	0.66

Table 4. Discriminatory Performance of Logistic Regression Models for Prostate Cancer Detection

Model	AUC	95% CI (Low)	95% CI (High)
Signal only	0.77	0.67	0.86
PSA only	0.83	0.72	0.93
Signal + PSA	0.88	0.76	0.9

metrics, particularly from Sensors 5 and 2.

Predictive modeling with VOC signal and PSA

Although several demographic and clinical variables (Table 1) showed significant group-level differences, these were treated as background descriptors rather than diagnostic predictors. Age, BMI, renal function, and smoking status, while statistically different, were non-specific. PSA was retained as the sole clinical predictor for modeling, given its role as the established biomarker and diagnostic benchmark. This approach allowed direct comparison of VOC-derived features against PSA without confounding effects from demographic variables.

Following the univariate screening, Gap_S5_i2 was selected as the representative VOC-derived signal feature for model development, based on its highest discriminative performance (AUC 0.77, $\rho \approx 0.40$). To benchmark its clinical relevance, predictive models were constructed using three configurations: (i) signal-only (Gap_S5_i2), (ii) PSA-only, and (iii) a combined model integrating both signal and PSA. Logistic regression was used to derive risk equations for each configuration, and model probabilities were evaluated using ROC curve analysis.

Logistic regression equations and risk stratification

For each model, predicted probabilities were translated into binary risk categories using the optimal cutoff determined by Youden's J statistic, maximizing the balance between sensitivity and specificity. This enabled direct comparison of discrimination and clinical utility across the three predictor sets. Performance outcomes are summarized in Table 3. Supplementary data provide details of the regression equations.

Decision thresholds determined by Youden's J showed that the signal-only model (Gap_S5_i2) achieved an AUC of 0.77, with sensitivity 0.63 and specificity 0.69 (Youden's J = 0.44). The PSA-only model achieved a higher AUC of 0.83, with sensitivity 0.63 and specificity 0.95 (Youden's J = 0.65). The combined Signal+PSA model yielded the highest performance (AUC 0.88), balancing sensitivity (0.77) and specificity (0.88) at a cutoff of 0.34 (Youden's J = 0.66) (Table 3).

The ROC–AUC curves in Figure 2 display the three logistic regression models evaluated on the full dataset without resampling. At the Youden-optimized cutoffs, the Signal-only model using Gap_S5_i2 showed moderate discrimination, PSA alone performed better, and the combined Signal+PSA model provided the highest AUC with a favorable balance of sensitivity and specificity.

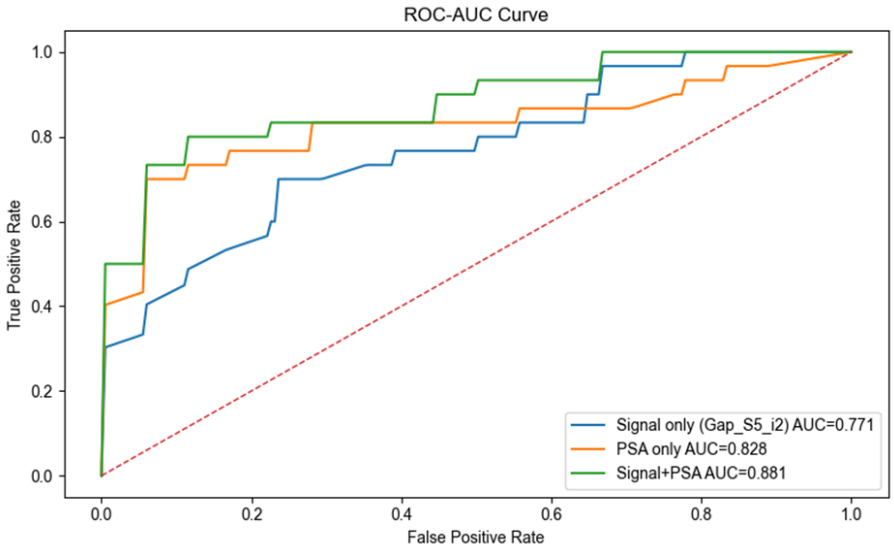

Figure 2 ROC curves comparing the three logistic regression models for prostate cancer detection. The ROC curves compare the apparent (in-sample) performance of three models: a signal-only model based on Gap_S5_i2 (blue line, AUC = 0.771), a PSA-only model (orange line, AUC = 0.828), and a combined signal + PSA model (green line, AUC = 0.881). The diagonal dashed line indicates the reference line for random classification. The combined model achieved the highest AUC, showing improved discriminatory ability compared with either predictor alone. Abbreviations: AUC, area under the curve; PSA, prostate-specific antigen; ROC, receiver operating characteristic.

These represent apparent (in-sample) performance and may be optimistic relative to out-of-sample estimates.

Logistic regression equations were derived for each model to quantify the probability of prostate cancer as a function of the predictors. Detailed formulas are presented in Supplementary data, enabling computation of patient-specific probabilities.

Diagnostic Performance of the Models

To evaluate diagnostic performance, three logistic regression models were directly fitted to the dataset: (1) signal-only, (2) PSA-only, and (3) signal combined with PSA. The discriminatory ability of each model was assessed using the area under the ROC curve (AUC) with corresponding 95% confidence intervals (CI). As summarized in Table 4, the signal-only model achieved an AUC of 0.77 (95% CI: 0.67–0.86). The PSA-only model showed higher discrimination with an AUC of 0.83 (95% CI: 0.72–0.93). The combined model that included both PSA and the signal-derived feature (Gap S5_i2) produced the highest point estimate of discrimination, with an AUC of 0.88 (95% CI: 0.76–0.90).

Although the confidence interval of the combined model overlapped with those of the single-feature models reflecting the modest sample size and inherent variability the higher AUC estimate of the combined model indicates an incremental gain in discrimination when signal-derived features are incorporated with PSA. These results demonstrate that PSA alone retains strong discriminatory ability, but the integration with VOC-based signal features can improve diagnostic performance. The relatively wide confidence intervals underscore that these are apparent (in-sample) estimates and require external validation in larger cohorts to establish robustness and generalizability.

Discussion

Demographics and Clinical Parameters

In the descriptive analysis, comparing the clinical categories of normal, BPH, and prostate cancer revealed a marked difference in age distributions. However, prostate cancer is typically rare in younger men, with the average age at diagnosis reported around 67 years [27]. BPH also shows a strong age-related prevalence, affecting approximately half of men by age 60 and up to 80% by age 80 [28]. Nevertheless, younger individuals can still present with either prostate cancer or BPH due to early onset or genetic predispositions, indicating that the age distinction between groups is less clear-cut in real-world clinical practice [29]. The eGFR values also reflected this limitation, as they are strongly associated with age. Consequently, it is difficult to distinguish whether the observed differences arise from true pathophysiological variation or simply from the age effect embedded in eGFR calculation [17, 30, 31]. Other parameters, including BMI, smoking status, and prostate volume, have been reported to influence prostate cancer risk in previous studies [32-34]. However, these variables often confound each other, reducing the clarity of causal interpretation. For this reason, PSA was chosen as the principal clinical predictor in the present analysis, since it is the most widely accepted biomarker for prostate cancer screening and diagnosis [35].

VOCs in Prostate Cancer Detection

Previous research on urological cancers, including kidney, ureter, and bladder (KUB) malignancies, highlighted methane, iso-butane, ethanol, and hydrogen as volatile compounds with discriminatory potential in disease detection [10, 11]. Traditional approaches relied on comparing mean signal differences between groups to identify candidate features [36]. However, such methods are inherently biased toward features that increase with disease status, while potentially overlooking negatively correlated signals that decrease in affected individuals.

To address this limitation, the present study applied a correlation-based approach. Specifically, Spearman's correlation coefficients were used to rank VOC-derived features according to both the strength and direction of their association with prostate cancer, thereby considering features with positive as well as negative correlations. The top ten ranked features were subsequently selected for modeling. Interestingly, these top-ranking correlations were predominantly derived from sensors 2, 3, and 5, which correspond to target gases similar to those implicated in earlier research [10, 11], namely hydrogen, ethanol, ammonia, hydrogen sulfide (H_2S), trimethylamine, and methylmercaptan.

Further evaluation using AUC analysis revealed that the sensors targeting trimethylamine and methylmercaptan showed the greatest potential to evolve into reliable predictor gases. Importantly, this predictive capacity was specific to the heating cycle at 2500 mV, highlighting both sensitivity and specificity in the detection process. This finding does not contradict earlier work: prior studies primarily demonstrated associations, whereas the present analysis advances the field by employing a correlation-driven strategy that integrates directionality of expression into statistical modeling, thereby enabling risk prediction rather than simple group-level association.

Comparison of PSA, VOCs, and Combined Models

Logistic regression analysis demonstrated that the signal-only model yielded a moderate AUC of 0.77, while the PSA-only model achieved a higher AUC of 0.83. Importantly, the combined model incorporating both PSA and VOC-derived features provided the best performance, with an AUC of 0.88. These findings highlight that PSA remains the principal biomarker for prostate cancer detection [37, 38], but integration with VOC fingerprints enhances specificity and may help reduce false-positive results. From a clinical perspective, the combined approach suggests that VOC patterns could serve as a confirmatory tool, supporting the interpretation of elevated PSA values as cancer-related rather than being attributable to benign conditions such as prostatic hyperplasia or inflammation.

Notably, within the VOC features, gases such as trimethylamine and methylmercaptan emerged as promising discriminators, particularly at specific heating cycles. Unlike traditional GC–MS–based approaches, semiconductor sensors allow real-time, low-cost

recognition of these compounds with greater practical applicability. This suggests a potential pathway for future work to refine target-gas–based detection, whereby incorporating trimethylamine and methylmercaptan may further improve diagnostic specificity when combined with PSA testing.

Clinical Significance and Future Directions

The present findings support the role of VOC fingerprints as a complementary tool to PSA in prostate cancer screening. While PSA remains the cornerstone of detection, combining sensor-derived VOC features with PSA values offers a more practical triage strategy by improving specificity and reducing the burden of unnecessary diagnostic procedures. From a clinical perspective, the integration of VOC sensor data into existing clinical workflows could provide a rapid, non-invasive adjunct that enhances the reliability of early prostate cancer detection.

Future work should extend beyond single-center analyses and include multicenter, prospective validation to establish generalizability across diverse patient populations. Moreover, the incorporation of sensor-based VOC profiles into AI-driven decision-support systems represents a logical next step, enabling real-time risk stratification and potentially guiding individualized diagnostic pathways. Together, these advances highlight the promise of VOC biosensing as a scalable and clinically relevant complement to conventional biomarkers in prostate cancer detection.

Limitations

This study has several limitations. The sample size was modest, and data were obtained from a single center, which may limit the generalizability of the findings. In addition, model performance was evaluated within the study cohort, and further validation in independent populations will be important.

A notable limitation is the marked age difference between the normal control group (mean age 22.9 years) and the BPH and cancer groups (mean ages 67.5 and 74.0 years, respectively). Since metabolic profiles and urinary VOC patterns are known to vary with age, it is difficult to fully distinguish disease-specific biomarker signals from age-related metabolic changes. This age disparity may have influenced the observed differences in VOC features and eGFR values. Future studies should include age-matched control groups to better isolate disease-specific VOC signatures.

A second limitation concerns the definition of the normal control group. Normal controls were defined using clinical and imaging criteria (PSA <4 ng/mL, normal digital rectal examination, and prostate volume ≤20 mL) without histopathological confirmation by biopsy. Although these stringent criteria were applied to minimize the likelihood of misclassification, the possibility of occult prostate cancer cannot be entirely excluded. This may have resulted in a slight underestimation of the true diagnostic performance of the VOC-based models.

A third limitation relates to model validation. All reported performance metrics represent apparent (in-sample) estimates without cross-validation or external validation. The use of Youden's J statistic to determine optimal cutoffs within the same dataset used for model development may lead to optimistic estimates of diagnostic performance. Therefore, the true discriminative ability of these models in independent populations may be lower than reported here. Prospective validation in larger multicenter cohorts with predefined decision thresholds will be necessary before clinical implementation.

A fourth limitation relates to potential unmeasured confounding factors that may influence urinary VOC composition. Variables such as diet, medications, comorbidities, hydration status, and timing of sample collection were not systematically controlled in this exploratory study. These factors may contribute to between-group variation in VOC profiles. Future studies should incorporate standardized sample collection protocols and consider multivariable adjustment to better isolate disease-specific VOC signals.

Despite these limitations, the main contribution of this study remains the demonstration of the potential of VOC-based biosensing, both alone and in combination with PSA, as a complementary approach for prostate cancer detection.

In conclusion, this study demonstrates that urinary VOC fingerprints, particularly signals related to trimethylamine and methylmercaptan, can complement PSA testing in the detection of prostate cancer. While PSA remains the most established biomarker, the integration of VOC-derived features improved diagnostic performance and reduced the likelihood of false positives associated with benign prostatic conditions. The findings support the potential role of VOC biosensing as a rapid, non-invasive adjunct in clinical screening workflows. Future multicenter studies and external validation are warranted to confirm these results and to explore the integration of VOC profiles into AI-driven diagnostic platforms.

Author Contribution Statement

CP was responsible for conceptualizing the study, overseeing its execution, securing funding, and drafting the manuscript. AP and SA contributed to software development and data visualization. CP and SC were involved in data curation, statistical analysis, and methodological design. DK and WA provided support in project administration and methodological planning. All authors reviewed and approved the final version of the manuscript.

Acknowledgements

The authors would like to express their sincere gratitude to Dr. Chaianan Tantisatirapoon, former faculty member of the School of Surgery, Institute of Medicine, Suranaree University of Technology, for his foundational contributions and insightful guidance that led to the conception of this research project. We are also deeply grateful to all participants from the urology clinic at Suranaree University of Technology Hospital for their generous cooperation, including permission to collect

pathological data and urine samples for volatile organic compound (VOC) analysis. This work was supported by the Institute of Research and Development, Suranaree University of Technology, which provided funding for the publication of this article.

Data Availability Statement

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Ethical Approval

This study was approved by the Human Research Ethics Committee under protocol number EC-67-136 and was conducted in accordance with the Declaration of Helsinki.

Disclaimer

The views expressed in this manuscript are solely those of the authors and do not necessarily reflect the position of any affiliated institutions.

Declaration of Competing Interests

The authors declare no competing interests.

References

1. Chen J, Zhang D, Yan W, Yang D, Shen B. Translational bioinformatics for diagnostic and prognostic prediction of prostate cancer in the next-generation sequencing era. Biomed Res Int. 2013;2013:901578. https://doi.org/10.1155/2013/901578.
2. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin. 2022;72(1):7-33. https://doi.org/10.3322/caac.21708.
3. Johns LE, Houlston RS. A systematic review and meta-analysis of familial prostate cancer risk. BJU Int. 2003;91(9):789-94. https://doi.org/10.1046/j.1464-410x.2003.04232.x.
4. Parra-Soto S, Ahumada D, Petermann-Rocha F, Boonpoor J, Gallegos JL, Anderson J, et al. Association of meat, vegetarian, pescatarian and fish-poultry diets with risk of 19 cancer sites and all cancer: Findings from the UK Biobank prospective cohort study and meta-analysis. BMC Med. 2022;20(1):79. https://doi.org/10.1186/s12916-022-02257-9.
5. Rowles JL, 3rd, Ranard KM, Smith JW, An R, Erdman JW, Jr. Increased dietary and circulating lycopene are associated with reduced prostate cancer risk: A systematic review and meta-analysis. Prostate Cancer Prostatic Dis. 2017;20(4):361-77. https://doi.org/10.1038/pcan.2017.25.
6. Al-Fayez S, El-Metwally A. Cigarette smoking and prostate cancer: A systematic review and meta-analysis of prospective cohort studies. Tob Induc Dis. 2023;21:19. https://doi.org/10.18332/tid/157231.
7. Ferlay J, Parkin DM, Steliarova-Foucher E. Estimates of cancer incidence and mortality in Europe in 2008. Eur J Cancer. 2010;46(4):765-81. https://doi.org/10.1016/j.ejca.2009.12.014.
8. Altwaijry N, Somani S, Parkinson JA, Tate RJ, Keating P, Warzecha M, et al. Regression of prostate tumors after intravenous administration of lactoferrin-bearing polypropylenimine dendriplexes encoding TNF-α, TRAIL, and interleukin-12. Drug Deliv. 2018;25(1):679-89. https://doi.org/10.1080/10717544.2018.1440666.
9. Ozah E, Imasogie DE. The diagnostic accuracy of prostate-specific antigen and digital rectal examination in the diagnosis of prostate cancer at the University of Benin Teaching Hospital. J West Afr Coll Surg. 2023;13(3):91-5. https://doi.org/10.4103/jwas.jwas_32_23.
10. Kreepala C, Ayutaya VSN, Phatthanakun R, Thabsuwan K, Paewponsong J. The biosensor using the target urine volatile organic compounds for detecting diabetic kidney disease. Sci Rep. 2025;15(1):14738. https://doi.org/10.1038/s41598-025-00013-6.
11. Suthat Na Ayutaya V, Tantisatirapoon C, Aekgawong S, Anakkamatee W, Danjittrong T, Kreepala C. Urinary cancer detection by the target urine volatile organic compounds biosensor platform. Sci Rep. 2024;14(1):3551. https://doi.org/10.1038/s41598-024-54138-1.
12. Rath R, Farajikhah S, Oveissi F, Dehghani F, Naficy S. Chemiresistive sensor arrays for gas/volatile organic compounds monitoring: A review. Adv Eng Mater. 2022;25. https://doi.org/10.1002/adem.202200830.
13. Chen H, Chen H, Chen J, Song M. Gas sensors based on semiconductor metal oxides fabricated by electrospinning: A review. Sensors (Basel). 2024;24(10). https://doi.org/10.3390/s24102962.
14. Tzanetou E, Manea-Karga E, Baira E, Boutsikou T, Iliodromiti Z, Iacovidou N, et al. Gas and liquid chromatography mass spectrometry as a tool for elucidating volatile organic compounds (vocs) and metabolites in maternal milk: A perspective on infants' health risk assessment. Chemosensors. 2024;12:30. https://doi.org/10.3390/chemosensors12030030.
15. Liu Q, Fan Y, Zeng S, Zhao Y, Yu L, Zhao L, et al. Volatile organic compounds for early detection of prostate cancer from urine. Heliyon. 2023;9(6):e16686. https://doi.org/10.1016/j.heliyon.2023.e16686.
16. Singh S, Semi S, Poornima, Sreelekha G, Adak C, Shukla R, et al. Metal oxide-based gas sensor array for the vocs analysis in complex mixtures using machine learning. 2023. https://doi.org/10.48550/arXiv.2307.06556.
17. Kreepala C, Halloran P. Fundamental concepts regarding graft injury and regeneration: Tissue injury, tissue quality, and recipient factors. 2014.
18. NCCN. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®). 2017.
19. Jørn Erlandsen E, Randers E. Challenges in the measurement of plasma creatinine on the roche cobas c702. Scand J Clin Lab Invest. 2018;78(6):490-5. https://doi.org/10.1080/00365513.2018.1501090.
20. Cazanave Mora JM, Del Valle García R, Pérez López L, Bequer Ariza DC, Zulueta Rodríguez O, Melchor Rodríguez A, et al. Measurement of total and free prostate specific antigen (psa) in human serum samples using an ultra-microanalytical system. J Pharm Biomed Anal. 2022;208:114470. https://doi.org/10.1016/j.jpba.2021.114470.
21. Inker LA, Eneanya ND, Coresh J, Tighiouart H, Wang D, Sang Y, et al. New creatinine- and cystatin C-based equations to estimate GFR without race. N Engl J Med. 2021;385(19):1737-49. https://doi.org/10.1056/NEJMoa2102953.
22. Kramarow EA. Health of former cigarette smokers aged 65 and over: United States, 2018. Natl Health Stat Report. 2020(145):1-12.
23. Kim H, Moon J, Oh Y, Kong M, Park KK, Huh JS. Relationship between prostate volume and lower urinary tract symptom in health checkup subjects. Korean J Urol Oncol. 2020;18:53-60. https://doi.org/10.22465/kjuo.2020.18.1.53.
24. Statistics P. PASW Statistics version 18.0. Chicago, IL: SPSS

Inc.; 2009.. Inc [google scholar]. 2009.
25. R Development Core Team. R: A language and environment for statistical computing. 2010.
26. Van Rossum G, Drake FL Jr. The Python Language Reference. Python Software Foundation; 2014.
27. Rawla P. Epidemiology of prostate cancer. World J Oncol. 2019;10(2):63-89. https://doi.org/10.14740/wjon1191.
28. Roehrborn CG. Benign prostatic hyperplasia. In: Campbell-Walsh Urology. 2012:2570-610. E10.
29. Lange EM, Salinas CA, Zuhlke KA, Ray AM, Wang Y, Lu Y, et al. Early onset prostate cancer has a significant genetic component. Prostate. 2012;72(2):147-56. https://doi.org/10.1002/pros.21414.
30. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, 3rd, Feldman HI, et al. A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009;150(9):604-12. https://doi.org/10.7326/0003-4819-150-9-200905050-00006.
31. Denic A, Glassock RJ, Rule AD. Structural and functional changes with the aging kidney. Adv Chronic Kidney Dis. 2016;23(1):19-28. https://doi.org/10.1053/j.ackd.2015.08.004.
32. MacInnis RJ, English DR. Body size and composition and prostate cancer risk: Systematic review and meta-regression analysis. Cancer Causes Control. 2006;17(8):989-1003. https://doi.org/10.1007/s10552-006-0049-z.
33. Islami F, Moreira DM, Boffetta P, Freedland SJ. A systematic review and meta-analysis of tobacco use and prostate cancer mortality and incidence in prospective cohort studies. Eur Urol. 2014;66(6):1054-64. https://doi.org/10.1016/j.eururo.2014.08.059.
34. Ankerst DP, Till C, Boeck A, Goodman P, Tangen CM, Feng Z, et al. The impact of prostate volume, number of biopsy cores and American Urological Association symptom score on the sensitivity of cancer detection using the prostate cancer prevention trial risk calculator. J Urol. 2013;190(1):70-6. https://doi.org/10.1016/j.juro.2012.12.108.
35. Raychaudhuri R, Lin DW, Montgomery RB. Prostate cancer: A review. JAMA. 2025;333(16):1433-46. https://doi.org/10.1001/jama.2025.0228.
36. Sansone F, Tonacci A. Non-invasive diagnostic approaches for kidney disease: The role of electronic nose systems. Sensors (Basel). 2024;24(19). https://doi.org/10.3390/s24196475.
37. Catalona WJ, Smith DS, Ratliff TL, Dodds KM, Coplen DE, Yuan JJ, et al. Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. N Engl J Med. 1991;324(17):1156-61. https://doi.org/10.1056/nejm199104253241702.
38. Schröder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, et al. Screening and prostate-cancer mortality in a randomized European study. N Engl J Med. 2009;360(13):1320-8. https://doi.org/10.1056/NEJMoa0810084.