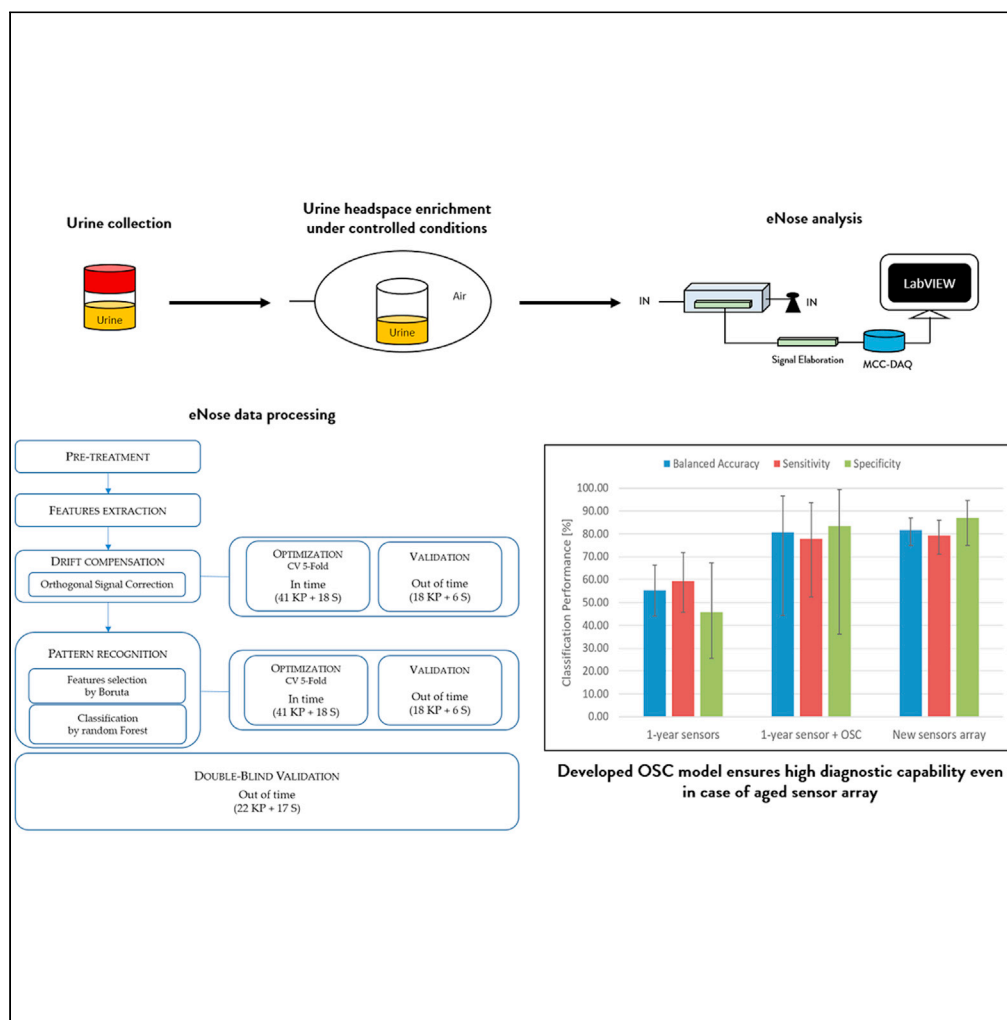


Article

Drift compensation on electronic nose data for non-invasive diagnosis of prostate cancer by urine analysis



Carmen Bax,
Stefano Prudenza,
Giulia Gaspari,
Laura Capelli,
Fabio Grizzi,
Gianluigi Taverna

laura.capelli@polimi.it

Highlights

Urine odor alteration due to prostate cancer can be detected by electronic noses

Sensors drift hinders electronic nose scaling up to large-scale diagnostic devices

OSC mitigates drift on 1-year-old sensors, restoring accuracy from 55% up to 80%



Article

Drift compensation on electronic nose data for non-invasive diagnosis of prostate cancer by urine analysis

Carmen Bax,¹ Stefano Prudenza,¹ Giulia Gaspari,¹ Laura Capelli,^{1,6,*} Fabio Grizzi,^{2,3} and Gianluigi Taverna^{4,5}

SUMMARY

Diagnostic protocol for prostate cancer (KP) is affected by poor accuracy and high false-positive rate. The most promising innovative approach is based on urine analysis by electronic noses (ENs), highlighting a specific correlation between urine alteration and KP presence. Although EN could be exploited to develop non-invasive KP diagnostic tools, no study has already introduced EN into clinical practice, most probably because of drift issues that hinder EN scaling up from research objects to large-scale diagnostic devices. This study, proposing an EN for non-invasive KP detection, describes the data processing protocol applied to a urine headspace dataset acquired over 9 months, comprising 81 patients with KP and 41 controls, for compensating the drift. It proved effective in mitigating drift on 1-year-old sensors by restoring accuracy from 55% up to 80%, achieved by new sensors not subjected to drift. The model achieved, on double-blind validation, a balanced accuracy of 76.2% (CI_{95%} 51.9–92.3).

INTRODUCTION

Prostate cancer (KP) is the second cancer by incidence in men (i.e., 1,276,106 new cases every year worldwide) and the fifth by mortality (Bray et al., 2018), with growing numbers expected for the next decades (Bax et al., 2018; Siegel et al., 2021).

The current diagnostic procedure for KP is based on prostate specific antigen (PSA) serum level, digital rectal examination, and prostate biopsy (Mottet et al., 2017). However, it is complex, time-consuming, scarcely accurate, and expensive because of periodical medical checks (Bax et al., 2018). Moreover, owing to the poor specificity of PSA testing, it leads to many unnecessary biopsies with associated morbidity and overtreatment of clinically insignificant cancers (Örtegren et al., 2019). Consequently, the current screening procedure results in a high health spending (D'Ambrosio et al., 2010).

Over the years, PSA-based algorithms (i.e., the ratio of free to total PSA [%fPSA], PSA density [PSAD], changes in PSA over time, termed PSA velocity [PSAV]) have been proposed. However, although they have slightly improved the specificity of PSA-based screening, they have not definitively resolved the limitations of current diagnostic and prognostic procedures (D'Ambrosio et al., 2010; Jansen et al., 2010).

In recent years, thanks to recent advances in the field of carcinogenesis about metabolic alterations associated with cancer (Bosland et al., 2015), the interest in chemical characterization of biological fluids, especially urine, to identify and quantify innovative biomarkers produced by cancer, has grown significantly (Bianchi et al., 2011; Heger et al., 2014; Jentzmik et al., 2010; Khalid et al., 2015; Sreekumar et al., 2009). Urinary KP biomarkers have emerged as ancillary tools to guide the clinical decision-making in different clinical scenarios. Several recently discovered markers, such as TMPRSS2 (Mosquera et al., 2009; Tomlins et al., 2005, 2008) or PCA3 (Hessels and Schalken, 2009; Wei et al., 2014), or metabolites involved in various pathways associated with cells' demand and production of energy (e.g., sarcosine, isoleucine, threonine, uracil, glutamine) (Bax et al., 2019), can aid in decreasing the number of unnecessary biopsies or over-detection of insignificant cancers.

However, it is undoubtable that these new markers are not without limitations, and their results should be interpreted with caution and in the context of the clinical information. Many of these markers have still not

¹Politecnico di Milano, Department of Chemistry, Materials and Chemical Engineering "Giulio Natta", piazza Leonardo da Vinci 32, Milan 20133, Italy

²Department of Immunology and Inflammation, IRCCS Humanitas Research Hospital, via Manzoni 56, Rozzano, Milan 20089, Italy

³Humanitas University, Via Rita Levi Montalcini 4, Pieve Emanuele, 20090 Milan, Italy

⁴Department of Urology, Humanitas Mater Domini Hospital, Via Gerenzano, 2, Castellanza, 21053 Varese, Italy

⁵Department of Urology, IRCCS Humanitas Research Hospital, via Manzoni 56, Rozzano, 20089 Milan, Italy

⁶Lead contact

*Correspondence:
laura.capelli@polimi.it

<https://doi.org/10.1016/j.isci.2021.103622>



been approved by the Food and Drug Administration and are still considered experimental by most major oncology guidelines.

Moreover, the critical investigation of the scientific literature in the field of metabolomics pointed out that results achieved until now are partial and in some cases contradictory (Bax et al., 2018). It seems that KP is more likely associated with the alteration of a pool of metabolites in terms of both concentration and composition rather than to the one of a single metabolite, thereby suggesting that urine should be analyzed as a whole (Bax et al., 2018, 2019).

An emerging approach, proposing a whole characterization of urine, focused on the analysis of odors emitted, directly relying on the sense of smell of trained dogs (Cornu et al., 2011; Elliker et al., 2014; Fischer-Tenhagen et al., 2018; Gordon et al., 2008; Protoshhak et al., 2019; Taverna et al., 2015a, 2015b; Willis et al., 2004). The results mainly achieved in this field by Taverna et al. (2015a, 2015b) demonstrated the existence of a relation between urine odor and KP. Indeed, highly trained dogs proved capable of differentiating patients with KP from controls with a diagnostic accuracy above 97% by simply sniffing urine samples (Cornu et al., 2011; Elliker et al., 2014; Taverna et al., 2015b).

Based on this evidence, some research groups (Asimakopoulos et al., 2014; Roine et al., 2014; Santonico et al., 2014) started investigating the possibility to transfer those results to an instrumental method based on the analysis of urine samples by electronic noses (ENs), i.e., instruments capable of mimicking mammalian olfaction through the combination of a chemical sensors array and a data processing unit based on machine learning techniques (Gardner and Bartlett, 1994).

Nevertheless, despite the high diagnostic accuracies reported (i.e., close to or above 90%), no study has already led to the introduction of innovative tools in the clinical practice. The main limitations are represented by a small investigated population and the fact that no study validated the classification performance through the execution of specific validation sessions with blind independent samples. Moreover, no study addressed the problem of sensor drift, i.e., the deviation over time of sensor response under the same conditions, currently representing the primary obstacle to the EN diffusion for long-term applications. Indeed, the drift causes a progressive worsening of the classification performance that makes EN unusable for long periods.

For this reason, the present study, conceived within a research project proposing an EN for the non-invasive KP diagnosis, focuses on the description of the data processing protocol specifically developed for compensating drift. In particular, we here propose the adoption of the orthogonal signal correction (OSC) algorithm to develop a specific drift correction model for the urine headspace dataset acquired over 9 months, according to the standardized experimental protocol patented in March 2019 (EP 19160856.1).

Moreover, we present the results of double-blind verification tests aimed to verify in an unbiased way the efficacy of the developed decisional model in compensating drift and assess its diagnostic performance. To do this, blind samples without any clinical information were analyzed and classified by the decisional model including or not the drift compensation step.

RESULTS

Electronic nose predictive model

This section describes the predictive model developed within this project concerning drift compensation and pattern recognition and reports results achieved in terms of diagnostic performance.

Drift compensation

With the purpose of developing the OSC drift compensation model and selecting the optimal number of OSC components, the classification performance achieved on the training set by compensation models including 1–15 OSC components was compared.

Figures 1 and 2 illustrate the trends of sensitivity, specificity, and accuracy as a function of increasing OSC components for the train and test sets, respectively. The classification performance of the decisional model on the train set, assessed by 5-fold cross-validation, keeps growing up to nine OSC components. Even

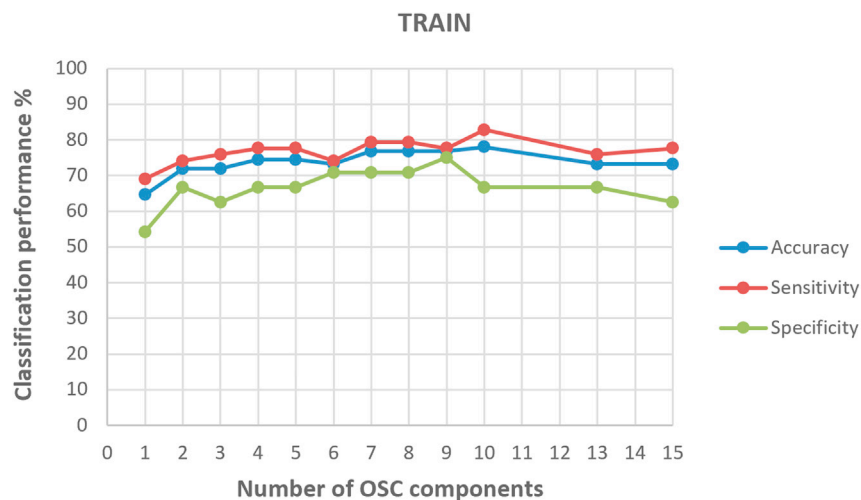


Figure 1. Classification performance on the train set in terms of Accuracy (blue line), sensitivity (red line), and specificity (green line) as a function of the number of OSC components removed

though the accuracy remains almost constant up to 15 OSC components, the specificity of the model reaches its maximum at nine OSC and then drops down to about 65% (Figure 1). Thus, based on the results of the internal validation, nine seemed to be the optimal number of OSC components to be removed from the system to eliminate variance not correlated to samples' clinical status and, thus, compensate for drift effects. Conversely, the results of the external validation, carried out by applying the correction model on more recent samples (i.e., the test set), highlighted a significant improvement of the classification performance after the removal of five OSC components. Then, the classification capability remained constant up to 10 OSC components, whereas a further increase of the OSC component worsened the classification performance: the specificity dropped to about 50% (Figure 2).

Given that the lower the number of OSC components the lower the complexity of the model and the lesser the risk of overfitting, five was selected as the optimal number of OSC components for the construction of the drift correction model to be applied on urine headspace data before pattern recognition.

Prostate cancer diagnosis

The KP diagnosis model was built on 47 steady-state and transient variables selected by Boruta: A4, B1, B5, C4, D5, E5, F5, G1, G2, G5, H1, H2, H5, I3, I5, J2, J3, J5, L3, M1, M2, M5, N2, N5, N6, P4, R4, S6, T2, T3, U1, U3, U5, W1, W2, X3, and Y3, where the letter indicates the type of feature, reported in Table 1, whereas the number refers to the position of the sensor in the EN array.

The KP diagnosis models were built on a subset of the urine headspace dataset (i.e., train), comprising samples analyzed from months 1 to 8, and its classification performance was assessed by 5-fold internal validation. Then, the remaining data (i.e., test), comprising the most recent analyses carried out on months 8 and 9, were processed as unknown samples to assess the model classification capability on data independent from training (external validation). The training set comprised 59 samples (41 KP and 18 S), whereas the independent test set consisted of 24 samples (18 KP and 6 S).

This double-step validation allowed simulating the real conditions of application of the classification model in a relevant environment.

A visual representation of the diagnosis model is reported in Figure 3, in which the two clusters relevant to controls (blue) and patients with KP (red) are clearly distinguishable. Indeed, samples from the control group cluster in the right portion of the plot, whereas most of the samples from the KP group cluster in the left portion of the principal-component analysis (PCA) score plot.

The classification performance achieved by both internal and external validation, summarized in Table 2, proved the capability of the proposed diagnostic tool, based on urine odor analysis, to detect KP with a

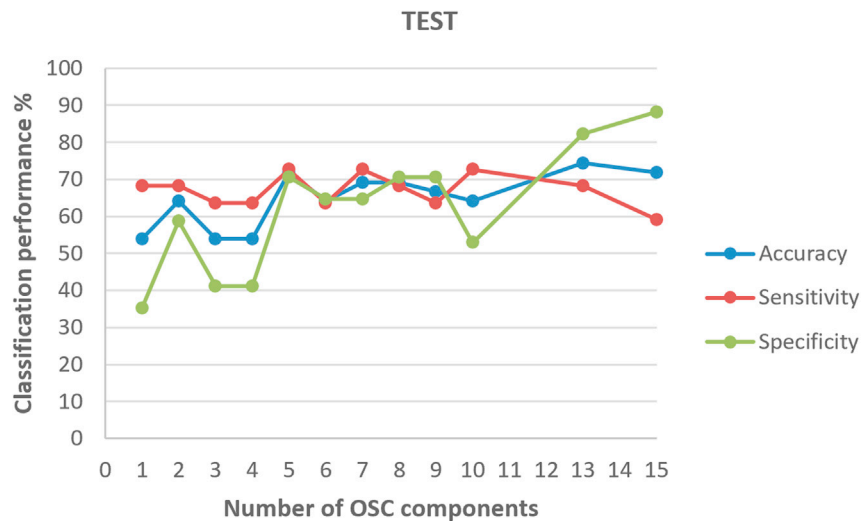


Figure 2. Classification performance on the test set in terms of Accuracy (blue line), sensitivity (red line), and specificity (orange line) as a function of the number of OSC components removed

balanced accuracy of about 81%. The double-step validation proved the robustness of the developed KP diagnosis model. Indeed, the external validation on 24 independent samples confirmed the classification performance achieved on train, which is the most optimistic evaluation.

A random choice classifier will achieve on the training set a classification accuracy of about 57%. Actually, this value is out of the confidence interval 95% associated with the balance accuracy achieved within this research (i.e., $CI_{95\%}$ 60.5–92.7). Therefore, the developed KP diagnosis model proved to be statistically effective in differentiating patients with KP from controls.

Evaluation of drift compensation

As a preliminary investigation, the model capability to compensate drift effects was evaluated visually by exploratory data analysis. PCA was applied to the dataset comprising all features extracted from the sensor responses before and after the OSC correction. Figure 4 and Figure 5 report the PCA score plots relevant to raw data collected over 9 months and corrected data by 5-OSC model, respectively. The PCA model was built on the train (i.e., older analyses), represented in red, where the test set (i.e., more recent data), represented in yellow, was projected on. In Figure 4, test data are separated from the cluster defined by the train. Conversely, after the 5-OSC correction, points belonging to the test set disperse in the train cluster (Figure 5), thereby confirming the efficacy of the drift compensation.

To quantitatively measure the efficacy of the developed data processing procedure in compensating drift effects, the classification performance, achieved including or not the OSC correction, was compared with the one achieved by a new sensor array not affected by drift issues (Figure 6).

The classification performance achieved by a 1-year-old sensor array without applying the OSC model for drift compensation was very poor (i.e., about 55%). Thus, after about 1-year exploitation, the EN resulted to be unusable as a diagnostic device. The short temporal validity of the decisional model represents a stringent limit for the scalability of the EN technology.

Instead, the implementation of the OSC correction allowed significant improvement of the classification performances of 1-year-old sensor array, thereby reaching an accuracy, a specificity, and a sensitivity comparable with the ones obtained by new sensors not subject to drift (i.e., about 80%).

To the best of our knowledge, such significant restoration of performances has been achieved only in a few cases (Fonollosa et al., 2016; Liu et al., 2021). However, it is important to underline that, despite some preliminary feasibility study (Bax et al., 2021a, 2021b), it is the first time that such type of correction algorithm

Table 1. Features extracted from EN signals

Number	Feature	P1	P2	P3	P4	α
A	$\frac{R(P3) + R(P4)}{2}$ $\frac{R(P1) + R(P2)}{2}$	1	2	1,500	1,800	
B	$\frac{R(P3) + R(P4)}{2}$ $\frac{R(P1) + R(P2)}{2}$	1	2	2,700	3,000	
C	$\frac{R(P3) + R(P4)}{2}$ $\frac{R(P1) + R(P2)}{2}$	60	61	1,500	1,800	
D	$\frac{R(P3) + R(P4)}{2}$ $\frac{R(P1) + R(P2)}{2}$	60	61	2,700	3,000	
E	$\frac{R(\text{end} - P1) + R(\text{end} - P2)}{2}$ $\frac{R(3000 + P3) + R(3000 + P4)}{2}$	5	0	5	0	
F	$\frac{R(\text{end} - P1) + R(\text{end} - P2)}{2}$ $\frac{R(3000 + P3) + R(3000 + P4)}{2}$	60	0	60	0	
G	$\frac{R(P3) + R(P4)}{2}$ - $\frac{R(P3) + R(P4)}{2}$	1	2	1,500	1,800	
H	$\frac{R(P3) + R(P4)}{2}$ - $\frac{R(P3) + R(P4)}{2}$	1	2	2,700	3,000	
I	$\frac{R(P3) + R(P4)}{2}$ - $\frac{R(P3) + R(P4)}{2}$	60	61	1,500	1,800	
J	$\frac{R(P3) + R(P4)}{2}$ - $\frac{R(P3) + R(P4)}{2}$	60	61	2,700	3,000	
L	$\frac{R(\text{end} - P1) + R(\text{end} - P2)}{2}$ $\frac{R(3000 + P3) + R(3000 + P4)}{2}$	5	0	5	0	
M	$\frac{R(\text{end} - P1) + R(\text{end} - P2)}{2}$ $\frac{R(3000 + P3) + R(3000 + P4)}{2}$	60	0	60	0	
N	Area of under the curve					
O	Area under the After phase curve					
P	$\frac{R(P1) + R(P2)}{2}$	1,500	1,800			
Q	$\frac{R(P1) + R(P2)}{2}$	2,700	3,000			
R	$\frac{R(P3) - R(P2)}{R(P2) - R(P1)}$	1	1,500	3,000		
S	$R(P2) - R(P1)$	1,500	3,000			
T	EMA maximum value on After phase $y[k] = (1 - \alpha)y[k - 1] + \alpha(x[k] - x[k - 1])$					0.001
U	EMA integral on After phase $y[k] = (1 - \alpha)y[k - 1] + \alpha(x[k] - x[k - 1])$					0.001
V	EMA minimum value on During phase $y[k] = (1 - \alpha)y[k - 1] + \alpha(x[k] - x[k - 1])$					0.001
W	EMA maximum value on After phase $y[k] = (1 - \alpha)y[k - 1] + \alpha(x[k] - x[k - 1])$					0.01
X	EMA integral on After phase $y[k] = (1 - \alpha)y[k - 1] + \alpha(x[k] - x[k - 1])$					0.01
Y	EMA minimum value on During phase $y[k] = (1 - \alpha)y[k - 1] + \alpha(x[k] - x[k - 1])$					0.01
Z	$\frac{R(P1)}{R(P2)}$	1	3,000			

has been applied to complex samples, as is the case for urine headspaces. In general, some studies (Fonollosa et al., 2016; Liu et al., 2021) proposed the application of drift correction models (e.g., component correction, domain adaptation, extreme learning) on synthetic dataset comprising the analysis of reference

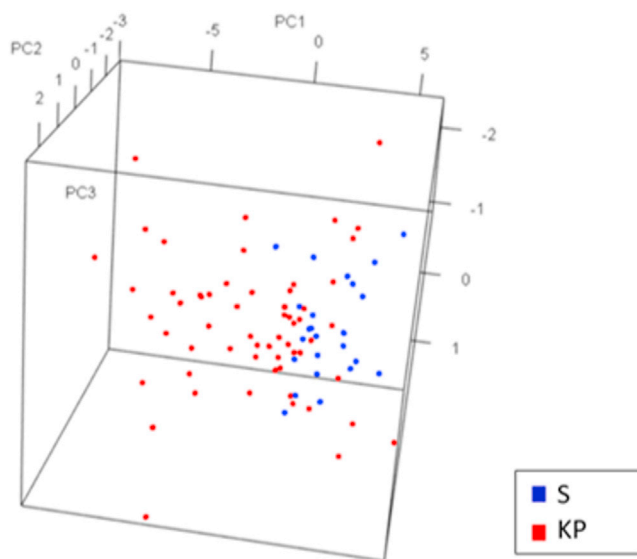


Figure 3. PCA score plot relevant to the diagnosis model built on the dataset corrected by 5-OSC drift correction model

KP, Prostate cancer; S, Control

substances at a known concentration (e.g., Ethanol, Ethylene, Acetone, Methane, Ammonia), thereby easing their implementation.

Therefore, results achieved within this project are unique and confirm the potentialities of the OSC as a powerful drift correction technique to be applied on EN data also in case of complex odor matrices, thereby showing the opportunity of developing a non-invasive, reliable, and cheap diagnostic tool for KP based on the analysis of urine odor.

Double-blind verification tests

This section presents the results of double-blind verification tests carried out within this research to assess, in an unbiased way, the classification performance of the developed decisional model and its efficacy in compensating drift effects on unknown and more recent samples. A total of 39 double-blind samples, independent from 83 samples involved in implementing the EN predictive model, were analyzed in month 9 and used for this purpose.

Table 2. Classification performance achieved by the KP diagnosis model assessed by internal (5-fold CV) and external validation

Diagnostic capability	Train 5-fold cross-validation (41 KP and 18 S)		Test External validation (18 KP and 6 S)	
	%	CI _{95%}	%	CI _{95%}
Balanced accuracy	80.7	60.5–92.7	80.6	44.2–96.6
Sensitivity	78.0	62.4–89.4	77.8	52.4–93.6
Specificity	83.3	58.6–96.4	83.3	36.0–99.5
NPV	62.5	40.6–81.2	55.6	21.2–86.3
PPV	91.4	76.9–98.2	93.3	68.0–99.8

NPV, negative predictive value; PPV, positive predictive value.

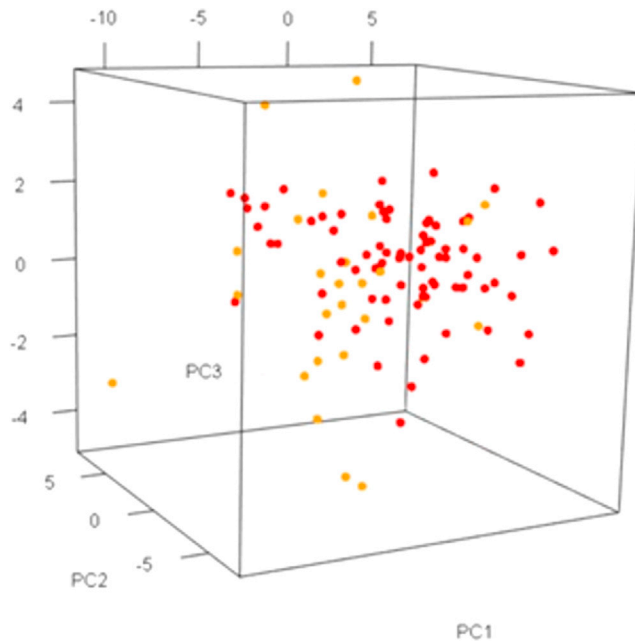


Figure 4. PCA score plot relevant to raw data before the application of the 5-OSC drift correction model

It is worthy to mention that, for patients with KP, the clinical condition was assessed based on the results of histopathological examination of the prostate after radical prostatectomy rather than prostate biopsy, aiming to limit the uncertainty associated with poor accuracy of the prostate biopsy.

The results of the KP diagnosis model, including the 5-OSC correction as the first step, were organized in a confusion matrix, reported in [Table 3](#), in order to evaluate the classification performance.

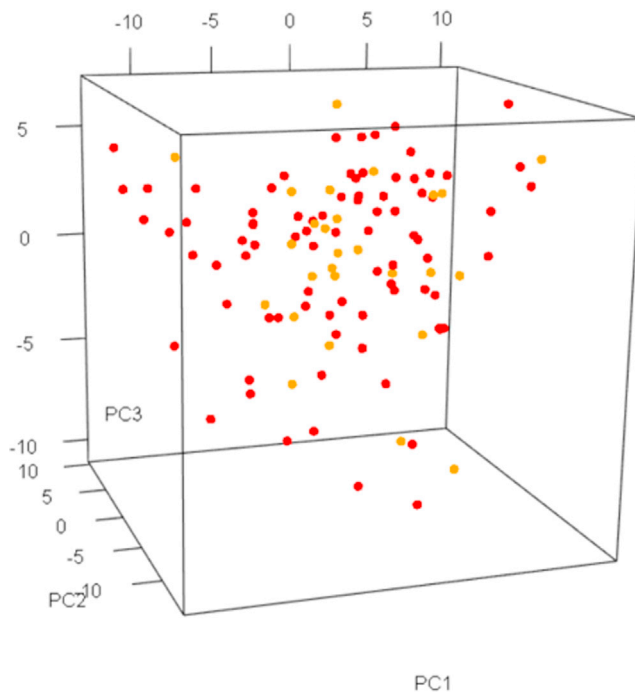


Figure 5. PCA score plot relevant to data corrected by 5-OSC drift correction model

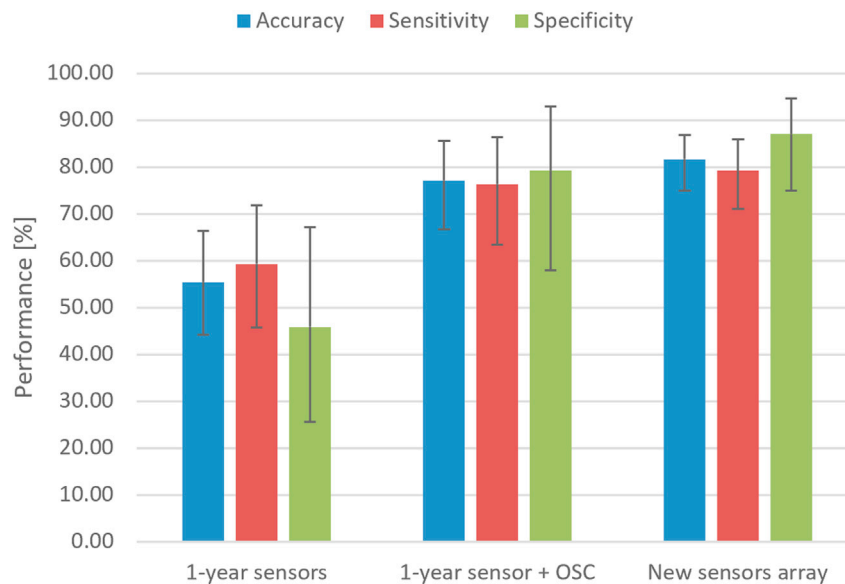


Figure 6. Comparison of the classification performances obtained with internal cross-validation of new and aged sensor arrays, before and after drift correction (error bars = 95% confidence interval)

The developed KP diagnosis model achieved on double-blind samples a balanced accuracy of 76.2% (CI_{95%} 51.9%–92.3%), which is considerably higher than that achieved by current protocols based on PSA serum level and prostate biopsy, whose accuracy is approximately 58%. Results, summarized in Table 4, proved the potentialities of the EN as a novel tool for the non-invasive diagnosis of KP. Indeed, the innovative tool achieved a specificity (i.e., about 71%) significantly higher than the one achieved by the current protocol (i.e., about 33%). Thus, it might represent in the future a solution to the problem of patients' overtreatment.

Most of all, the results of double-blind tests, simulating the real scenario of application of the EN if introduced in the clinical practice, proved its applicability for long periods. Indeed, the classification performance achieved, after 5-OSC drift compensation, by a 1-year-old sensor array on more recent and independent data was comparable with that of new sensors not subjected to drift (Figure 6). Although a small decrease in the specificity balanced by an increase in sensitivity was recorded, the confidence on both double-blind samples (Table 4) and training data (Table 2) overlap.

Table 4 also reports the classification performance achieved by the EN predictive model if the 5-OSC drift compensation is not applied prior to classification. In that case, the classification performance achieved is not acceptable: the EN completely loses the capability to recognize urine samples from patients with KP. To the best of our knowledge, no other literature study proposing the analysis of urine odors by EN for diagnostic purposes, which achieved diagnostic performance comparable with or even higher than that reported here, addressed the problem of drift. In particular, all studies published in the relevant scientific literature (Aggio et al., 2016; Asimakopoulos et al., 2014; D'Amico et al., 2012; Roine et al., 2014) reported the results of researches involving EN data collected within a relatively short period of time. In this condition, the instrument and the sensors did not experience any problem of aging, thereby easing EN capability to distinguish samples belonging to different classes (i.e., controls and patients with KP). Nevertheless, in real-life applications, aging and drift cannot be avoided. Thus, literary results not considering drift effects are hardly generalized. Conversely, the approach proposed here could provide, after validation on a larger population, a pathway to ensure stable classification performance over time and limit efforts associated with periodical recalibration of the EN, thereby enabling its scalability.

DISCUSSION

The present study proposes the analysis of urine odor by means of an EN for a non-invasive diagnosis of KP. The research focused on the description of the data processing procedure specifically developed to tackle

Table 3. Confusion matrix relevant to the classification of double-blind samples by the diagnosis model including 5-OSC correction (KP, Prostate cancer; S, Control)

Clinical condition	EN classification	
	KP	S
KP	18	4
S	5	12

drift problem. Because of the progressive worsening of EN classification performance over time, drift represents one of the main obstacles for EN to switch from research objects to large-scale diagnostic devices, despite the very promising results that have been reported in the scientific literature proposing EN for urine analysis.

Aiming to make results generalizable and limit efforts associated to EN periodical recalibrations, a drift correction model based on OSC algorithm was developed and validated by means of specific double-blind tests. A dataset, comprising the analysis of urine headspaces from 122 subjects (81 collected from men with KP and 41 collected from control donors), by the same EN sensor array over 9 months was used.

Our findings proved the efficacy of the proposed strategy in mitigating drift effects on 1-year-old sensors. More in detail, the 5-OSC model allowed restoration of the diagnostic accuracy of 1-year-old sensors from 55% up to about 80%, which was achieved by new sensors not subjected to drift issues (Figure 6). Those results proved the potentialities of the developed approach to overcome issues associated with sensor aging.

The classification performance appeared particularly encouraging and showed the opportunity of developing a novel tool for the non-invasive diagnosis of KP based on urine odor analysis by EN. Compared with the current diagnostic procedure for KP, based on PSA serum levels and prostate biopsy, the EN proved considerably powerful. Indeed, being considerably more accurate than current tools (i.e., accuracy of about 80% compared with 58% [Harvey et al., 2009]), it might provide an effective solution to patients' overtreatment associated with high false-positive rates of the PSA test.

Limitations of the study

Two major aspects should be addressed before EN may become a large-scale KP diagnostic tool: the transfer ability of prediction models and the validation in the relevant environment.

Indeed, because of poor reproducibility of gas sensors, prediction models developed on an instrument can be hardly transferred as they are to other devices. In general, a recalibration of the model is needed. Consequently, scaling of the EN up to an industrial level results in an expensive and time-intensive process. Future studies should, thus, focus on the development of a specific model for extending the validity of calibration models built on a master EN also on untrained devices.

Concerning results of validation, the execution of a multi-centric clinical trial to a more extended number of patients needs to be carried out to gain acceptance by the medical community. This will allow the achievement of TRL5 and make the introduction of EN in the clinical practice realistic.

Table 4. Classification performance relevant to double-blind verification tests achieved by diagnosis model

Diagnostic capability	5-OSC model applied		No drift compensation	
	%	CI _{95%}	%	CI _{95%}
Balanced accuracy	76.2	51.9–92.3	47.1	25.6–57.9
Sensitivity	81.8	59.7–94.8	0	0–12.7
Specificity	70.6	44.0–89.7	94.1	71.3–99.8
NPV	75.0	47.6–92.7	42.1	26.3–59.2
PPV	78.3	52.3–92.5	0	0–95.0

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCE TABLE](#)
- [RESOURCE AVAILABILITY](#)
 - Lead contact
 - Materials availability
 - Data and code availability
- [EXPERIMENTAL MODEL AND SUBJECTS DETAILS](#)
 - Population
- [METHODS DETAILS](#)
 - Electronic nose
 - Experimental protocol
 - Data processing
 - Pre-treatment
 - Features extraction
 - Drift compensation
 - Pattern recognition
 - Double-blind verification tests
- [ADDITIONAL RESOURCES](#)

AUTHOR CONTRIBUTIONS

C.B. conceived the structure of, drafted, and revised the manuscript; S.P. and G.G. carried out the experimental activity and drafted the paper; L.C. coordinated the study, helped with data interpretation, and revised the manuscript; F.G. contributed to the understanding of clinical aspects and revised the manuscript; G.T. coordinated the clinical research and revised the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 12, 2021

Revised: October 7, 2021

Accepted: December 10, 2021

Published: January 21, 2022

REFERENCES

- Aggio, R.B., de Lacy Costello, B., White, P., Khalid, T., Ratcliffe, N.M., Persad, R., and Probert, C.S. (2016). The use of a gas chromatography-sensor system combined with advanced statistical methods, towards the diagnosis of urological malignancies. *J. Breath Res.* *10*, 017106.
- Artursson, T., Eklöv, T., Lundström, I., Mårtensson, P., Sjöström, M., and Holmberg, M. (2000). Drift correction for gas sensors using multivariate methods. *J. Chemom.* *14*, 711–723.
- Asimakopoulos, A.D., Del Fabbro, D., Miano, R., Santonico, M., Capuano, R., Pennazza, G., D'Amico, A., and Finazzi-Agrò, E. (2014). Prostate cancer diagnosis through electronic nose in the urine headspace setting: a pilot study. *Prostate Cancer Prostatic Dis.* *17*, 206–211.
- Bax, C., Taverna, G., Eusebio, L., Sironi, S., Grizzi, F., Guazzoni, G., and Capelli, L. (2018). Innovative diagnostic methods for early prostate cancer detection through urine analysis: a review. *Cancers (Basel)* *10*, 123.
- Bax, C., Lotesoriere, B.J., Sironi, S., and Capelli, L. (2019). Review and comparison of cancer biomarker trends in urine as a basis for new diagnostic pathways. *Cancers* *11*, 1244.
- Bax, C., Bernasconi, R., Massironi, F., Magagnin, L., Grizzi, F., Capelli, L., and Taverna, G. (2021a). Inkjet printed ZnO sensors for early prostate cancer detection by means of urine odor analysis. *J. Electrochem. Soc.* *168*, 047513.
- Bax, C.P., Gutierrez-Galvez, A., and Capelli, L. (2021b). Drift compensation on electronic nose data relevant to the monitoring of odorous emissions from a landfill by opls. *Chem. Eng. Trans.* *85*, 13–18.
- Bianchi, F., Dugheri, S., Musci, M., Bonacchi, A., Salvadori, E., Arcangeli, G., Cupelli, V., Lanciotti, M., Masieri, L., Serni, S., Carini, M., Careri, M., and Mangia, A. (2011). Fully automated solid-phase microextraction-fast gas chromatography-mass spectrometry method using a new ionic liquid column for high-throughput analysis of sarcosine and N-ethylglycine in human urine and urinary sediments. *Anal. Chim. Acta* *707*, 197–203.
- Biehl, W., Hattesoehl, A., Jörres, R.A., Duell, T., Althöhn, U., Koczulla, A.R., and Schmetzer, H. (2019). VOC pattern recognition of lung cancer: a comparative evaluation of different dog- and eNose-based strategies using different sampling materials. *Acta Oncol.* *58*, 1216–1224.
- Bolón-Canedo, V., Sánchez-Marroño, N., and Alonso-Betanzos, A. (2013). A review of feature selection methods on synthetic data. *Knowl. Inf. Syst.* *34*, 483–519.
- Bosland, M.C., Ozten, N., Eskra, J.N., and Mahmoud, A.M. (2015). A perspective on prostate carcinogenesis and chemoprevention. *Curr. Pharmacol. Rep.* *1*, 258–265.
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* *68*, 394–424.
- Breiman, L. (2001). Random Forests. *Mach. Learn.* *45*, 5–32.

- Di Carlo, S., and Falasconi, M. (2012). In Drift Correction Methods for Gas Chemical Sensors in Artificial Olfaction Systems: Techniques and Challenges (Politecnico di Torino).
- Di Carlo, S., Falasconi, M., Sanchez, E., Scionti, A., Squillero, G., and Tonda, A. (2011). Increasing pattern recognition accuracy for chemical sensing by evolutionary based drift compensation. *Pattern Recognit. Lett.* 32, 1594–1603.
- Cornu, J.N., Cancel-Tassin, G., Ondet, V., Girardet, C., and Cussenot, O. (2011). Olfactory detection of prostate cancer by dogs sniffing urine: a step forward in early diagnosis. *Eur. Urol.* 59, 197–201.
- D'Ambrosio, G.G., Campo, S., Cancian, M., Pecchioli, S., and Mazzaglia, G. (2010). Opportunistic prostate-specific antigen screening in Italy: 6 years of monitoring from the Italian general practice database. *Eur. J. Cancer Prev.* 19, 413–416.
- Distante, C., Sicilian, P., and Persaud, K. (2002). Dynamic cluster recognition with multiple self-organising maps. *Pattern Anal. Appl.* 5, 306–315.
- D'Amico, A., Santonico, M., Pennazza, G., Capuano, R., Vespasiani, G., Fabbro, D., Paollesse, R., Natale, C., Martinelli, E., and Agrò, E. (2012). A novel approach for prostate cancer diagnosis using a gas sensor array. *Proc. Eng.* 47, 1113–1116.
- Elliker, K.R., Sommerville, B.A., Broom, D.M., Neal, D.E., Armstrong, S., and Williams, H.C. (2014). Key considerations for the experimental training and evaluation of cancer odour detection dogs: lessons learnt from a double-blind, controlled trial of prostate cancer detection. *BMC Urol.* 14, 22.
- Fearn, T. (2000). On orthogonal signal correction. *Chemometrics Intell. Lab. Syst.* 50, 47–52.
- Fischer-Tenhagen, C., Johnen, D., Nehls, I., and Becker, R. (2018). A proof of concept: are detection dogs a useful tool to verify potential biomarkers for lung cancer? *Front. Vet. Sci.* 5, 52.
- Fonollosa, J., Fernández, L., Gutiérrez-Gálvez, A., Huerta, R., and Marco, S. (2016). Calibration transfer and drift counteraction in chemical sensor arrays using Direct Standardization. *Sensors Actuators B: Chem.* 236, 1044–1053.
- Fryder, M., Holmberg, M., Winqvist, F., and Lundstrom, I. (1995). A calibration technique for an electronic nose. *Proc. Int. Solid-State Sensors Actuators Conf. - TRANSDUCERS 95*, 25–29.
- Gardner, J.W., and Bartlett, P.N. (1994). A brief history of electronic noses. *Sens. Actuators B Chem.* 18, 210–211.
- Gardner, J.W., and Bartlett, P.N. (2000). Electronic noses. Principles and applications. *Meas. Sci. Technol.* 11, 1087.
- Gordon, R.T., Schatz, C.B., Myers, L.J., Kosty, M., Gonczy, C., Kroener, J., Tran, M., Kurtzhals, P., Heath, S., Koziol, J.A., et al. (2008). The use of canines in the detection of human cancers. *J. Altern. Complement Med.* 14, 61–67.
- Guest, C., Harris, R., Sfanos, K.S., Shrestha, E., Partin, A.W., Trock, B., Mangold, L., Bader, R., Kozak, A., McLean, S., et al. (2021). Feasibility of integrating canine olfaction with chemical and microbial profiling of urine to detect lethal prostate cancer. *PLoS One* 16, e0245530.
- Gutierrez-Osuna, R. (2000). Drift reduction for metal-oxide sensor arrays using canonical correlation regression and partial least squares. *Electron. Noses Olfaction*, 147–152.
- Harvey, P., Basuita, A., Endersby, D., Curtis, B., Iacovidou, A., and Walker, M. (2009). A systematic review of the diagnostic accuracy of prostate specific antigen. *BMC Urol.* 9, 14.
- Heger, Z., Cernei, N., Gumulec, J., Masarik, M., Eckschlager, T., Hrabec, R., Zitka, O., Adam, V., and Kizek, R. (2014). Determination of common urine substances as an assay for improving prostate carcinoma diagnostics. *Oncol. Rep.* 31, 1846–1854.
- Hessels, D., and Schalken, J.A. (2009). The use of PCA3 in the diagnosis of prostate cancer. *Nat. Rev. Urol.* 6, 255–261.
- Hui, D., Jun-Hua, L., and Zhong-Ru, S. (2003). Drift reduction of gas sensor by wavelet and principal component analysis. *Sens. Actuators B Chem.* 96, 354–363.
- Jansen, F.H., van Schaik, R.H., Kurstjens, J., Horninger, W., Klocker, H., Bektic, J., Wildhagen, M.F., Roobol, M.J., Bangma, C.H., and Bartsch, G. (2010). Prostate-specific antigen (PSA) isoform p2PSA in combination with total PSA and free PSA improves diagnostic accuracy in prostate cancer detection. *Eur. Urol.* 57, 921–927.
- Jentzmik, F., Stephan, C., Miller, K., Schrader, M., Erbersdobler, A., Kristiansen, G., Lein, M., and Jung, K. (2010). Sarcosine in urine after digital rectal examination fails as a marker in prostate cancer detection and identification of aggressive tumours. *Eur. Urol.* 58, 12–18.
- Khalid, T., Aggio, R., White, P., De Lacy Costello, B., Persad, R., Al-Kateb, H., Jones, P., Probert, C.S., and Ratcliffe, N. (2015). Urinary volatile organic compounds for the detection of prostate cancer. *PLoS One* 10, e0143283.
- Kokocińska-Kusiak, A., Woszczyło, M., Zybała, M., Maciocha, J., Barłowska, K., and Dzieciół, M. (2021). Canine olfaction: physiology, behavior, and possibilities for practical applications. *Animals* 11, 2463.
- Kursa, M.B., and Rudnicki, W.R. (2010). Feature selection with the boruta package. *J. Stat. Softw.* 1, 11.
- Laref, R., Ahmadou, D., Losson, E., and Siadat, M. (2017). Orthogonal signal correction to improve stability regression model in gas sensor systems. *J. Sensors* 2017, 9851406.
- Liaw, A., and Wiener, M. (2001). Classification and regression by Random Forest. *Forest* 23, 18–22.
- Liu, T., Li, D., and Chen, J. (2021). An active method of online drift-calibration-sample formation for an electronic nose. *Measurement* 171, 108748.
- Llobet, E., Brezmes, J., Ionescu, R., Vilanova, X., Al-Khalifa, S., Gardner, J.W., Bårnsan, N., and Correig, X. (2002). Wavelet transform and fuzzy ARTMAP-based pattern recognition for fast gas identification using a micro-hotplate gas sensor. *Sens. Actuators B Chem.* 83, 238–244.
- Martinelli, E., Magna, G., De Vito, S., Di Fuccio, R., Di Francia, G., Vergara, A., and Di Natale, C. (2013). An adaptive classification model based on the Artificial Immune System for chemical sensor drift mitigation. *Sens. Actuators B Chem.* 177, 1017–1026.
- Mazzola, S.M., Pirrone, F., Sedda, G., Gasparri, R., Romano, R., Spaggiari, L., and Mariangela, A. (2020). Two-step investigation of lung cancer detection by sniffer dogs. *J. Breath Res.* 14, 026011.
- Mosquera, J.-M., Mehra, R., Regan, M.M., Perner, S., Genega, E.M., Bueti, G., Shah, R.B., Gaston, S., Tomlins, S.A., Wei, J.T., et al. (2009). Prevalence of TMPRSS2-ERG fusion prostate cancer among men undergoing prostate biopsy in the United States. *Clin. Cancer Res.* 15, 4706–4711.
- Mottet, N., Bellmunt, J., Bolla, M., Briers, E., Cumberbatch, M.G., De Santis, M., Fossati, N., Gross, T., Henry, A.M., Joniau, S., et al. (2017). EAU-ESTRO-SIOG guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative intent. *Eur. Urol.* 71, 618–629.
- Muezzinoglu, M.K., Vergara, A., Huerta, R., Rulkov, N., Rabinovich, M.I., Selverston, A., and Abarbanel, H.D.I. (2009). Acceleration of chemo-sensory information processing using transient features. *Sens. Actuators B Chem.* 137, 507–512.
- Di Natale, C., Martinelli, E., and D'Amico, A. (2002). Counteraction of environmental disturbances of electronic nose data by independent component analysis. *Sens. Actuators B Chem.* 82, 158–165.
- Örtengren, J., Holmberg, J.T., Lekås, E., Mana, S., Mårtensson, S., Richthoff, J., Sundqvist, P., Kjölhede, H., Bratt, O., and Liedberg, F. (2019). A randomised trial comparing two protocols for transrectal prostate repeat biopsy: six lateral posterior plus six anterior cores versus a standard posterior 12-core biopsy. *Scand. J. Urol.* 53, 217–221.
- Padilla, M., Perera, A., Montoliu, I., Chaudry, A., Persaud, K., and Marco, S. (2010). Drift compensation of gas sensor array data by orthogonal signal correction. *Chemometrics Intell. Lab. Syst.* 100, 28–35.
- Protoshhak, V.V., Andreev, E.A., Karpushhenko, E.G., Slepcev, A.V., Ovchinnikov, D.V., Alentev, S.A., Lazutkin, M.V., Mamaenko, A.V., and Mamaenko, T.V. (2019). Prostate cancer and dogs sense of smell: opportunities of noninvasive diagnostics. *Urologia* 5, 22–26.
- Roine, A., Veskimäe, E., Tuokko, A., Kumpulainen, P., Koskimäki, J., Keinänen, T.A., Häkkinen, M.R., Vepsäläinen, J., Paavonen, T., Leikkala, J., et al. (2014). Detection of prostate cancer by an electronic nose: a proof of principle study. *J. Urol.* 192, 230–234.
- Santonico, M., Pennazza, G., Asimakopoulos, A.D., Del Fabbro, D., Miano, R., Capuano, R., Finazzi-Agrò, E., and D'Amico, A. (2014). Chemical sensors for prostate cancer detection oriented to non-invasive approach. *Proc. Eng.* 87, 320–323.

Siegel, R.L., Miller, K.D., Fuchs, H.E., and Jemal, A. (2021). Cancer statistics, 2021. *CA Cancer J. Clin.* *71*, 7–33.

Sreekumar, A., Poisson, L.M., Rajendiran, T.M., Khan, A.P., Cao, Q., Yu, J., Laxman, B., Mehra, R., Lonigro, R.J., Li, Y., et al. (2009). Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* *457*, 910–914.

Taverna, G., Tidu, L., and Grizzi, F. (2015a). Sniffing out prostate cancer: a new clinical opportunity. *Cent. Eur. J. Urol.* *68*, 308–310.

Taverna, G., Tidu, L., Grizzi, F., Torri, V., Mandressi, A., Sardella, P., La Torre, G., Coccione, G., Seveso, M., Giusti, G., et al. (2015b). Olfactory system of highly trained dogs detects prostate cancer in urine samples. *J. Urol.* *193*, 1382–1387.

Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., et al. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* *310*, 644–648.

Tomlins, S.A., Laxman, B., Varambally, S., Cao, X., Yu, J., Helgeson, B.E., Cao, Q., Prensner, J.R., Rubin, M.A., Shah, R.B., et al. (2008). Role of the TMPRSS2-ERG gene fusion in prostate cancer. *Neoplasia* *10*, 177–188.

Vergara, A., Vembu, S., Ayhan, T., Ryan, M.A., Homer, M.L., and Huerta, R. (2012). Chemical gas sensor drift compensation using classifier ensembles. *Sens. Actuators B: Chem.* *166–167*, 320–329.

Wang, Y., Yang, A., Chen, X., Wang, P., Wang, Y., and Yang, H. (2017). A deep learning approach for blind drift calibration of sensor networks. *IEEE Sensors J.* *17*, 4158–4171.

Wei, J.T., Feng, Z., Partin, A.W., Brown, E., Thompson, I., Sokoll, L., Chan, D.W., Lotan, Y., Kibel, A.S., Busby, J.E., et al. (2014). Can urinary PCA3 supplement PSA in the early detection of prostate cancer? *J. Clin. Oncol.* *32*, 4066–4072.

Willis, C.M., Church, S.M., Guest, C.M., Cook, W.A., McCarthy, N., Bransbury, A.J., Church, M.R., and Church, J.C. (2004). Olfactory detection

of human bladder cancer by dogs: proof of principle study. *BMJ* *329*, 712.

Wold, S., Antti, H., Lindgren, F., and Öhman, J. (1998). Orthogonal signal correction of near-infrared spectra. *Chemometrics Intell. Lab. Syst.* *44*, 175–185.

Zeaiter, M., Roger, J., and Bellon-Maurel, V. (2005). Robustness of models developed by multivariate calibration. Part II: the influence of pre-processing methods. *Trends Anal. Chem.* *24*, 437–445.

Zhang, J., Zhang, Z., Xiang, Y., Dai, Y., and Harrington Pde, B. (2011). An emphatic orthogonal signal correction-support vector machine method for the classification of tissue sections of endometrial carcinoma by near infrared spectroscopy. *Talanta* *83*, 1401–1409.

Zuppa, M., Distante, C., Siciliano, P., and Persaud, K. (2004). Drift counteraction with multiple self-organising maps for an electronic nose. *Sens. Actuators B Chem.* *98*, 305–317.

STAR★METHODS

KEY RESOURCE TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Urine samples	Humanitas Mater Domini Hospital in Castellanza (VA)	N/A
Raw and analyzed data	This paper	N/A
Software and algorithms		
Standard Normal Variate	(Zeaiter et al., 2005)	N/A
Orthogonal Signal Correction	(Fearn, 2000)	N/A
Boruta	(Kursa & Rudnicki, 2010)	Rstudio, package = Boruta
Random Forest	(Breiman, 2001)	Rstudio, package = randomForest

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, prof. Laura Capelli (laura.capelli@polimi.it).

Materials availability

This study did not use commercial reagents, but only real urine samples provided by the Humanitas Mater Domini Hospital in Castellanza (VA).

Data and code availability

Data and code cannot be uploaded to a public repository, but remain available upon reasonable request to the lead contact, prof. Laura Capelli (laura.capelli@polimi.it).

EXPERIMENTAL MODEL AND SUBJECTS DETAILS

Population

This study has involved 122 subjects: 81 men with KP, 35 healthy young males (age: 18–25 years) with negative family history of KP and prostate specific antigen (PSA) < 1 ng/ml, and 6 adult men (age > 45 years) with negative family history of KP, negative digital rectal examination < 2.5 ng/ml, stable over time. Each subject reviewed an information sheet and provided a written consent for the KP VOC test. The study was approved by the Ethical Committee at Humanitas Clinical and Research Center, where patients were treated (Approval no. CE-ICH260/11). No exclusion criteria regarding medical history, hormonal status, diet, clinical treatments, or tobacco consumption were considered for the selection of participants. Although some of these could be hypothesized as confounding factors for the olfactory discrimination, previous researches (Biehl et al., 2019; Guest et al., 2021; Kokocińska-Kusiak et al., 2021; Mazzola et al., 2020) demonstrated that, if well-standardized protocols are involved for recruitment, conservation of urine samples and analysis, they do not affect the classification performance.

Urine samples from both healthy young and adult males (control group) and KP patients were collected at the Humanitas Mater Domini Hospital in Castellanza (VA). Immediately after collection, samples were frozen to prevent alterations due to bacterial activity. Then, frozen samples were transported to the Department of Chemistry, Materials and Chemical Engineering “Giulio Natta” and stored until use. Clinical information, including clinical and pathological stage (The clinical stage is defined considering the results the Gleason score.) (The pathological stage is based on the evaluation of the entire prostate removed with prostatectomy. It provides information about tumour location and extension.), pathological Gleason score (GS) (The Gleason score provides information about the type of cancer cells constituting the tumour mass and it is assessed by means of a histopathological evaluation of the prostate tissue.), PSA serum levels and concomitant pathologies were provided whether available.

METHODS DETAILS

Electronic nose

EN involved for the study is a lab-scale prototype developed at the Politecnico di Milano, equipped with six n-type doped metal oxide semiconductor (MOS) sensors produced at the Department of Chemistry, Materials and Chemical Engineering of the Politecnico di Milano by inkjet printing (Bax et al., 2021a; 2021b). During exercise, the MOS sensors were maintained at a constant temperature of about 400°C by the 5 V powered Pt heater, and the resistance of the metal oxides active layers was acquired continuously by means of a custom-made circuitry at a frequency of 1Hz and recorded for further processing.

The sensor chamber is made of stainless steel and has a parallelepiped shape with length and depth of 20 cm, height of 3 cm, for a global volume of 60 cm³. The sensor chamber is equipped with 2 holes on the lateral sides, which served as inlet and outlet of the odour sample.

For the analysis, the urine headspace bag is connected to one of lateral holes of the chamber, and a constant flowrate of 50 cm³/min is sucked into the chamber by means of a vacuum pump. Then, odourless air, obtained by filtration with an activated carbon filter, is sucked into the chamber to restore sensors' signals for the next measurement session.

Experimental protocol

The standardized experimental protocol involved for sample preparation and analysis by EN, developed within the project and detailed in the European Patent EP 19160856.1 filed in March 2019, consists of 4 steps (Figure):



Figure. Standardized experimental protocol involved for urine sample preparation and analysis

- Thawing: urine samples, stored at -18°C , are thawed in a water bath at about 40°C .
- Urine headspace enrichment: 10 mL of liquid urine are put in a NalophanTM bag filled of odourless air and conditioned at 60°C and 20%RH for 1 h to favour the enrichment of the gaseous phase with urine volatiles.
- Urine headspace conditioning: the gaseous phase is separated from the liquid and conditioned at 60°C and 20%RH for 1.30 h to reduce the moisture content of the headspace before the EN analysis.
- EN analysis: the urine headspace is analysed at a fixed concentration, recording the variations of resistance related to absorption, i.e. *During*, and desorption, i.e. *After*, of VOCs on the sensors surface. The analysis lasts 80 min: 50 min *During* followed by 30 min *After*.

Data processing

In addition to sample preparation and analysis, data processing constitutes the most critical step for the development of an instrument capable of detecting, identifying, and measuring volatile compounds. The data processing pathway involved in the study consisted of 4 steps (as schematized in Figure): signal pre-treatment, feature extraction, drift compensation and pattern recognition. The following subsections provide detailed description of each step.

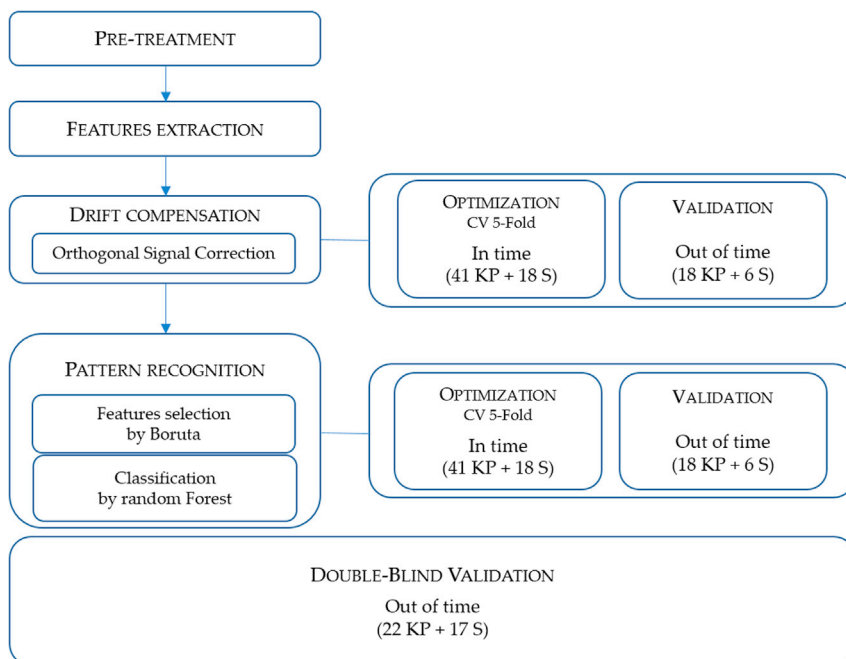


Figure. Data processing pathway involved for the study

Pre-treatment

The raw resistance curves recorded by gas sensors of the EN array during the urine headspace analyses were processed by Standard Normal Variate (SNV) (Zeaiter et al., 2005) for compensating baseline shift among urine headspace analyses carried out on different days, most probably related to external factors (e.g., oscillation of environmental temperature or humidity) not useful for sample classification. To do this, each spectrum was centred and scaled by its standard deviation, according to the following equation:

$$x_{ij}^* = \frac{x_{ij} - x_i^{\text{mean}}}{\sqrt{\sum_{j=1}^n (x_{ij} - x_i^{\text{mean}})^2 / n - 1}} \quad (\text{Equation 1})$$

where x_{ij} is the original intensity in the i^{th} row and j^{th} column of the spectral matrix, x_{ij}^* is the intensity after pre-processing in the i^{th} row and j^{th} column of the spectral matrix, x_i^{mean} is the mean value of the i^{th} spectra and n is the number of variables (line intensities) of each spectrum, respectively.

This pre-treatment allowed compensating for baseline shift among urine headspace analyses carried out on different days, thereby resulting in a better discrimination among samples from the Control S and KP Groups (Figure).

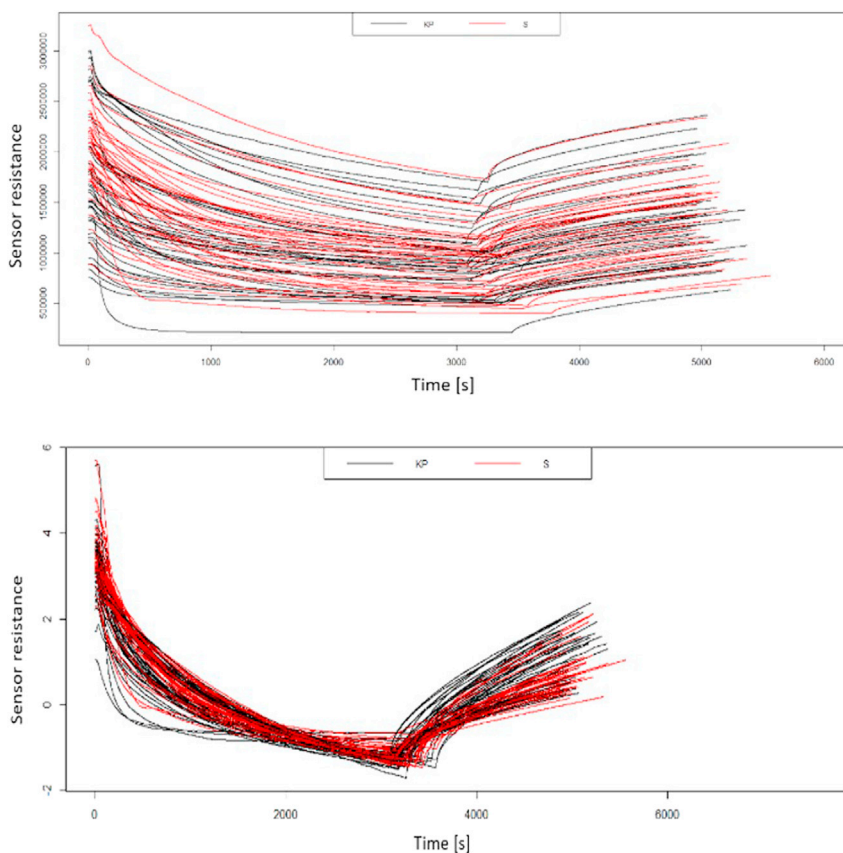


Figure. Resistance curves relevant to the analyses of urine headspaces from control and KP groups before (top) and after (bottom) SNV correction

KP, Prostate cancer; S, Control.

Features extraction

After pre-processing, a pool of features, representative for both plateau conditions and transient EN responses relevant to urine headspace analysis, was extracted from each sensor response curve for further processing. Table 1 summarizes features included in the urine headspace dataset. Specifically, it reports the equation and the parameters used for features calculation. 25 features were extracted from each sensor response, therefore the resulting dataset consisted of 150 variables.

The feature set was then autoscaled prior further processing.

Drift compensation

The orthogonal signal correction. Drift currently represents one of the main obstacles for EN to switch from research objects to large-scale diagnostic devices, since the progressive worsening of their classification performance over time inhibits their use for long-term applications by a non-technical staff.

Drift is mainly related to physical changes of the sensor active layer, very likely induced by thermal stress, poisoning or environmental contamination, occurring over sensor exploitation (Di Carlo and Falasconi, 2012).

Several methods for drift compensation, which tackle the problem from different perspectives depending on the specific application, have been proposed in the scientific literature (Di Carlo and Falasconi, 2012):

- Pre-processing techniques, as Baseline manipulation (Gardner and Bartlett, 2000) and Frequency domain filtering (Hui et al., 2003; Llobet et al., 2002), have been proposed for a first correction of

the raw sensor signals aimed at the removal of noise related to external factors (e.g., environmental conditions), which results also in a compensation of drift.

- Periodic calibration methods, such as Multiplicative drift correction (Fryder et al., 1995) and Component Deflation (Gutierrez-Osuna, 2000), estimate the drift direction to be removed from data by monitoring changes in sensor response to one or more reference standards (i.e., calibrants), that are analysed periodically with equal intervals of time.
- Attuning methods, as Independent component correction (Di Natale et al., 2002) and Orthogonal signal correction (OSC) (Padilla et al., 2010), perform component correction without resorting to the use of calibration samples, but deducing drift components directly from the training data.
- Adaptive method, as Neural networks (Distante et al., 2002; Zuppa et al., 2004) and Evolutionary algorithms (Di Carlo et al., 2011), adapt the Pattern Recognition Model (PaRC) based on pattern changes due to drift effects that are deduced from training data. This allows for an increasing time validity of the PaRC model, which in turn reduces the request for calibration (Martinelli et al., 2013; Vergara et al., 2012).

The attuning methods have been largely applied in different fields, as diagnostic, biological or environmental, for both mitigating drift effects and improving system classification performances (Artursson et al., 2000; Laref et al., 2017; Padilla et al., 2010; Zhang et al., 2011). The Orthogonal Signal Correction (OSC), which has been already proved to act efficiently on gas sensors array (Laref et al., 2017; Padilla et al., 2010), was selected for the specific application thanks to its several advantages: it does not require the use of a calibrant, it does not need excessive large training data, and, above all, it removes from the original data only information that are not relevant for class discrimination. More in detail, this technique, firstly proposed by Wold et al. (Wold et al., 1998), removes from the data matrix X all the information not correlated to a vector (or matrix) Y containing information relevant for classification (e.g., concentration, toxicity, belonging class etc ...) by imposing the condition of orthogonality between X and Y.

The procedure for data correction can be summarized as a sequence of the following operations, where X is the matrix relevant to data to be corrected and Y the matrix providing extra data information:

1. Centring and scaling data
2. $M = I - \frac{X^T Y Y^T X}{Y^T X X^T Y}$
3. $Z = XM$
4. $PCA(Z) \rightarrow p$
5. $w_{\perp} = p$
6. $t_{\perp} = Xw_{\perp}$
7. $p_{\perp} = \frac{X^T t_{\perp}}{t_{\perp}^T t_{\perp}}$
8. $X_{OSC} = X - t_{\perp} p_{\perp}^T$

Specifically, Tom Fearn's algorithm (TFosc) starts by centring and scaling both X and Y matrices (step 1), and a matrix Z, orthogonal to X^TY, is calculated (step 2 and 3). Then, a PCA is performed on Z (step 4) to assess the loading vector p to be used as weight vector w_⊥ (step 5). In steps 6 and 7 the final score and loading vectors t_⊥ and p_⊥ are calculated by PCA, and in step 8 the correction is performed (Wang et al., 2017).

Concerning new samples correction, new data (i.e., test) should be centred and scaled as training data X. Then, using the orthogonalized weights w_⊥ obtained on training set, the passages from 6 to 8 are repeated, substituting X with the test matrix.

Implementation and validation. A specific model based on Orthogonal Signal Correction (OSC) algorithm by Tom Fearn (Fearn, 2000) was implemented and applied to data collected by the same EN sensor array over 9 months.

A urine headspace dataset, comprising 83 samples, was divided into a train set (i.e., 59 oldest analyses carried out in months 1–8) and a test set (i.e., 24 more recent data acquired in months 8 and 9), in order to implement the drift correction model and test its performance on independent data. The proportion between healthy subjects and KP patients has been kept constant in train and test sets: control subjects represented about the 30% of the samples both in train and test sets.

The OSC model was applied to the $n \times 150$ feature matrix (X) to remove the variance orthogonal to a binary $n \times 2$ matrix (Y), containing information of the belonging class of each analysis. The Y matrix is a binary matrix of two columns, representing the KP and Control Group respectively, whose rows report 1 at the corresponding column that designate the class membership, while the other terms of the row are set to 0.

To select the optimal number of OSC components to be removed for compensating drift, the classification performance achieved on 5-Fold Cross Validation on the train set and external validation on the test set by drift correction model including 1 to 15 OSC-components was compared. The number of components that maximized the classification performance was chosen for building the correction model. The classification performance of a diagnostic tool was assessed in terms of sensitivity, specificity, and accuracy. In general, sensitivity and specificity express the capability of the device to correctly classify KP patients and controls, respectively. Conversely, the accuracy of the system expresses the rate of correctly predicted samples out of all the analyses considered.

Pattern recognition

Features selection. To reduce data dimensionality and identify the most significant features for best describing differences between samples from the Control and KP Groups, a features selection model based on Boruta algorithm was implemented. Boruta algorithm is a wrapper method (Bolón-Canedo et al., 2013; Muezzinoglu et al., 2009), which lists features included in the dataset according to their information gain, measuring the importance of features for classification purposes. More in detail, the algorithm creates for each feature a so called “shadow attribute”, consisting in the randomly shuffled value among measures. In this way, there is no more correlation between the variable value and the belonging class of the samples, and the importance of these “shadow attributes” can be non-zero only due to random fluctuations. By comparing the classification performances of each feature to the one of the best “shadow attribute” it is possible to identify which features are truly significant for improving samples classification. Features achieving an information gain significantly higher than the best of the “shadow attributes” were selected as important (Kursa and Rudnicki, 2010).

The Boruta algorithm was implemented fixing a maximum number of runs equal to 350. Given the unbalanced number of samples in the Control and KP Groups and the scarce number of the controls, with the purpose of ensuring the stability of the feature selection results, we repeated 15 times the algorithm imposing an equal number of samples from the Control and KP groups in the training dataset.

Classification. Finally, in order to build the KP diagnosis model, the Random Forest (RF) algorithm, proposed by Breiman et al. (Breiman, 2001), was properly trained. RF is a popular ensemble method that can be used to build predictive models for both classification and regression problems.

To provide a classification of unknown samples, RF creates an entire forest of random uncorrelated decision trees that are used to list data belonging to different classes in a dataset (Liaw and Wiener, 2001). Specifically, the RF algorithm splits the initial dataset into two subsets: the “Bootstrap Dataset” (BD) and the “Out Of Bootstrap Dataset” (OOB). The BD constitutes the first tree of classification forest with randomly selected samples of the original dataset. To build the tree, data have been splitted at each node, using the feature providing the best classification performance at dividing samples by class of belonging. The choice of the feature is based on the comparison of the performance of various random variables, selected among all the variables present in the dataset. The tree stops growing when the last node has a worst classification performance of the samples than the previous one. Then, the OOB set, including samples of the original dataset not considered to build the classification tree, is used to test its classification performance. This operation is repeated many times to build the entire forest. Once the forest has been created, the model can be used to classify samples from an independent dataset, whose classification is based on the majority of vote of trees in the random forest (Breiman, 2001; Liaw and Wiener, 2001).

For the specific application, the RF was trained by imposing the number of tree to be used in the forest (ntree) equal to 500. This value allowed achieving satisfactory classification performance whilst maintaining low computational time.

The same validation scheme involved for the development of the drift compensation model was used to tune the RF model. In this case, the 5-Fold cross validation focused on the definition of the optimal number of variables to be considered at each node of the tree (mtry). Based on results achieved by comparing mtry from 1 to 15, an mtry equal to 5 was selected for the prediction model.

Double-blind verification tests

In order to provide an unbiased validation of the developed drift correction model, double-blind validation tests were performed. For this purpose, 39 urine samples collected from control subjects and KP patients were provided without any information about their clinical status. They were analysed in the last two weeks of month 9 and classified according the path summarized in Figure.

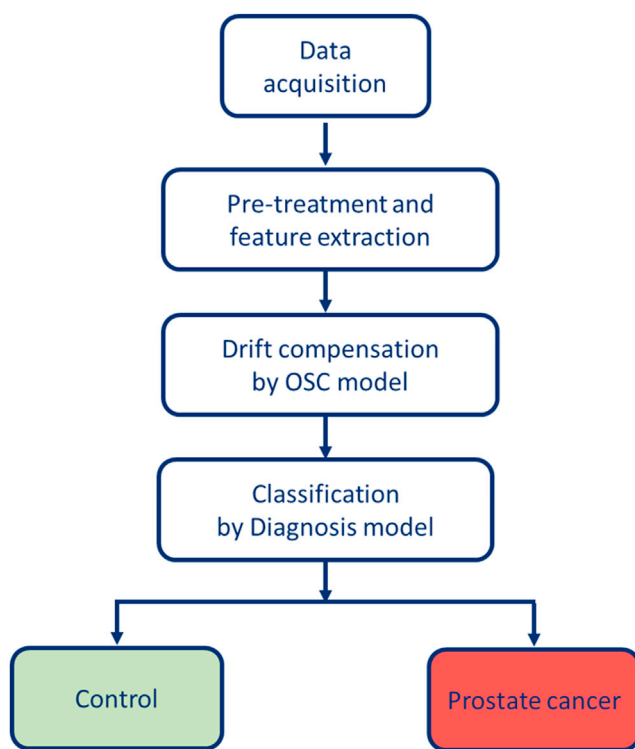


Figure. Block flow diagram of blind sample analysis

The results of the classification by diagnosis model were organized in a confusion matrix for evaluating the diagnostic capability of the proposed innovative tool, which was assessed in terms of Accuracy, Specificity, Sensitivity, Negative Predicted Value (NPV) and Positive Predicted value (PPV).

ADDITIONAL RESOURCES

The study was approved by the Ethical Committee at Humanitas Clinical and Research Center, where patients were treated (Approval no. CE-ICH260/11).