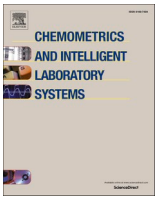
Chemometrics and Intelligent Laboratory Systems 234 (2023) 104752

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics



An analytical approach based on excitation-emission fluorescence spectroscopy and chemometrics for the screening of prostate cancer through urine analysis: A proof–of–concept study

Eleonora Mustorgi [a], Caterina Durante [b,*], Cristina Malegori [a], Piergiorgio Greco [c], Riccardo Bartoletti [c], Marina Cocchi [b], Monica Casale [a]

[a] Department of Pharmacy, University of Genoa, Viale Cembrano, 4 – 16148 Genoa, Italy
[b] Department of Chemical and Geological Sciences, University of Modena e Reggio Emilia, via Campi 103, Modena, Italy
[c] Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, via Savi, Pisa, Italy

ARTICLE INFO

Keywords:
Prostate cancer
Excitation-emission fluorescence spectroscopy
Cancer marker
PARAFAC

ABSTRACT

In the present feasibility study, excitation-emission fluorescence spectroscopy has been investigated, as a rapid and accurate analytical method for the development of a tentative model for the early screening of prostate cancer directly through urine analysis in order to provide reliable results while improving patient compliance.

Sixty-nine urine samples (46 samples from patients with histologically proven prostate cancer and 23 from healthy donors) were provided, by the University of Pisa, Urology Unit. The excitation-emission fluorescence measurements were performed on centrifugated urine samples at room temperature on a Perkin-Elmer LS55B luminescence spectrometer and the corresponding data array was analysed with parallel factor analysis (PARAFAC).

From a synergistic analysis of the obtained results, four main fluorophores, corresponding to four selected PARAFAC factors, were recognizable in the urine excitation-emission matrices (EEMs) and the respective species could be potential markers in the differentiation among healthy and cancer samples. PARAFAC results, in terms of extracted scores, coupled with discriminant algorithms, allowed to develop a first attempt of healthy/cancer discrimination model. The chemometrics models show promising correlation between some of the depicted fluorophores and the disease state. However, considering the limited cohort (not only in terms of number but also of representativeness), this study must be considered as a proof of concept; a more sound and statistically relevant sampling must be performed in order to consider the confounding factors in the cohort treated and to develop an analytical approach applicable in real scenarios.

1. Introduction

Prostate cancer epidemiology is raising in the last few years likely due to both the increasing age of male patients and the technical innovations related to its early diagnosis. Prostate cancer mortality is second to the lung cancer mortality with 358.989 new cases in 2018 [1].

Prostate cancer is more frequent in male patients over 60 years and due to this reason early diagnosis is often not easy due the overlap of two different diseases on the same organ. The differential diagnosis has to be determined between benign prostate hyperplasia (BPH) and cancer (PCa) which affect two different zones of the same gland and can be also considered as co-existing diseases. Higher prostate specific antigen (PSA) levels may be generated by both the diseases because influenced also by prostate gland volume and patient age. To date, the protocol adopted for PCa early clinical diagnosis includes the digital rectal evaluation, the PSA levels determination and the multi-parametric Magnetic Resonance Imaging (mpMRI) test, although the certain disease identification and characterization is obtained only by prostate biopsy. Prostate biopsy is an invasive method that needs to be performed with local or general anaesthesia and shows several limitations such as the presence of co-morbidities, the medical treatment with anti-platelet therapies, the risk of developing infections and sepsis [2]. Due to these reasons the perspective of a "liquid biopsy" seems to be captivating and prospectively useful for the daily clinical practice.

* Corresponding author. University of Modena and Reggio Emilia, Via Campi 103, Modena, Italy.
E-mail address: caterina.durante@unimore.it (C. Durante).

https://doi.org/10.1016/j.chemolab.2023.104752
Received 21 November 2022; Received in revised form 4 January 2023; Accepted 5 January 2023
Available online 10 January 2023

Based on the latest scientific evidence, it is of current interest to have rapid and accurate analytical methods for early screening of prostate cancer directly by urine analysis, in order to provide reliable results while improving patient compliance. Indeed, several studies showed the richness of information available in the human urine able to support cancer detection [3–5]. As regards the development of fast and accurate model, non-destructive fingerprint spectroscopy methods, such as Fluorescence Spectroscopy, have proved to be extremely efficient for the analysis of biological fluids, including urine, in the clinical context [6,7].

In 2010, Masilamani et al. [8] showed the results of a novel study in which the native or intrinsic fluorescence of urine was used to aid the diagnosis of several types of cancer. In their study fluorescence emission spectra and Stokes shift spectra of the first voided urine samples were acquired for 100 healthy controls and those of 50 cancer patients of different aetiology. They concluded that flavoproteins and porphyrins released into urine can act as generic biomarkers of cancer with a specificity of 92%, a sensitivity of 76%, and an overall accuracy of 86.7%. A weak point of that study was that authors performed the statistical analysis to discriminate diseased patients from healthy patients using only seven a priori selected descriptors (ratios of intensity peaks) without considering, and benefiting from, the whole spectral information embodied into the fluorescence spectra.

In 2013, a study conducted by Zvarik et al. [9] dealt with assessing the differences in terms of metabolites, which can be detected by excitation-emission fluorescence, in urine samples of patients affected by ovarian cancer compared to healthy volunteers. They observed changes in the spectral profiles that were interpreted as reduction of pyridoxic acid content, whereas blue-fluorescing pteridines became dominant in urine samples of cancer patients with respect to healthy donors. Thus pteridines, which are related to cellular metabolism, could be suitable candidates for neoplasia-associated fluorescent markers in human urine. The observed changes in intrinsic fluorescence were studied by plotting as images (intensity represented by colour coding) the three-dimensional fluorescence excitation-emission landscapes, where the characteristic circular patterns, highlighting high emission, were used to identify particular wavelengths regions which could be attributes to specific fluorophores. Also in this case, a chemometrics approach to extract the information from the whole excitation-emission landscapes for all samples (i.e. the raw measured complex analytical data) was not attempted.

In the same year, Rajasekaran et al. [10], studied native fluorescence characteristics of human urine samples using excitation–emission matrices (EEMs) over a range of excitation and emission wavelengths in order to discriminate patients with cancer from normal subjects. A total of 80 urine samples from normal subjects and 90 from pathologically confirmed cancerous patients were collected and analysed. EEMs were acquired in the following ranges, 250–450 nm for excitation and 270–750 nm for emission. However, only the spectra corresponding to fluorescence emission at 405 nm have been considered both for visual spectral comparison and to perform the discriminant analysis of normal form cancerous subjects. In more detail, stepwise multiple linear discriminant analysis was performed by the authors considering 19 ratio variables calculated using fluorescence intensities at emission wavelengths which represent characteristic spectral features of different groups of subjects, at 405 nm excitation.

In the present study, for the first time, excitation-emission Fluorescence Spectroscopy, has been investigated as a rapid and accurate analytical method for the development of a tentative model for the early screening of prostate cancer directly through urine analysis in order to provide reliable results while improving patient compliance.

A further element of novelty, with respect to this state or art, consists in processing the whole EEM landscapes with a suitable multiway analysis method, i.e., parallel factor analysis (PARAFAC) which allow resolution of the spectral profiles corresponding to single fluorophores and their relative concentration estimation [11]. Indeed, the resolved profiles could be associated to chemical compounds (metabolites) that were recognised to have a role in oncological pathologies in literature. Thus, could be suggested as potential markers of prostatic oncological pathologies and subject to further investigation.

2. Materials and methods

2.1. Patient information

69 urine samples provided by the Center of Urology University Hospital Cisanello of Pisa (Italy), were analysed; the samples belong to the following categories.

- 46 samples of patients whose prostate biopsy and subsequent histological examination have diagnosed a malignant prostate cancer;
- 23 control samples (healthy donors).

Detailed information of patients is shown in Table 1.

All the patients candidate to prostate biopsy had at least one clinical (raised Prostate Specific Antigen levels and/or induration/s at Digital Rectal Examination) and/or radiological suspicion of prostate cancer at multiparametric MRI test. Multiparametric MRI test was considered as suspected for prostate cancer just in case of PIRADS V2 score $\geq$4 [12].

Patients with histologically proven prostate cancer were classified according to the Gleason score scale and the D'Amico classification as low, intermediate and high risk prostate cancer [13,14].

2.2. Sampling

Urine samples were taken with a sterile procedure after pathological diagnosis of prostate cancer and before any kind of medical and/or surgical treatment. Urine samples were immediately frozen at −80 °C until analysis and stored according to hospital protocol; then, the samples were transported to the analytical laboratory in homologated packaging with inside dry ice pellets at −80 °C and they were stored in a special cold room at −80 °C.

Before being analysed, the samples were kept for 17 h in the refrigerator (3–5 °C) and for 1 h in the thermostat room of the laboratory at a temperature of 20 °C; since the samples in many cases had sediment, they were centrifuged for 30 min at 3000 rpm before analysis.

Some urine samples were analysed in replicate to assess the repeatability of the method, and therefore the final number of EEM spectra that have been processed were 66 from cancer patients and 29 from healthy donors.

2.3. Instrumental settings

The EEM fluorescence measurements were performed on centrifugated urine samples at room temperature on a Perkin-Elmer LS55B luminescence spectrometer (Waltham, MA, USA). The excitation-emission matrices of the urine were recorded using the standard cell holder in a 10 mm quartz SUPRASIL® cell with cell volume of 3.5 mL by PerkinElmer. The excitation spectra were recorded between 250 nm and 530 nm each 5 nm (29 recorded points), whereas the emission wavelengths ranged from 270 nm to 650 nm each 0.5 nm (761 points). The

Table 1
Patient characteristics.

Patients n.°	46
Median Age (±SD)	69.5 (±7.8)
Patients with positive DRE	31
Patients with PIRADS 4–5 at mp MRI	28
Median PSA levels ng/ml (±SD)	8.7 (±2.9)
Number of core biopsies per patient, median value (±SD)	16 (±1.3)
Patients with pathological low risk PCa (Gleason score ≤6)	5
Patients with pathological intermediate risk Pca (Gleason score 3 + 4)	19
Patients with pathological high risk Pca (4 + 3 and more)	17

excitation and the emission monochromator slits were set to 10 nm. The FL WinLab software (PerkinElmer) was used to register the fluorescent signals.

2.4. Data analysis

2.4.1. Data sets

The acquired EEM landscapes were arranged in a three-way data array of dimensions 95 (human urine samples) x 761 (emission wavelengths) x 29 (excitation wavelengths).

The data, prior to PARAFAC decomposition, were split in calibration (70%, i.e. 67 samples) and validation (30%, i.e. 28 samples) sets by using Kennard Stone Duplex algorithm [15] distinct per category in order to keep the same percentage of calibration and validation samples for each class (proportional splitting).

2.4.2. PARAFAC analysis

According to the specific nature of EEM data, organised in a 3D data array (sample × λ emission × λ excitation), Parallel Factor Analysis algorithm [16], was applied to model directly the n-way data [17].

PARAFAC decomposes a data tensor $\underline{\mathbf{X}}$ with dimension $I \times J \times K$ into three loading matrices A, B and C. The trilinear PARAFAC model is expressed as follows:

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk} \quad i = 1, 2, \ldots, I;\ j = 1, 2, \ldots, J;\ k = 1, 2, \ldots, K$$

where x_{ijk} is an element of the $\underline{\mathbf{X}}$ array in position *i, j, k* and e_{ijk} is its residual value. F is the number of PARAFAC factors. In the case of fluorescence data, $\mathbf{a_f}$, $\mathbf{b_f}$ and $\mathbf{c_f}$ are vectors holding for each of the *f*-th fluorophore (factor) concentration (in each sample), excitation and emission profiles, respectively [18].

EEM data were first preprocessed in order to minimize the non-relevant artifacts. In particular, Rayleigh scatter, normally present in this kind of data [19], was removed using a first and a second order Rayleigh filters (half-width: 20 nm) and replaced with interpolated data. Zeros were assigned to sub-Rayleigh wavelengths [20]. Then, filtering in the second mode (eleven points window) and despiking were applied.

In order, to select the proper number of PARAFAC factors four parameters have been considered: the total variance explained by the model (fit); the core consistency; the similarity of fit and core consistency for replicate runs, i.e. PARAFAC has been restarted 5 time for each model dimensionality explored (from 1 to 6); and, the congruence of mode 2 and mode 3 loadings profiles obtained by split half analysis. Split half analysis consists in dividing the data set in two parts with respect to samples mode and calculating a distinct PARAFAC model on each, the congruence is then estimated as the covariance between corresponding loadings of the two halves. These criteria are implemented in the PLS Toolbox (see software section), namely *nvalidate* function.

2.4.3. Discriminant analysis

Both Linear discriminant analysis (LDA) [21] and Partial least squares discriminant analysis (PLS-DA) [22] were applied in order to discriminate cancer patients from healthy donors using as independent variables the first mode scores (relative concentrations) of PARAFAC model. The number of PLS-DA components was chosen according to minimum root mean squares error in cross validation (venetian blind with five splits).

2.4.4. Software

Data were imported to Matlab v. R2019a (The MathWorks, Inc., Natick, MA, US). PARAFAC and PLS-DA were performed using PLS Toolbox v. 8.9 (Eigenvector Research, Inc., Manson, WA, US). Linear discriminant analysis was performed by using the *fitcdiscrim* and *classify* functions of the Statistical and Machine Learning Toolbox of Matlab.

3. Results

3.1. EEM landscape analysis and fluorophores resolution

The analysis of EEM data shows that interesting differences in urine fluorescence excitation/emission spectra from patients with prostate cancer in comparison to healthy subjects can be already appreciated starting from the collected data, after a pre-processing step applied to remove the interference from Rayleigh scattering (as described in *Section 2.3.2*). As an example, two of the fluorescence excitation-emission landscapes of urine matrices coming from a healthy person (H16 sample) and a prostate cancer patient (P03) are presented in Fig. 1a and **b**, as contour plots.

From Fig. 1, it is quite evident the presence of strongly overlapping fluorescence bands in both EEM landscapes; however, some differences between the two samples could be highlighted. Firstly, in the healthy sample spectrum, two main bands could be visualized at 390 (excitation)/490 (emission) nm and at 470/510 nm with a shoulder at 490/520. These characteristics were observed in almost all the urine samples coming from healthy controls (Figure A1, Supplementary Materials). Conversely, there is a fading of the peak at 490/520 nm in prostate cancer sample, coupled with the emergence of the band at 350/420 nm. Furthermore, the reported cancer sample landscape is characterized by a depression of fluorescence at 390/490 nm with respect to the one of healthy sample. However, not all the urine EEM landscapes from the 46 cancer patients showed the same trend (Figure A2a and A2b, Supplementary Materials). Due to the complexity of the matrix, it is very hard to decipher the different emission bands and the presence of overlapping bands. Therefore, a multiway resolution method, i.e. PARAFAC, has been applied, benefiting from the second order advantage, in this way a clearer interpretation of the several overlapped bands could be achieved.

3.2. PARAFAC analysis

The EEM data were arranged in a three-way data array of dimensions I x J x K, where I is the number of investigated samples (i.e. the 67 human urines samples forming the calibration set), J the number of emission wavelengths (761 points) and K number of excitation wavelengths (29 points). Before decomposition by PARAFAC, the data were appropriately preprocessed in order to minimize the non-relevant artifacts (as reported in *Section 2.3.2*). A four-factors PARAFAC model was selected according to criteria based on residuals, core consistency and split-half analysis [23], indeed four different fluorophores were detected in the investigated urine samples.

The results from the PARAFAC model are reported in Figs. 2 and 3. In particular, Fig. 2 shows the emission (mode 2 loadings) and Fig. 3 the excitation (mode 3 loadings) profiles of the four resolved factors in urine samples. These two modes represent the underlying pure spectra of characteristic fluorophores present in the investigated urine samples and which can be putatively identified based on literature.

The excitation/emission wavelengths corresponding to the maximum fluorescent intensity of the first factor (blue) is 360/440 nm. According to several authors, the band at around 370/440 nm could be ascribed to one or more chemical species, such as pteridines [9,24]. These compounds could play an important role in the synthesis of some vitamins, as well as they are important intermediates in anabolic and catabolic reactions. Furthermore, Masilamani et al. [8] reported that the 444 nm band in emission could be due to NADH bound to a protein, even if it could be difficult to envisage the presence of this molecule in healthy urine samples [9]. Although shifts are present with respect to the reference for both NADH and pteridines, they could be acceptable since it is well known that the fluorescent emission signal from a fluorophore can be strongly dependent on the surrounding environment [25,26].

The second (red) factor has excitation and emission maxima at 330 and 420 nm, respectively, and could be attributed to several

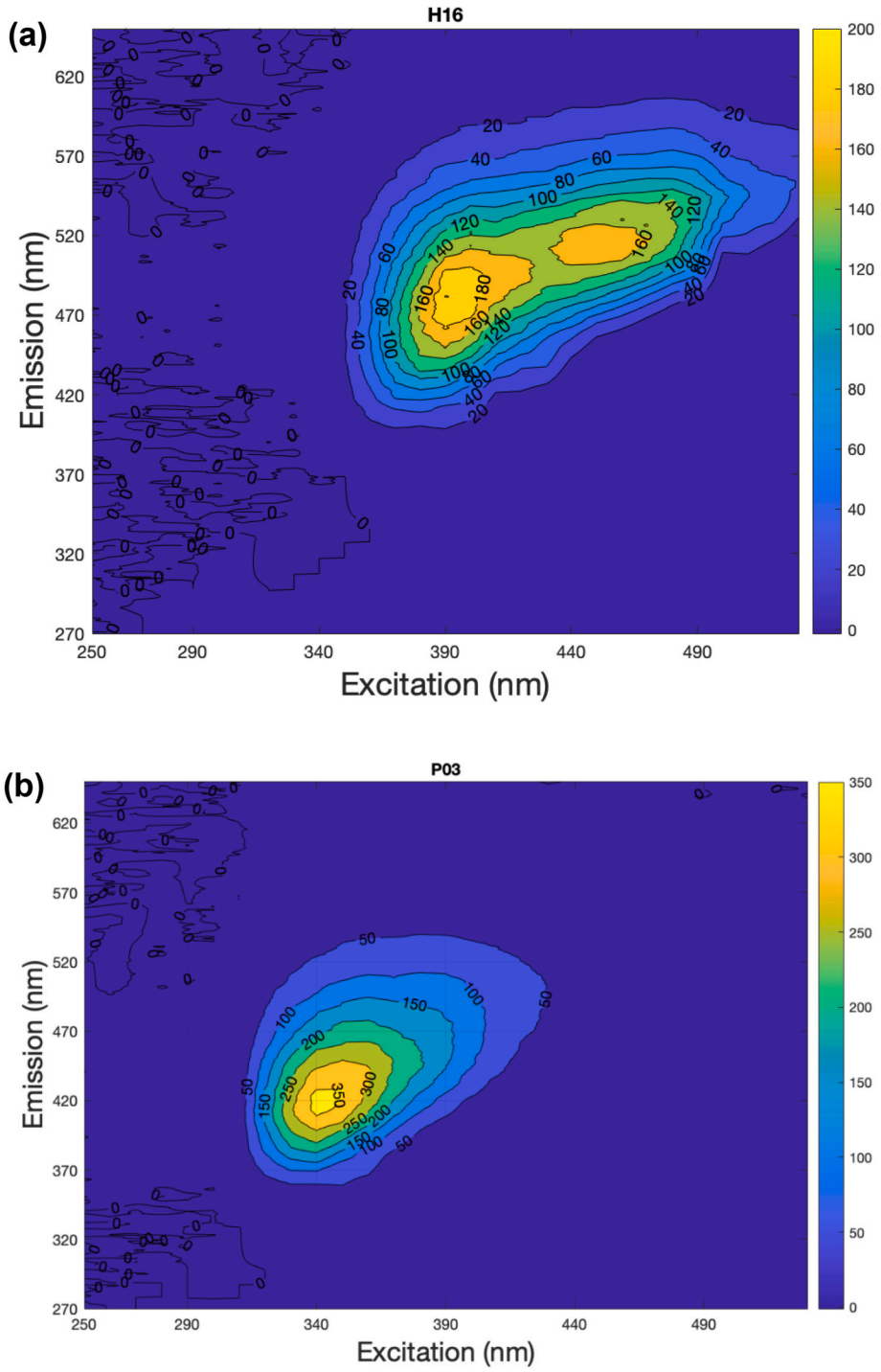


Fig. 1. Excitation-emission contour plot of a representative human urine sample from (a) a healthy person and (b) an oncological patient.

fluorophores such as pyridoxic acid and uric acid [9]. In particular, pyridoxic acid is excreted in the urine as a catabolic product of vitamin B6 and it is involved in many enzymatic reactions as pyridoxal-phosphate active form of vitamin B6.

The excitation/emission maxima positioned at 390/470 nm characterizes the third factor. Free NADH has an excitation/emission maximum at around 400/485 nm [8], hence, considering a potential shift of the emission when the polarity of the microenvironment changes, this band could be tentatively associated with the third PARAFAC factor. Furthermore, from a closer inspection of the shape of this factor, it is worth noticing a little, hardly visible "bump'' with emission maximum at 620 which is in agreement with literature values for porphyrins [8].

Finally, excitation/emission loadings (450/530 nm) of the fourth factor could fit well with excitation/emission of flavins and their metabolites [9,24]. In particular, riboflavin presents an excitation/emission maximum at around 450/550 nm [27]. Furthermore, Masilamani et al. [8] reported that bilirubin could also contribute to the fluorescence band at these wavelengths and its band could overlap in this region. However, flavin is more fluorescent than bilirubin, therefore, the latter could not apport a different and prominent contribution.

The role of the four PARAFAC factors to distinguish healthy and prostate cancer samples can be inspected by the first mode loading plots shown in Fig. 4a and b.

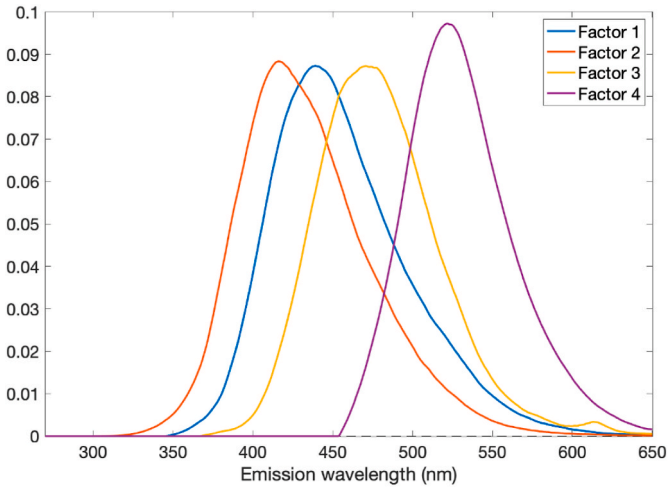


Fig. 2. Emission loading vectors from four factors PARAFAC model.

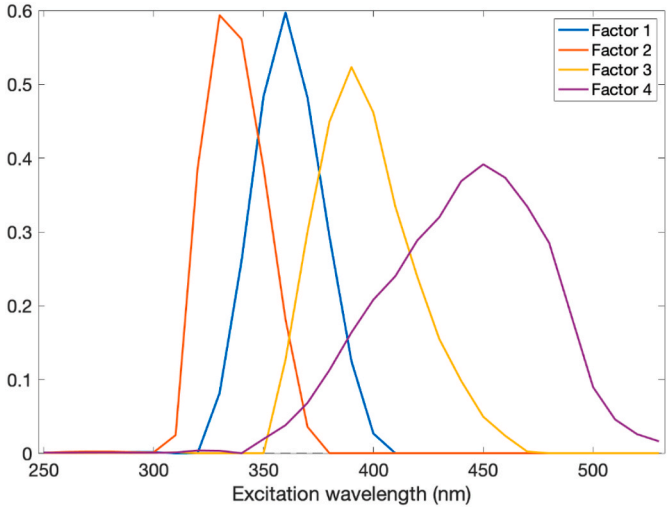


Fig. 3. Excitations loading vectors from four factors PARAFAC model.

In particular, Fig. 4a displays the first mode loading plot of the first vs. the second factors. Validation set samples were also projected on the model. First, a good reproducibility within the different replicates (same number in the label) is observed along both the factors, on the other hand the samples spread is not related to a clear cancer/non-cancer distinction. Rather, a differentiation in two groups arises: one heterogeneous group, includes almost all cancer samples that, together with samples H14, H15, H17 and H18 belonging to healthy controls, and it lies at high values for both factors, i.e., pteridines and pyridoxic acid; the second one, is formed by several samples positioned at factor 2 (pyridoxic acid) values, close to zero. Although no prior information about the healthy status of the patients has been used for building the PARAFAC model (i.e., it is an unsupervised method), it was possible to highlight a partial separation between healthy and diseased individuals (Fig. 4b) considering the third and fourth PARAFAC factors. In particular, it is worth noting an increase of both the third (putative free form of NADH) and the fourth (flavins) factors passing from urine samples of cancer patients to healthy donors. Four samples (H11, H14, H15 and H17) of healthy donors, nonetheless present an 'anomalous' trend and further investigation are ongoing to better understand their behaviour.

3.3. Classification results

The PARAFAC results discussed above showed that factors 3 and 4 are the most significative in differentiating between healthy volunteers and prostate cancer patients. Thus, linear discriminant analysis was applied considering factors 3 and 4 as descriptors to obtain a classification model. The results are shown in Fig. 5.

As concern the calibration set, four healthy samples are misclassified, three of these, H11, H15 and H17, have been already noticed in the exploratory analysis and discussed above; on the other hand, all the prostate cancer samples are correctly classified. For the test set, one healthy (H14, the fourth previously discussed) and one cancer (M33) are respectively misclassified. The overall performance of the LDA model is reported in Table 2.

For comparative purposes, a PLS-DA analysis has been also performed by considering all the four PARAFAC factors. The model was cross-validated using venetian blind with five splits, two latent variables were selected and the results in terms of sensitivity and specificity are reported in Table 3.

The area under the ROC curve (AUC in CV) is rather high (Table 3), however three of the healthy samples were wrongly assigned to cancer class and four of the latter were incorrectly predicted as healthy ones, both in fit and cross-validation. It has to be noticed that the three healthy samples accepted by the cancer class are the same H11, H15 and H17, misclassified by the linear discriminant model. The prediction capability improved in validation where only one sample per class was misclassified; finally, the healthy misclassified sample is H14 and the cancer one is the same one misclassified by LDA (M33). More trustworthy results could be surely obtained by enlarging the number of samples, anyhow the results are very consistent.

A biplot for the PLS DA model is reported in Fig. 6. As expected from the sensitivity and specificity values, there is an almost good separation between the two classes (except for four healthy samples).

In particular, there is a tendency towards differentiation along the first PLS-DA component, whilst on the second it is possible to note a difference in the intra cancer class which is split into two groups, the first one with positive second component scores and the second group with negative ones. In the biplot, it is also possible to see which variables are important for this separation. In particular, the samples that are positively correlated to the ''cancer direction'' present a lower amount for both factors 3 and 4 (i.e. free-NADH and flavins) with respect to the control samples. Finally, factors 1 and 2, pteridines and pyridoxic acid respectively, contribute to differentiate the two groups among cancer samples. However, almost all cancer samples, with both LVs negative scores, generally present a lower amount of the intensity bands for all the four fluorophores.

4. Discussion

In this study, excitation emission matrix fluorescence measurements on human urine combined with multivariate data analysis were proposed as a potential fast routine analysis to screen potential prostate cancer patients by urine sampling. Previous studies with fluorescence spectroscopy have shown optimal results of EEM in the investigation of cancer patients of different aetiology, but this work, to the best of author's knowledge, represents the first attempt in the discrimination of prostate cancer patients.

In the analysis of raw EEM, two different behaviours seem to be present among cancer patients: one characterized by one main high intensity band at around 350/420 nm and another characterized by a low intensity of the whole excitation/emission landscape. However, it was very hard to decipher the different emission bands due to the complexity of the matrix and the presence of overlapping bands. Therefore, a multiway resolution method, PARAFAC, was applied and a clearer interpretation of the highlighted bands was achieved.

PARAFAC results highlighted the presence of four fluorophores,

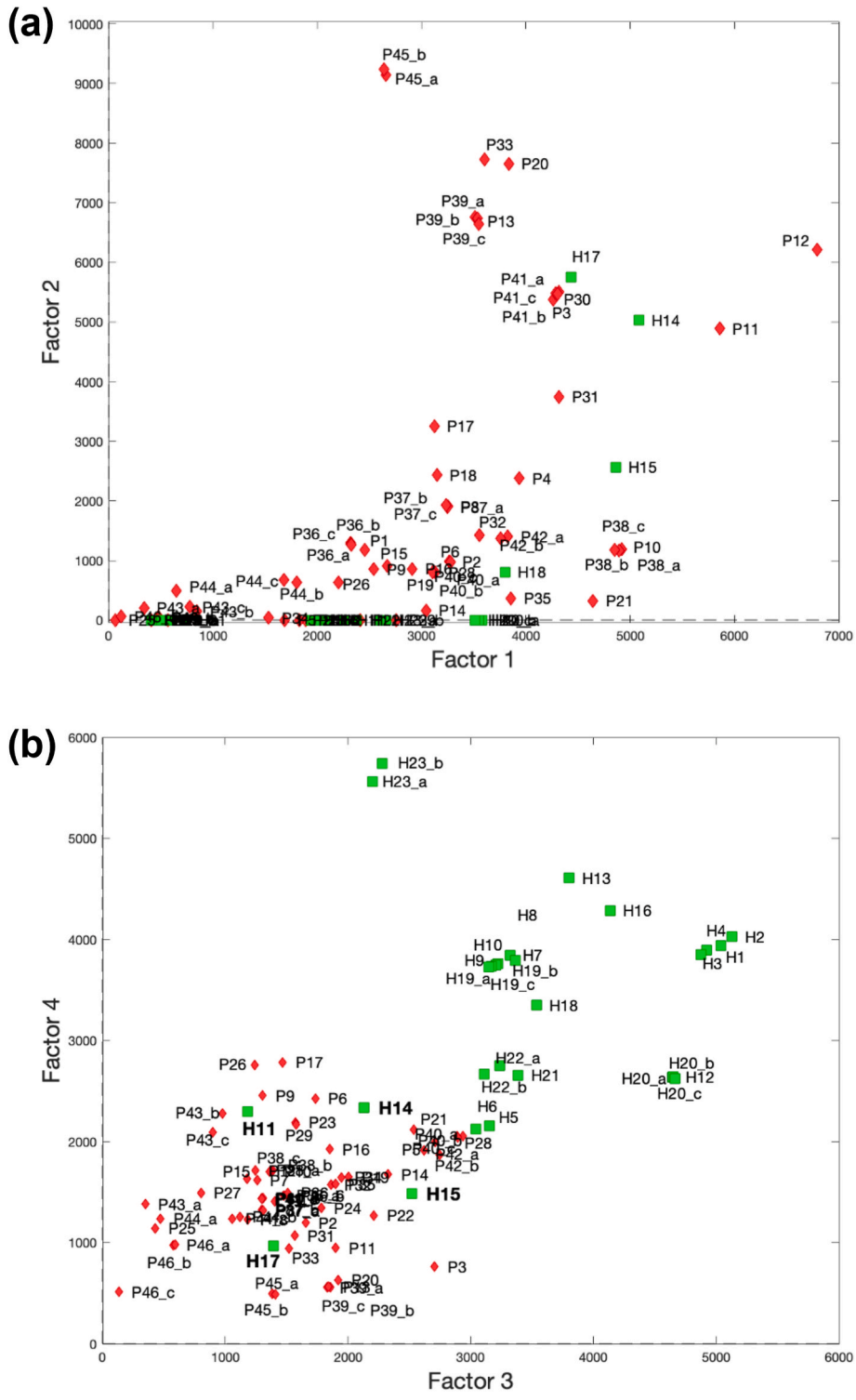


Fig. 4. Mode 1 PARAFAC loadings plot of (**a**) first vs. second PARAFAC factors and (**b**) third vs. fourth ones.

corresponding to the four PARAFAC factors, that could act as potential markers in the differentiation between healthy and cancer urine samples. By a comparison with the results of previous studies performed on urine, these could be putatively assigned as: pteridines and/or bounded NADH at 360/460 nm (excitation/emission), pyridoxic acid at 330/420 nm and, free-NADH and flavins, in the regions at 390/470 nm and 450/530 nm, respectively. The results suggested that the first two compounds could be characteristic only of a subgroup of cancer samples with higher concentrations, while the latter two contributed to some extent in the differentiation of healthy from cancer samples, which present lower values of concentration of both the fluorophores. Pteridines have been already shown to be an interesting neoplasia marker in human urine since their biosynthesis could be altered by the presence of malignant tumours which lead to a change of their concentration [9,28]. As regards pyridoxic acid, it is one of the catabolic products of vitamin B6 thus, its higher amount in cancer samples could reflect a decrease of the vitamin B6, whose concentration has been hypothesized to reduce cancer risk [29].

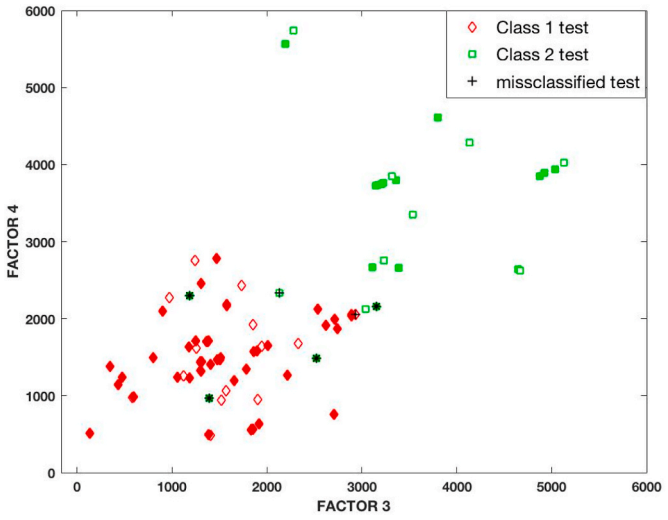


Fig. 5. Plot of mode 1 PARAFAC loadings: factor 3 vs. factor 4; filled circles indicate calibration set and empty triangles validation set. The asterisks and crosses indicate samples misclassified by LDA for calibration and validation set, respectively.

PARAFAC allowed enhanced interpretation of the results, thanks to its capability to furnish resolved factors chemically interpretable. The analysis can be further improved by using standard addition of putative analytes (species) identified in the characterization phase.

Moreover, PARAFAC scores, which represent the relative concentrations of the resolved species, when fed to LDA or to PLS DA, furnished a first evaluation of the capability to achieve healthy/cancer discrimination. The obtained results were extremely encouraging, since all the cancer samples were corrected classified and only four healthy samples were misclassified in calibration. . These samples, also pointed out in PARAFAC analysis, showed rather different EEM landscapes with respect to the others and should be further investigated. Finally, both LDA and PLS-DA cancer model well performed in prediction with 94,5% and 89,7% in Sensitivity and Specificity as well.

Furthermore, the classification results confirmed that cancer samples present a lower amount for both free-NADH and flavins with respect to the control samples. On the other hand, pteridines and pyridoxic acid contributed to differentiate the two groups among cancer samples.

5. Conclusion

In this paper, a proof-of-concept research is proposed, in order to provide preliminary evidences that excitation-emission fluorescence spectroscopy combined with multivariate analysis allows to obtain rapid and accurate analytical method for the early screening of prostate cancer. Indeed, this tool could be able to provide reliable results directly through urine analysis improving patient compliance.

However, the limited cohort prevents to be conclusive about the robustness of the obtained results. In fact, there might be hidden confounding factors in the cohort treated (like the age, gender, diet, alcohol and cigarettes intake, etc, etc.) that have to be considered for a solid conclusion.

Though further investigation is required, findings provide a promising step toward the long-term goal of translating important insights gleaned from this research into analytical strategies that can be applied to clinical population.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Consent for publication

Written informed consent for publication was obtained from all participants.

Novelty statement

This work, to the best of author's knowledge, represents the first attempt in the discrimination of prostate cancer patients by using

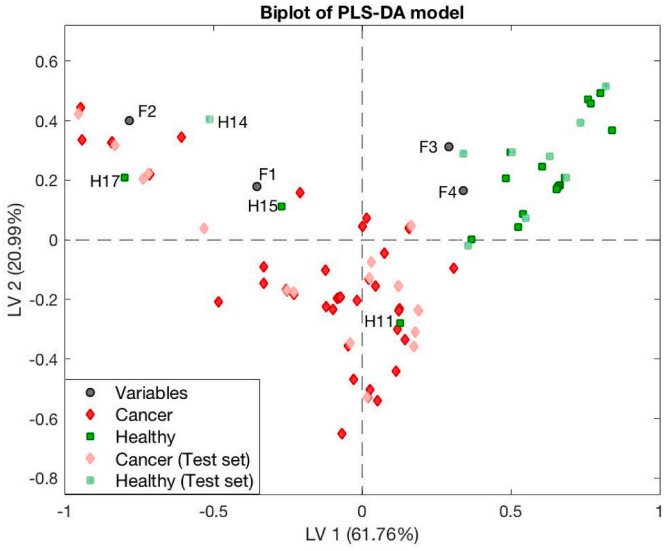


Fig. 6. PLS-DA biplot of the first vs. second PLS-DA components. Cancer patients (red diamonds); healthy persons (green circles). Variables, i.e. mode 1 PARAFAC loadings (gray circles), F1, F2, F3 and F4 stand for first, second, third and fourth components, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 2
LDA classification results.

	Calibration set	Validation set	Sensitivity training	Specificity training	Sensitivity test	Specificity test
Cancer class	47	19	100%	80%	94,7%	88,9%
Healthy class	20	9	80%	100%	88,9%	94,7%

Table 3
Results from PLS-DA model for classification of cancer and healthy classes based on PARAFAC scores.

	Calibration set	Validation set	LVs	AUC (CV)	Sensitivity CV	Specificity CV	Sensitivity predict	Specificity predict
Cancer class	47	19	2	0,92	91,5%	85%	94,7%	88,9%
Healthy class	20	9			85%	91,5%	88,9%	94,7%

excitation-emission fluorescence spectroscopy of urine coupled with chemometrics.

Four potential prostate cancer markers were putatively identified by means chemometrics analysis.

CRediT authorship contribution statement

Eleonora Mustorgi: Conceptualization, Investigation, Methodology. **Caterina Durante:** Conceptualization, Software, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Cristina Malegori:** Investigation. **Piergiorgio Greco:** Resources. **Riccardo Bartoletti:** Resources. **Marina Cocchi:** Software, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Monica Casale:** Conceptualization, Methodology, Project administration, Formal analysis, Software, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgement

The authors thank Dr Francesca Molinari for her support in analytical acquisitions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemolab.2023.104752.

References

[1] P. Rawla, Epidemiology of prostate cancer, World J. Oncol. 10 (2019) 63, https://doi.org/10.14740/wjon1191.
[2] European Association of Urology (EAU), Guidelines on Prostate Cancer, 2021, pp. 21–34. Edition Arnhem (NL). Pagg.
[3] S. Gamagedara, S. Gibbons, Y. Ma, Investigation of urinary pteridine levels as potential biomarkers for noninvasive diagnosis of cancer, Clin. Chim. Acta 412 (2011) 120, https://doi.org/10.1016/j.cca.2010.09.015.
[4] S. Gamagedara, A.T. Kaczmarek, Y. Jiang, X. Cheng, M. Rupasinghe, Y. Ma, Validation study of urinary metabolites as potential biomarkers for prostate cancer detection, Bioanalysis 4 (2012) 1175, https://doi.org/10.4155/bio.12.92.
[5] Tyler Cook, Yinfa Ma, Sanjeewa Gamagedara, Evaluation of statistical techniques to normalize mass spectrometry-based urinary metabolomics data, J. Pharm. Biomed. Anal. 177 (2020), 112858, https://doi.org/10.1016/j.jpba.2019.112854.
[6] S. Madhuri, N. Vengadesan, P. Aruna, D. Koteeswaran, P. Venkatesan, S. Ganesan, Native fluorescence spectroscopy of blood plasma in the characterization of oral malignancy, Photochem. Photobiol. 78 (2003) 197–204, https://doi.org/10.1562/0031-8655(2003)0780197NFSOBP2.0.CO2.
[7] Guanghua Zhu, Huangxian Ju, Hong Zheng, Fluorescence spectroscopic determination of dipyridamole binding on pancreas-1 tumor cell membrane, Clin. Chim. Acta 348 (2004) 101, https://doi.org/10.1016/j.cccn.2004.05.001.
[8] V. Masilamani, T. Vijmasi, M. Al Salhi, K. Govindaraj, A.P. Vijaya-Raghavan, B. Antonisamy, Cancer detection by native fluorescence of urine, J. Biomed. Opt. 15 (2010) 57003-1–57003-9, https://doi.org/10.1117/1.3486553.
[9] M. Zvarik, D. Martinicky, L. Hunakova, I. Lajdova, L. Sikurova, Fluorescence characteristics of human urine from normal individuals and ovarian cancer patients, Neoplasma 60 (2013) 533–537, https://doi.org/10.4149/neo_2013_069.
[10] R. Rajasekaran, P.R. Aruna, D. Koteeswaran, L. Padmanabhan, K. Muthuvelu, R. R. Rai, P. Thamilkumar, M.K. Chilakapati, S. Ganesan, Characterization and diagnosis of cancer by native fluorescence spectroscopy of human urine, Photochem. Photobiol. 89 (2013) 483–491, https://doi.org/10.1111/j.1751-1097.2012.01239.x.
[11] A. Gardlo, A.K. Smilde, K. Hron, M. Hrda, R. Kalikova, D. Friedeckỳ, T. Adam, Normalization techniques for PARAFAC modeling of urine metabolomic data, Metabolomics 12 (2016) 117, https://doi.org/10.1007/s11306-016-1059-9.
[12] L.C. Brown, H.U. Ahmed, R. Faria, A. El-Shater Bosaily, R. Gabe, R.S. Kaplan, et al., Multiparametric MRI to improve detection of prostate cancer compared with transrectal ultrasound-guided prostate biopsy alone: the PROMIS study, Health Technol. Assess. 22 (2018) 1, https://doi.org/10.3310/hta22390.
[13] J. I Epstein, M. J Zelefsky, D. D Sjoberg, J.B. Nelson, L. Egevad, C. Magi-Galluzzi, et al., A **contemporary prostate cancer grading system: a validated alternative to the Gleason score**, Eur. Urol. 69 (2016) 428, https://doi.org/10.1016/j.eururo.2015.06.046.
[14] A.V. D'Amico, Risk-based management of prostate cancer, N. Engl. J. Med. 365 (2011) 169, https://doi.org/10.1056/NEJMe1103829.
[15] R.D. Snee, Validation of regression models: methods and examples, Technometrics 19 (1977) 415–428, https://doi.org/10.1080/00401706.1977.10489581.
[16] R. Bro, H.A.L. Kiers, A new efficient method for determining the number of components in PARAFAC models, J. Chemom. 17 (2003) 274–286, https://doi.org/10.1002/cem.801.
[17] C.M. Andersen, G. Mortensen, Fluorescence spectroscopy: a rapid tool for analyzing dairy products, J. Agric. Food Chem. 56 (2008) 720–729, https://doi.org/10.1021/jf072025o.
[18] K.R. Murphy, C.A. Stedmon, D. Graeber, R. Bro, Fluorescence spectroscopy and multi-way techniques, **PARAFAC**, Anal. Methods 5 (2013) 6557–6566, https://doi.org/10.1039/C3AY41160E.
[19] J.R. Lakowicz, Principles of Fluorescence Spectroscopy, Springer, New York, 2006, ISBN 978-0-387-46312-4.
[20] R. Bro, M. Vidal, EEMizer: automated modeling of fluorescence EEM data, Chemometr. Intell. Lab. Syst. 106 (2011) 86–92, https://doi.org/10.1016/j.chemolab.2010.06.005.
[21] R.A. Fisher, The use of multiple measurements in taxonomic problems, Ann. Eugen 7 (1936) 179–188, https://doi.org/10.1111/j.1469-1809.1936.tb02137.x.
[22] M. Barker, W. Rayens, Partial least squares for discrimination, J. Chemom. 17 (2003) 166–173, https://doi.org/10.1002/cem.785.
[23] R. Bro, PARAFAC. Tutorial and applications, Chemometr. Intell. Lab. Syst. 38 (1997) 149–171, https://doi.org/10.1016/S0169-7439(97)00032-4.
[24] M.J.P. Leiner, M.R. Hubmannn, O.S. Wolfbeis, The total fluorescence of human urine, Anal. Chim. Acta 198 (1987) 13–23, https://doi.org/10.1016/S0003-2670(00)85002-3.
[25] O.O. Abugo, R. Nair, J.R. Lakowicz, Fluorescence properties of rhodamine 800 in whole blood and plasma, Anal. Biochem. 279 (2000) 142–150, https://doi.org/10.1006/abio.2000.4486.
[26] A.J. Lawaetz, R. Bro, M. Kamstrup-Nielsen, I.J. Christensen, L.N. Jørgensen, H. J. Nielsen, Fluorescence spectroscopy as a potential metabonomic tool for early detection of colorectal cancer, Metabolomics 8 (2012) S111–S121, https://doi.org/10.1007/s11306-011-0310-7.
[27] L. Lenhardt, R. Bro, I. Zeković, T. Dramicanin, M.D. Dramicanin, Fluorescence spectroscopy coupled with PARAFAC and PLS DA for characterization and classification of honey, Food Chem. 175 (2015) 284–291, https://doi.org/10.1016/j.foodchem.2014.11.162.
[28] S. Gamagedara, S. Gibbons, Y. Ma, Investigation of urinary pteridine levels as potential biomarkers for noninvasive diagnosis of cancer, Clin. Chim. Acta 412 (2011) 120–128, https://doi.org/10.1016/j.cca.2010.09.015.
[29] S.C. Larsson, N. Orsini, A. Wolk, Vitamin B6 and risk of colorectal cancer. A meta-analysis of prospective studies, JAMA 303 (2010) 1077–1083, https://doi.org/10.1001/jama.2010.263.