Noninvasive detection of bladder cancer using mid-infrared spectra classification

Siouar Bensaid, Amar Kachenoura, Nathalie Costet, Karim Bensalah, Hugues Tariel, Lotfi Senhadji

Noninvasive detection of bladder cancer using mid-infrared spectra classification[☆]

Siouar Bensaid[a,b], Amar Kachenoura[a,b], Nathalie Costet[a,b], Karim Bensalah[a,b,c], Hugues Tariel[d], Lotfi Senhadji[a,b,*]

[a] *Université de Rennes 1, LTSI, F-35000 Rennes, France*
[b] *Inserm, UMR 1099, F-35000 Rennes, France*
[c] *CHU de Rennes, Service d'Urologie, F-35033 Rennes, France*
[d] *DIAFIR, Avenue Chardonnet, Parc Lorans 26J, F-35000 Rennes, France*

Abstract

In this paper, we focus on the detection of Bladder Cancer (BC) via mid infrared spectroscopy. Two main contributions, material and methods, are presented. In terms of material, a new minimally invasive technology, combining fiber evanescent wave spectroscopy and newly patented biosensors, is used for the first time to acquire mid-infrared spectra from voided urine/bladder wash. This new machine promises practicality, cheapness and high-quality of spectrum acquisition. As for classical systems, the data acquired using the new system was highly correlated, resulting in a poor classification performance using classical methods. Therefore, the second contribution consists in developing statistical methods that alleviate the problem. Three new statistical methods based on Partial Least Square Discriminant Analysis algorithm (PLSDA) are proposed. PLSDA is a supervised classifier well-known for its ability to process correlated data. The key point is the choice of the most discriminant latent variables in the training step. In this work, we propose three new decision rules in order to select the most relevant latent variables. These decision rules give rise to three algo-

[☆]This work was supported by the AMNIFIR project through a grant from the French National Research Agency (ANR), TecSan Programme, grant number ANR-12-TECS-0012.

[*]Corresponding author
Email addresses: siouar.bensaid@univ-rennes1.fr (Siouar Bensaid), amar.kachenoura@univ-rennes1.fr (Amar Kachenoura), nathalie.costet@univ-rennes1.fr (Nathalie Costet), karim.bensalah@chu-rennes.fr (Karim Bensalah), hugues.tariel@diafir.com (Hugues Tariel), lotfi.senhadji@univ-rennes1.fr (Lotfi Senhadji)

rithms, namely bayesian, joint and best model PLSDA. A comparative study between the proposed methods and standard ones, namely SVM, K-MEANS and classical PLSDA, confirms clearly the efficiency of the former. The best performance in terms of accuracy is achieved by joint and best model PLSDA (82.35%). Besides, by embedding the proposed statistical methods in the new machine, we are able to provide a new medical device that is very promising in terms of automatic bladder cancer detection.
Keywords: bladder cancer, variable selection, infrared spectroscopy, automatic detection, chalcogenide glass fibers, PLSDA, SVM

1. Introduction

In 2012, the International Agency for Research on Cancer (IARC) has estimated 429.800 new cases and 165.100 deaths from Bladder Cancer (BC) worldwide (Torre et al., 2015). In the United States, BC ranks 3^{rd} in prevalence and 7^{th} in mortality among men (American Cancer Society, 2015). Although BC is sometimes discovered incidentally on imaging studies, the most common symptom is hematuria, that occurs in 80-90% of the cases (Goodison et al., 2013). BC has a high risk of recurrence that requires a lifelong follow-up, which makes BC the most costly of all cancers from diagnosis to death (Smith & Guzzo, 2013). For that reason, there is a great need of accurate minimally invasive and cost-effective screening methods.

The gold standard to detect BC is cystoscopy, which is an examination of the bladder with a rigid or flexible endoscope inserted through the urethra (Badalament et al., 1987). Magnified images of the bladder and/or the tumour allow the doctor to make an immediate visual diagnosis that will require further pathologic confirmation. This test succeeds in detecting most BCs. However, it can be less conclusive in patients with bladder inflammation in case of an indwelling catheter or an infection. It can also fail to detect, flat lesions (carcinoma *in situ*) and small papillary tumors (Issaq et al., 2008; Smith & Guzzo, 2013). Furthermore, fibroscopy can be cumbersome and painful which can make

patient follow-up hazardous. Generally, cystoscopy is combined with voided urinary cytology since the latter is efficient in detecting carcinoma in situ. Though simple and noninvasive, urinary cytology is not used as a screening test because of low sensitivity and intraobserver and interobserver variability (Badalament et al., 1987; Moonen et al., 2006). The direct contact with bladder mucosa and easy collection of urine have raised interest in the evaluation of urine-based biomarkers. A number of molecular tests have been developed and approved by the US Food and Drug Administration (FDA) (Goodison et al., 2013; Smith & Guzzo, 2013; Schmitz-Dräger et al., 2015; Kamat et al., 2013; Moreira et al., 2010; Shariat et al., 2008). These tests are more sensitive for lower-grade tumors than cytology and detect carcinoma *in situ* and low grade BC. However, they still lack diagnosis precision and are approved to be used only in conjunction with cytology and cystoscopy.

In an effort to explore different pathways, optical-based methods have been described: *In vivo* optical imaging technologies as Photodynamic Diagnosis (PDD), Narrow Band Imaging (NBI), Confocal Laser (CLE) and Optical Coherence Tomography (OCT) (Liu et al., 2012). PDD and NBI have been shown to increase the ability to detect carcinoma in situ (Palmer et al., 2013). CLE and OCT improve the ability to diagnose muscle-invasive BC. These techniques can only be performed through a cystoscope and have not been conceived to replace it. Additional flaws have been also reported such as high false-positive rates for PDD and NBI, patients layoff time caused by the administration of contrast agents for PDD and CLE, and unaddressed cost-effectiveness issue (Palmer et al., 2013). An optical method that has also received much interest is Raman Spectroscopy (RS) which is a vibrational spectroscopy technique based on the inelastic scattering of light after its interaction with tissue (Bensalah et al., 2010; Fleureau et al., 2011; Couapel et al., 2013). RS has been proven more specific than PDD in discriminating cancerous lesions from benign inflammations (Draga et al., 2010). The specificity can be enhanced by replacing conventional high-volume probe with confocal probe, allowing deep imaging of tissues with less background noise (Barman et al., 2012). A specificity of

100% has been achieved for confocal RS compared to 79% for high-volume RS. Despite these interesting results, several reasons have limited RS incorporation into clinical practice: length of acquisition, high cost and strong fluorescence background that obscures weak Raman signals. Recent works has proposed to circumvent these setbacks (De Luca et al., 2015; Li et al., 2015). RS has been tested on tissue (De Jong et al., 2006), urine (De Luca et al., 2015) and blood specimens (Li et al., 2015).

Another optical method that has been a major focus of many researches is Fourier-Transform Infrared (FTIR) spectroscopy which is a vibrational spectroscopy based on infrared light absorption. High sensitivity, high speed, and simple and self-calibrated instrumentation are major advantages of FTIR over the dispersive techniques. FTIR is mostly used in the Mid-InfraRed (MIR) region where fundamental vibrational modes of most of the interesting biomolecules are located (Hocdé et al., 2004). Promising results showed the ability of FTIR spectroscopy, to characterize BC using blood, tissue and urine specimens (Ollesch et al., 2014; Hughes et al., 2013; Pezzei et al., 2013; Bird et al., 2008). The use of optical fibers, transparent in the MIR range, allowed to make measurements in the Attenuated Total Reflectance (ATR)-mode at a remote location. This is called Fiberoptic Evanescent Wave Spectroscopy (FEWS). The recent development of chalcogenide-glass optical fiber helped efficient implementation of FEWS and provided more informative recorded spectra (Bureau et al., 2005; Keirsse et al., 2004, 2003). This new technology was proved promising in detecting metabolic alterations in patients (Anne et al., 2009; Le Corvec et al., 2016; Le Corvec et al., 2016; Albert et al., 2016), but to our knowledge, it has not been tested on BC yet.

The first contribution of this study consisted in testing a new device that implements FEWS and chalcogenide biosensors technologies in a practical and inexpensive way, in order to explore the automatic diagnosis of BC. The new device was used to acquire a high-quality MIR spectra from voided urine (non-invasive) and (minimally-invasive) bladder wash samples. Statistical methods should then be developed and embedded in the machine in order to classify the

acquired spectra. Nevertheless, one flaw was the high-correlation measured between control and BC groups. Therefore, the challenge was to propose new classifiers able to "optimally" process highly-correlated data. In statistical analysis, the Principal Component Analysis (PCA) has been prevalently used to analyze bladder samples as a main unsupervised classifier (Barman et al., 2012; De Luca et al., 2015; Bird et al., 2008; Pezzei et al., 2013; Hughes et al., 2013) or as a preprocessing step for another classifier such as Linear Discriminant Analysis (LDA) (Draga et al., 2010; De Jong et al., 2006). Nevertheless, it was shown in (Issaq et al., 2008) that Partial Least Square Discriminant Analysis (PLSDA) outperforms PCA in the classification of urinary bladder samples using high performance liquid chromatography coupled online with a mass spectrometer metabolomic approach. Though Both PCA and PLSDA are efficient in classifying highly-correlated data, the latter was proved more advantageous as it incorporates the labeling of observations in the computation of the uncorrelated components. Moreover, a key point in PLSDA is the decision rule used to select the most discriminant components. As a second contribution in this paper, we proposed to use three new decision rules: the "Bayesian", the "joint" and the "best model", giving rise to three new PLSDA methods, denoted PLSDA$_B$, PLSDA$_J$ and PLSDA$_{BM}$, respectively. We conjecture that they are more efficient than the classical decision rules as they rank the components according to their classification performance in the training step before selecting them in the prediction step. A comparative study between the proposed methods and the classical PLSDA as well as standard methods, namely support vector machine (SVM) and K-MEANS, were conducted. The proposed methods were shown to perform better in terms of classification scores. The best accuracy (82.35%) was achieved by joint PLSDA and best model PLSDA. Hence, the obtained results suggest that the new acquisition system combined with the new proposed methods presents a potential noninvasive and cost-effective method for BC screening and follow-up. It is noteworthy that we are the first to experiment the new patented MIR bio-sensors developed by the company DIAFIR (Tariel & Charpentier, 2015).

2. Data acquisition

2.1. Patients and sample preparation

A prospective pilot study was carried out in the urology department of CHU of Rennes, following a given protocol designed and validated therein. Volunteer patients were fully informed about the study and gave their written consent. They were aged over 18 (men or women). The tumoural group included BC-diagnosed patients that were scheduled for endoscopic tumour resection or total cystectomy. The control group included patients hospitalized for urolithiasis surgery, without indwelling double-J stent. Patients who did not fully satisfy these criteria or were suffering from other types of cancer were excluded from this study, as well as pregnant and breastfeeding women.

Urine specimens were collected twice from each patient. In a first phase, voided urine samples were collected in the morning fasting, before getting in the operation room of urology department in CHU of Rennes (preoperative time). This noninvasive procedure takes only about 15 minutes. In a second phase, a bladder washing was achieved in the operation room before the surgical operation (intraoperative time). Collected samples were frozen at -80°C in the Biological Resource Center of Rennes before the analysis.

The final cohort included 40 equally distributed patients (20 diagnosed with BC and 20 diagnosed with urolithiasis). We denote by PREOP and INTRAOP the voided urine and the bladder washing samples, respectively. Both were examined by MIR spectroscopy.

2.2. FEWS using mid-infrared biosensors

FEWS is an infrared remote spectroscopy that combines the principle of ATR and the technology of fiber sensors to allow rapid and *in-situ* analysis of samples. In FEWS, an infrared signal is transmitted in an optical fiber by total internal reflection. ATR occurs when a chemical sample is brought into contact with the fiber, thereby generating a partial absorption of the infrared signal in the fiber interface. The spectrum of the attenuated infrared wave picked up at the fiber

Figure 1: SPID™ spectrometer and LS23™ sensor patented by DIAFIR

output provides the metabolic characteristics of the sample. Chalcogenide glass fibers enjoy rheological properties that made them an optimal choice for FEWS implementation. They are flexible and exhibit a large optical transparency in the MIR spectral range from 2 to 12 μm with optical losses below 1 $dB.m^{-1}$ in the 6-9 μm region (Keirsse et al., 2004). Their hydrophobic behavior makes them especially suitable for application in biology and medicine where water is a nuisance to detect relevant information. It is shown that the sensitivity of sensors is highly improved when the diameter of the fiber is locally reduced, creating thus a tapered sensing zone that is brought into contact with the sample to be analyzed (Lucas et al., 2004). DIAFIR managed to develop optical fibers embedded into a disposable cover. Their fabrication procedure is detailed in (Houizot et al., 2014). These sensors were used in our experiments, they are single use and require roughly 7 μl volume of urine per analysis.

2.3. Instrumentation

The experimental setup consisted of an SPID™ non cooled spectrometer, coupled with a chalcogenide glass fiber enclosed in LS23™ sensor developed by DIAFIR (see figure 1) (Tariel & Charpentier, 2015). SPID™ is a FTIR spectrometer that was especially conceived to accommodate bio-sensors developed by DIAFIR. It is made compact compared to classical spectrometers so that it can be used in point of care application. For spectrum acquisition, a clean dry propette was used to place a drop from one thawed aliquot of a given sample

on the LS23™ sensor.

The FTIR spectra were acquired in the absorption mode in the [4000,400] cm^{-1} frequency range. The nominal spectral resolution was set to 4 cm^{-1} and a zero-filling factor of 2 was employed, yielding a discrete spectral point spacing of 2 cm^{-1}. A Blackman Harris three-term apodisation function was used for Fourier transformation. Sixty-four scans were recorded and averaged to yield the sample absorption intensity. The optical signal was recorded at the output extremity of the fiber, providing the infrared single beam spectra. A single beam reference spectrum was obtained for the background (without sample) before each sample analysis, yielding for every wavenumber a reference intensity. This step is necessary to take into consideration possible side effects in the experiment such as the entrance/exit conditions of the infrared beam, the interaction and attenuation along the optical signal transportation section, the transition of modes during the taper to the sensing zone, and effects related to fiber bending or surface roughness (Bureau et al., 2005). The sample and reference spectra of each patient were processed to derive only one "intensity ratio" spectrum that was used as input data in the statistical analysis presented in section 3.

2.4. Datasets

As described in table 1, there were two datasets (PREOP and INTRAOP) of 40 signals each. In each dataset, BC and control groups were equally distributed (20 spectra per group). The underlying spectra were displayed, outliers were visually spotted and ruled out. Three and 6 outliers were omitted from PREOP and INTRAOP datasets, respectively. We ended up with 37 spectra in PREOP (18 BC and 19 control) and 34 spectra in INTRAOP (18 BC and 16 control).

The resulting INTRAOP and PREOP raw spectra are presented in figure 2. Initial portion before 900 cm^{-1} was removed due to the limited visibility of the optical fiber (>800 cm^{-1}) and the high effect of water. The portion between 2800 $cm^{-1}$and 1800 cm^{-1} was also omitted to eliminate the contribution of the environmental CO_2. Visual inspection of the two datasets did not reveal

any remarkable difference between INTRAOP and PREOP spectra. Moreover,
195 we noticed that, for both, spectra of BC and control groups were similar in

Table 1: A two explored MIR signals datasets

40 Patients			
40 PREOP		40 INTRAOP	
20 BC	20 control	20 BC	20 control
Outliers omission			
18 BC	19 control	18 BC	16 control
37 PREOP		34 INTRAOP	

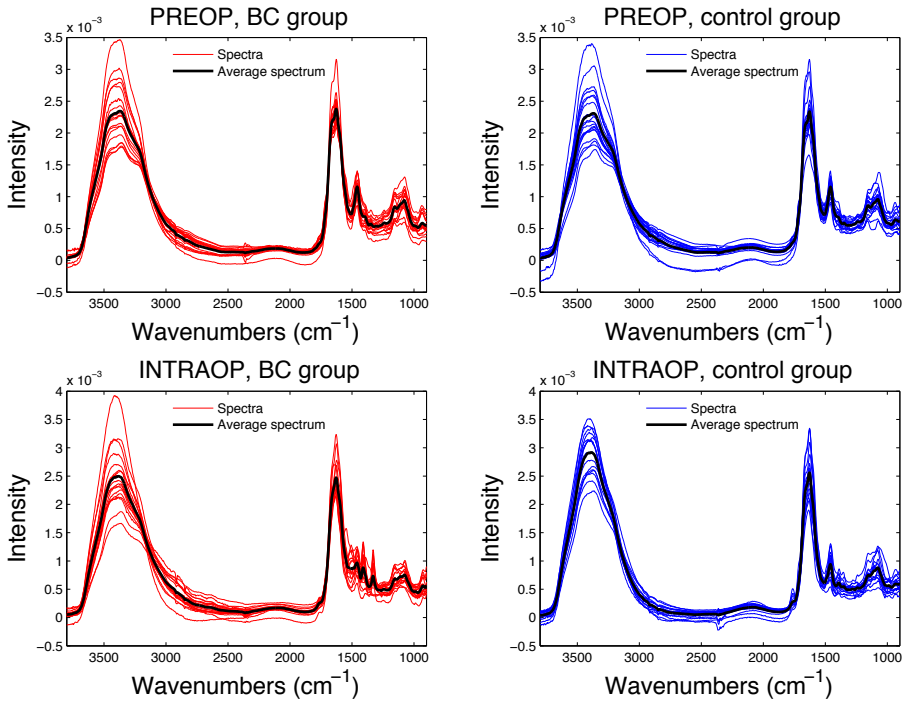

Figure 2: PREOP (top) and INTRAOP (bottom) MIR spectra of BC (red) and control (blue) groups, and their corresponding averages (black).

average. Therefore intergroup correlations were expected to be high. The hydrophobic property of chalcogenide glass fibers helps to minimize the effect of water, but it does not remove it completely, especially when the analyzed samples are biofluids as in our context. The influence of water was noticed in the presence of a baseline (slow wave) and a significant lobe after 3000 cm^{-1} in all spectra. This internal artifact was very disturbing since it might screen small details that would be important for discrimination. Moreover, it increased the intergroup correlations and brought a spurious similarity into the intragroup samples. In figure 3, we clearly notice the homogeneity between the diagonal and the off diagonal blocks (intragroup and intergroup correlations, respectively) in the PREOP correlation matrix. This homogeneity was less present in the INTRAOP correlation matrix where the blocks were distinguishable. However, the correlation was still considered significant which made the discrimination task very complicated.

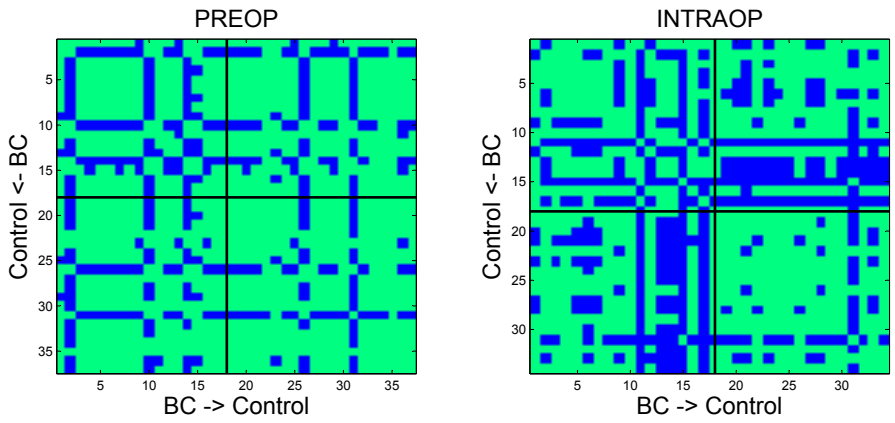

Figure 3: Correlation matrix of PREOP (left) and INTRAOP (right) spectra: correlation coefficients > 0.95 are set to 1 (green), 0 otherwise (blue).

3. Classification methods

We built a classification scheme based on three main stages involving a tunable preprocessing step, an appropriate classifier and a strategy to evaluate the classification performance. Let's denote by $\boldsymbol{X}$ the $M \times N$ matrix of dataset, where M and N are the number of patients and the length of one spectrum, respectively. $\boldsymbol{y}$ denotes the M-length discrete vector of labels assigned to patients.

3.1. Preprocessing

First, we normalized the signals by dividing each spectral point by the area of the total intensity of the spectrum to ensure their comparability. We then considered baseline correction using a classical method. The latter consisted in computing the first derivative of signals as the difference between two consecutive samples, followed by a 5 point Salvitzky-Golay smoothing. The classification procedure was separately applied on raw data and its first derivatives.

3.2. Partial Least Square Discriminant Analysis

3.2.1. Overview of the classical approach

PLSDA is a linear supervised classifier that was proposed by Sjötröm et al. as an extension of the PLS Regression algorithm (PLSR) to the discriminant analysis paradigm (Sjöström et al., 1986). PLSR aims to predict a set of continuous dependent variables (the responses) from a set of independent variables (the predictors) by extracting from the latter a reduced-dimension set of orthogonal components called Latent Variables (LVs) that have the best predictive power (Abdi, 2010). This algorithm is advantageous when predictors are numerous and redundant, which generally happens when the number of predictors is larger than the number of observations. In this case, classical regression algorithms such as multiple linear regression are no longer feasible. Moreover, the main asset of PLSR is that LVs are constructed using the covariance between both responses and predictors, whereas in principal component regression, defined as a truncated PCA (Jolliffe, 1982), they are determined using solely the

covariance of predictors. In PLSDA, responses are no longer continuous but discrete variables. In the binary case, the discrete response vector $\boldsymbol{y}$ is transformed into an $M \times 2$ dummy matrix $\boldsymbol{Y}$ where each binary entry $\boldsymbol{Y}(m, c)$ represents the membership of the m^{th} sample to the c^{th} class ($c = \{1, 2\}$); i.e. if the m^{th} sample belongs to class c then $\boldsymbol{Y}(m, c) = 1$, otherwise 0 (Ballabio & Consonni, 2013).

As a supervised method, the PLSDA requires a training step where a regression model $\mathcal{M}$ is formed between predictors $\boldsymbol{X}$ and responses $\boldsymbol{Y}$ as it follows

$$\boldsymbol{X} = \boldsymbol{T}\,\boldsymbol{P}^T + \boldsymbol{E} \tag{1}$$

$$\boldsymbol{Y} = \boldsymbol{T}\,\boldsymbol{Q}^T + \boldsymbol{F} \tag{2}$$

where $\boldsymbol{T}$ and $\boldsymbol{P}$ denote the $M \times L$ scores matrix and the $N \times L$ loadings matrix of predictors, respectively. $\boldsymbol{Q}$ is the $2 \times L$ loadings matrix of responses ($\boldsymbol{Y}$-weights). The scalar L refers to the number of LVs retained in the model. $\boldsymbol{E}$ and $\boldsymbol{F}$ are considered as residuals. The computation of this model is achieved using algorithm 1. We used the algorithm implemented in the PLSDA classification toolbox (Ballabio & Consonni, 2013). Notice the common use of the score matrix $\boldsymbol{T}$ in decomposing both $\boldsymbol{X}$ and $\boldsymbol{Y}$ (regression mode of PLSDA). It is also noteworthy that LVs in algorithm 1 are extracted in a descending explanation power, i.e. the elements $\mathcal{M}\{:, l\} = \{\boldsymbol{W}(:, l), \boldsymbol{Q}(:, l), \boldsymbol{T}(:, l), \boldsymbol{P}(:, l)\}$ explain the cross-covariance of $(\boldsymbol{X}, \boldsymbol{Y})$ more than $\mathcal{M}\{:, l + 1\}$. Here, $\boldsymbol{W}$ denotes the $N \times L$ loadings matrix of predictors ($\boldsymbol{X}$-weights). Once built, the model $\mathcal{M}$ is used to label new samples $\boldsymbol{X}_{new}$ in the prediction step as described in algorithm 2.

There are two focal "parallel" steps in building the PLSDA classifier: i) the estimation of model complexity aiming to find out the number of selected LVs that best describe the data without redundancy, and ii) the assessment of the overall quality of the model (Szymáska et al., 2012). In step i), the parameter L is estimated by corssvalidation where several models with different dimensions are generated and evaluated. The optimal LVs number corresponds to the one of the best performing model in terms of classification scores. The step ii) is motivated by the properties of regression models where predicted responses $\hat{\boldsymbol{Y}}$

are not perfectly binary. Therefore, a decision rule is established to deduce class assignments using properly estimated thresholds. Several decision rules were derived and it was shown that the choice of the decision rule may deeply affect the discrimination results (Szymáska et al., 2012).

A pivotal point in PLSDA is the choice of the most discriminant LVs. In the conventional strategy described before, LV selection is tackled in terms of optimal number (how many LVs should be kept?), which implies that LVs are selected in the same order of their extraction. In other words, if L is the optimal estimated number, then the L first extracted LVs are retained for the prediction step. This strategy is justified by the fact that the explanation power of a given LV is correlated to its discrimination power, which is not always true. Indeed, in some cases, discriminative features may lie in the small details captured by distant LVs (Fargeas et al., 2013). This observation motivated us to develop new strategies of LVs selection.

Algorithm 1 $\mathcal{M}$ = plsr $(\boldsymbol{X}, \boldsymbol{Y}, L)$

for $l = 1$ to L **do**
  Initialize $\boldsymbol{U}(:,l)$ and $\boldsymbol{T}(:,l)$ with random column vectors from $\boldsymbol{Y}$ and $\boldsymbol{X}$ respectively, initialize $\boldsymbol{T}_{old}$ with random vector $>$ than $\boldsymbol{T}(:,l)$ in norm, $\boldsymbol{E} = \boldsymbol{X}$ and $\boldsymbol{F} = \boldsymbol{Y}$
  while $\boldsymbol{T} \neq \boldsymbol{T}_{old}$ or number of iteration max is not reached yet **do**
    $\boldsymbol{T}_{old} \leftarrow \boldsymbol{T}(:,l)$
    $\boldsymbol{W}(:,l) \leftarrow \boldsymbol{E}^T\, \boldsymbol{U}(:,l)/(\boldsymbol{U}(:,l)^T\boldsymbol{U}(:,l))$ (compute $\boldsymbol{X}$-weights)
    $\boldsymbol{W}(:,l) \leftarrow \boldsymbol{W}/(\boldsymbol{W}(:,l)^T\boldsymbol{W}(:,l))$
    $\boldsymbol{T}(:,l) \leftarrow \boldsymbol{E}\, \boldsymbol{W}(:,l)/(\boldsymbol{W}(:,l)^T\boldsymbol{W}(:,l))$ (compute $\boldsymbol{X}$-scores)
    $\boldsymbol{Q}(:,l) \leftarrow \boldsymbol{F}^T\, \boldsymbol{T}(:,l)/(\boldsymbol{T}(:,l)^T\boldsymbol{T}(:,l))$ (compute $\boldsymbol{Y}$-weights)
    $\boldsymbol{U}(:,l) \leftarrow \boldsymbol{F}\, \boldsymbol{Q}(:,l)/(\boldsymbol{Q}(:,l)^T\boldsymbol{Q}(:,l))$ (compute $\boldsymbol{Y}$-scores)
  end while
  $\boldsymbol{P}(:,l) \leftarrow \boldsymbol{E}^T\, \boldsymbol{T}(:,l)/(\boldsymbol{T}(:,l)^T\boldsymbol{T}(:,l))$
  $\boldsymbol{E} \leftarrow \boldsymbol{E} - \boldsymbol{T}\, \boldsymbol{P}^T$ and $\boldsymbol{F} \leftarrow \boldsymbol{F} - \boldsymbol{T}\, \boldsymbol{Q}^T$
end for
$\mathcal{M} \leftarrow \{\boldsymbol{W}, \boldsymbol{Q}, \boldsymbol{P}\}$

Algorithm 2 $\hat{\boldsymbol{Y}}$ = plsr_pred $(\boldsymbol{X}_{new}, \mathcal{M})$

Initialize $\hat{\boldsymbol{Y}} \leftarrow 0$
for $l = 1$ to L **do**
  $\boldsymbol{T}_{new}(:,l) \leftarrow \boldsymbol{X}_{new}\, \boldsymbol{W}(:,l)/(\boldsymbol{W}(:,l)^T\boldsymbol{W}(:,l))$
  $\hat{\boldsymbol{Y}} \leftarrow \hat{\boldsymbol{Y}} + \boldsymbol{T}_{new}(:,l)\, \boldsymbol{Q}(:,l)^T$
  $\boldsymbol{X}_{new} \leftarrow \boldsymbol{X}_{new} - \boldsymbol{T}_{new}(:,l)\, \boldsymbol{P}(:,l)^T$
end for

3.2.2. New strategies of LV selection in PLSDA

We proposed a modified PLSDA algorithm with a new cross-validation step where LVs were ranked according to their discrimination (and not explanation) power before selecting the most relevant for an optimal model. As in the classical approach, the dataset was divided into three complementary subsets $\boldsymbol{X}_{tr}$, $\boldsymbol{X}_v$

and $\boldsymbol{X}_{ts}$ that were used for training, validation and testing, respectively.

In the training step, a regression model $\mathcal{M}$ of L_{max} LVs was built using algorithm 1, where L_{max} is the maximum number of extracted components. It is generally set to the smallest dimension of $\boldsymbol{X}_{tr}$. In a three-stage validation step, the extracted LVs were *i*) evaluated in terms of discrimination power, *ii*) ranked, and then *iii*) combined to find out an optimal subset for classification. In order to assess the discrimination power of each LV $l \in [1, \cdots, L_{max}]$, the elementary responses $\{\hat{\boldsymbol{Y}}_l\}_{l=1:L_{max}}$ were first estimated by algorithm 2 using the validation set $\boldsymbol{X}_v$ and the elementary model $\mathcal{M}\{:, l\}$. A decision rule was then applied to assign classes and to evaluate the underlying performances in terms of sensibility, specificity and accuracy. We proposed to use three decision rules: the "Bayesian", the "joint" and the "best model" decision rules.

In the "Bayesian" decision rule, two thresholds λ_1^B and λ_2^B are separately defined for classes 1 and 2, respectively (Ballabio & Consonni, 2013). They are calculated on the basis of Bayes theorem, assuming that the estimated values in each class follow a relatively normal distribution. Each threshold is selected at the point where the number of false positives and false negatives of the underlying class is minimized. When $\hat{\boldsymbol{Y}}(m, c)$ is greater than λ_c^B, the m^{th} sample is assigned to the c^{th} class, otherwise not. When a given sample m is assigned to both classes or conversely to none of them, it is defined as 'not_assigned'. In the proposed strategy, the "Bayesian" decision was applied to each response matrix $\hat{\boldsymbol{Y}}_l$, which resulted in estimating $2 \times L_{max}$ thresholds λ_{l1}^B and λ_{l2}^B and their corresponding performances.

The "best model" and the "joint" decision rules are based on ROC curves computation where each sample is always affected to one class at the end of the process. In both approaches that we denoted by the "best model" decision and the "joint" decision, two thresholds are defined for classes 1 and 2 and are calculated on the basis of the ROC curves plotted for the estimated responses $\hat{\boldsymbol{Y}}(:, 1)$ and $\hat{\boldsymbol{Y}}(:, 2)$, respectively. In the "best model" decision, thresholds λ_1^{BM} and λ_2^{BM} are defined as the optimal thresholds that independently maximize the performances of responses $\hat{\boldsymbol{Y}}(:, 1)$ and $\hat{\boldsymbol{Y}}(:, 2)$, respectively. Whereas in

the "joint" decision, the optimal thresholds λ_1^J and λ_2^J are jointly fixed such that they maximize the performance of both $\hat{\boldsymbol{Y}}(:,1)$ and $\hat{\boldsymbol{Y}}(:,2)$. Applying the "best model" decision (the "joint" decision) to each response matrix $\hat{\boldsymbol{Y}}_l$ resulted in estimating $2 \times L_{max}$ thresholds λ_{l1}^{BM} and λ_{l2}^{BM} (λ_{l1}^J and λ_{l2}^J), and their corresponding performances.

$$\begin{cases} \lambda_{l1}^{BM} = \arg\max \text{Perf}_{l1} \\ \lambda_{l2}^{BM} = \arg\max \text{Perf}_{l2} \end{cases} \qquad l = 1, \cdots, L_{max}$$

and

$$(\lambda_{l1}^J, \lambda_{l2}^J) = \arg\max(\text{Perf}_{l1} + \text{Perf}_{l2}), \quad l = 1, \cdots, L_{max}$$

where Perf_{l1} and Perf_{l2} denoted the performances deduced from $\hat{\boldsymbol{Y}}_l(:,1)$ and $\hat{\boldsymbol{Y}}_l(:,2)$, respectively.

In the ranking stage ii), the different components are sorted in a descending order of discrimination power, which amounts to building a mapping function that assigns a new index l' to component l. Unlike the "Bayesian" and "joint" decisions where only one rank is assigned to a given component l, the "best model" decision assigns two different ranks l'_1 and l'_2 as the two responses $\hat{\boldsymbol{Y}}_l(:,1)$ and $\hat{\boldsymbol{Y}}_l(:,2)$ are independently used. Hence, the index mapping of the "Bayesian" and "joint" decisions denoted respectively by $\boldsymbol{i}^B$ and $\boldsymbol{i}^J$ are L_{max}-length vectors, whereas, the index mapping of the "best model" decision is an $L_{max} \times 2$ matrix of two columns $\boldsymbol{i}_1^{BM}$ and $\boldsymbol{i}_2^{BM}$ (see figure 4).

The final stage of validation step consists in estimating the optimal number L of rearranged components in the final model. To this end, the ranked LVs are sequentially combined, evaluating each time the classification scores of the generated model on $\boldsymbol{X}_v$. The optimal model corresponds to the most performing in terms of classification scores. In "the best model" scenario, the use of two rankings, $\boldsymbol{i}_1^{BM}$ and $\boldsymbol{i}_2^{BM}$, yields two different independent performances; the optimal model is then given by the best performance.

We denote by PLSDA_B, PLSDA_J and PLSDA_{BM} the PLSDA-based algorithms when using the "Bayesian", the "joint" and the "best model" decision rules, respectively, and by $\mathcal{M}^B$, $\mathcal{M}^J$ and $\mathcal{M}^{BM}$ the underlying generated opti-

mal models. The latter are evaluated in the testing step, using $\boldsymbol{X}_{ts}$ subset, in terms of prediction power. Obviously, the decision rule used in this step is the same rule used to construct the model in the validation step.

3.3. Performance evaluation

Due to the limited cohort size (less than 20 patients for BC and control groups), it seemed less relevant to divide the dataset into three complementary sets. Therefore, a training/testing scheme was rather used where the dataset was only divided into two complementary sets $\boldsymbol{X}_{tr}$ and $\boldsymbol{X}_{ts}$, and the training dataset $\boldsymbol{X}_{tr}$ was used for both training and validation. Moreover, a leave-one-out cross-validation (LOOCV) scheme was implemented to evaluate the classification scores. In LOOCV, one patient is removed from the given dataset, in turn, and used as a testing set.

3.4. Significant predictors selection

For a given spectrum and after omitting irrelevant intervals (see section 3.1), there were $N = 933$ wavenumbers (predictors) used in classification, which is relatively a high number compared to the available cohort size M (37 for PREOP, 34 for INTRAOP). There is also a probability that the effective information is confined in only few discriminative wavenumbers. In order to check this possibility, we proposed a second scenario (test II) where an intermediate step of wavenumbers selection was established inbetween preprocessing and classification. In this additional step, the One-Way ANOVA test was implemented to help identifying the most discriminant predictors. In this test, we compared the intergroup variation to the intragroup variation for each given wavenumber $n \in [1, \cdots, N]$. If the ratio of the intragroup variation to the intergroup variation is significantly high, then we conclude that the group means are significantly different from each other and the underlying wavenumber is expected to deeply influence the classification result. The significance of each test was measured with the P-value parameter where the latter was compared to a fixed

Table 2: The different achieved tests

	Raw spectra	First derivative
without One-Way ANOVA	Test I.A	Test I.B
with One-Way ANOVA	Test II.A	Test II.B

λ_p cutoff. A wavenumber n was selected when its P-value is lower than λ_p. The λ_p cutoff was fixed to one of the standard values, namely $\{0.05, 0.02, 0.01, \ldots\}$.

Using different preprocessing steps with tests I and II yielded several scenarios that are summarized in table 2: tests I.A and I.B refer to running classifiers on raw data and its first derivatives, respectively, without significant predictors selection; and tests II.A and II.B correspond to running classifiers on raw data and its first derivatives, respectively, with significant predictors selection. An overview of the proposed classifiers and the different scenarios is given in figure 4.

3.5. Comparative study

In order to assess the improvement introduced by the proposed algorithms, we compared the latter with the classical PLSDA (PLSDA_C) implemented in the the PLSDA classification toolbox where the LVs are arranged according to their extraction order and the classification scores are computed using the "Bayesian" decision rule (Ballabio & Consonni, 2013). To guarantee the comparability of scores between the different methods, samples labeled as "not_assigned" in PLSDA_C and PLSDA_B were switched to the misclassified categories so that not-assigned samples amongst BC samples and control samples were assigned to false negative and false positive categories, respectively.

The PLSDA-based algorithms were as well compared to two other standard classifiers. The first classifier, K-MEANS, is an unsupervised method where a dataset is partitioned into k clusters by minimizing a certain distance (Bishop, 2006). In our case, we had two clusters ($k = 2$) and the criterion was set to the Euclidean distance. The second classifier was a support vector machine (SVM).

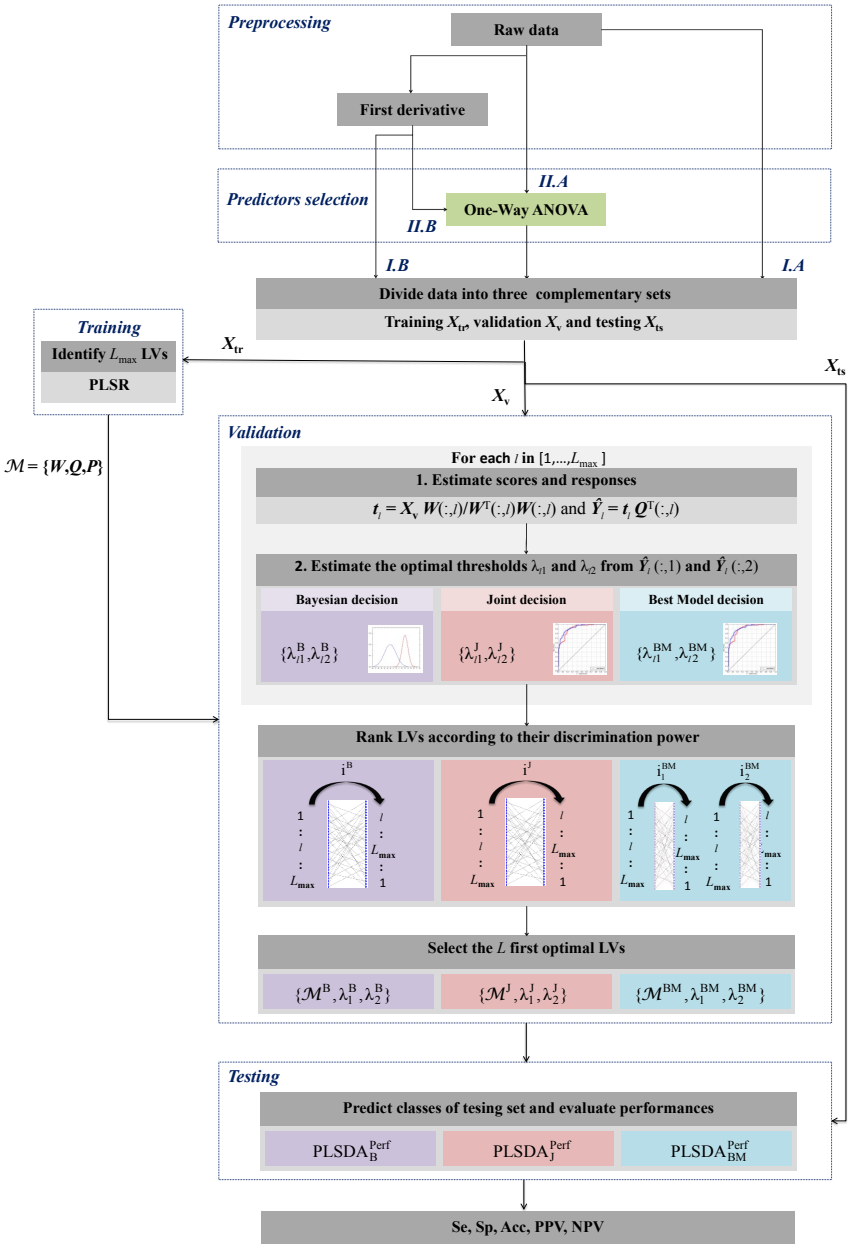

Figure 4: Overview diagram

The latter is based on learning models (kernel functions) in the training step to build a hyperplane separating data in two classes (Bishop, 2006). In our tests, we set a linear kernel and applied a LOOCV scheme to evaluate classification scores.

To sum up, six classifiers (K-MEANS, SVM, PLSDA$_C$, PLSDA$_B$, PLSDA$_J$ and PLSDA$_{BM}$) were involved in all the tests enumerated in table 2. Classification scores were compared and discussed in the next section.

4. Results and discussion

In this section, we applied the proposed procedure in section 3 to PREOP and INTRAOP databases. Performance evaluation was achieved via the computation of usual classification scores, namely, Specificity (Sp), Sensibility (Se), Accuracy (Acc), Positive Predictive Value (PPV) and Negative Predictive Value (NPV). The purpose of these experiments was, first, to confirm our conjecture that FEWS can be considered as a BC diagnostic tool. Secondly, we were also interested to find out which of the two samples, PREOP and INTRAOP, would be more informative. All the obtained results were summarized in figure 5.

4.1. Test I: without One-Way ANOVA

Looking to the bar plots of tests I.A and I.B (figure 5), we clearly see that all the classifiers failed to achieve good scores on both raw and first derivative PREOP spectra (top and second left plots). The best achieved accuracy was limited to 51% (test I.A, K-MEANS). In the meanwhile, good performances were globally observed for INTRAOP data (third ans bottom left plots), which suggests that the latter are more informative than the PREOP samples. This finding was expected since bladder wash samples are richer with exfoliated cells than voided urine. Therefore, more information can be collected about the uroepithelium state.

Focusing more on the results obtained for INTRAOP spectra, PLSDAs were obviously performing better than standard methods, K-MEANS and SVM, except for test I.B (bottom left) where SVM outperformed PLSDA$_C$ and PLSDA$_{BM}$.

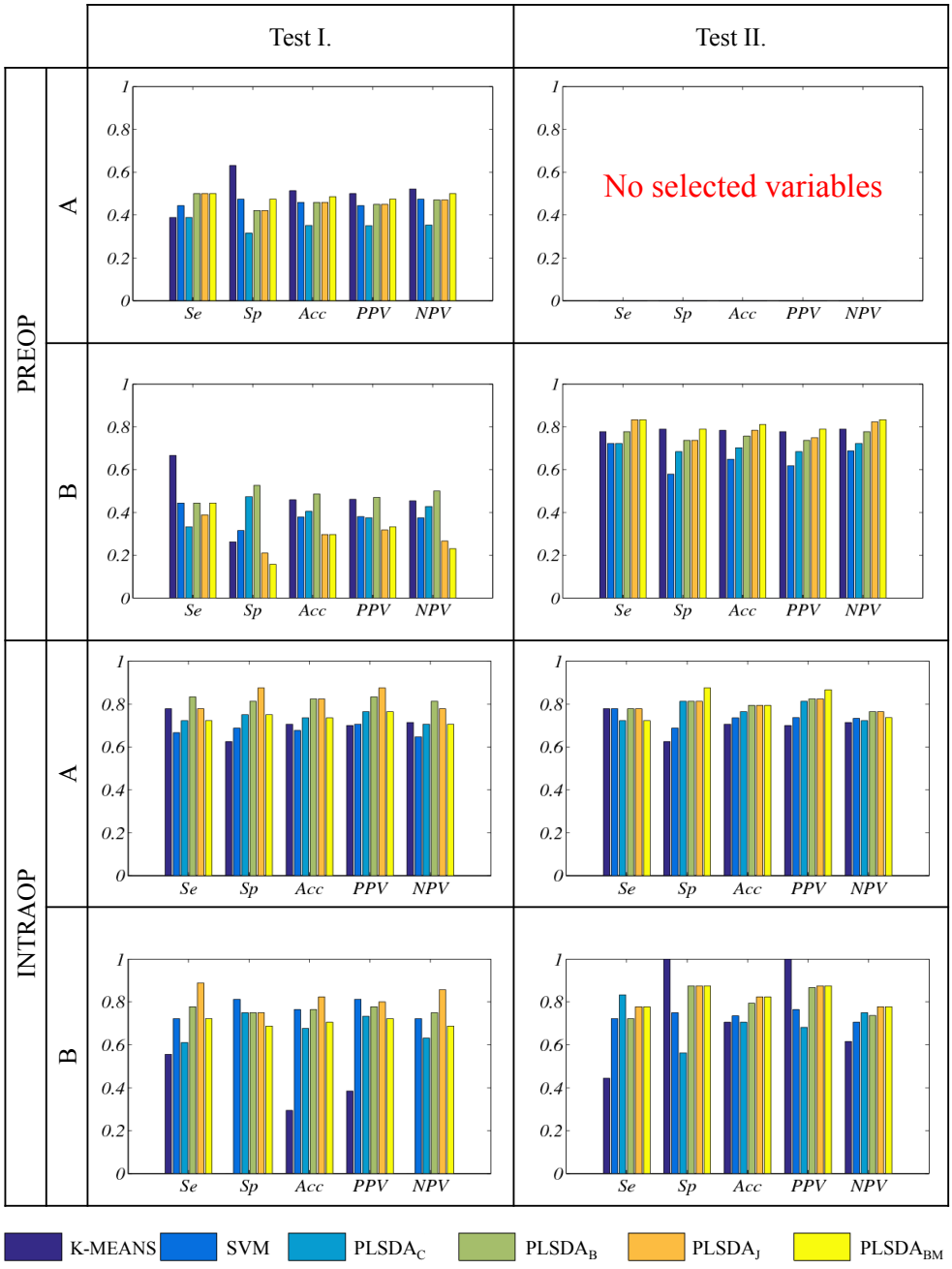

Figure 5: Classification scores of K-MEANS, SVM, PLSDA$_C$, PLSDA$_B$, PLSDA$_J$ and PLSDA$_{BM}$ obtained in the tests described in table 2 for PREOP and INTRAOP datasets

The new proposed methods achieved better classification scores than PLSDA$_C$ proving the efficiency of the new strategy for LVs selection. More precisely, for the three proposed methods, we noticed in one hand that the first extracted LV was always ruled out from the optimal model. On the other hand, the second extracted LV always ranked first in performance and therefore was usually selected in the model. The rejection of the first extracted LV from the optimal model was quietly expected. In fact, this component incorporates the high correlation existing between the spectra and brought by the presence of water. Contrariwise, in PLSDA$_C$, the first component was unavoidably selected in the model resulting in slight decrease of classification scores. Generally, the new proposed methods generated optimal models composed of utmost two LVs ($L = 2$) where the second extracted LV always ranked first (in performance) and the third extracted LV ranked second in 80% of cases. The optimal model size L in the case of PLSDA$_C$ was estimated to three components in 90% of cases. It is noteworthy that in the case of PREOP database, the selected LVs from one LOOCV to another was quasi-random whatever the used method was, which explains again the underlying low classification scores.

The best classification scores were obtained in test I.A using PLSDA$_J$ and PLSDA$_B$ where the former has achieved the highest specificity and accuracy (87.5% and 82.35% respectively), while the latter had a good sensitivity reaching 83.3%. Nevertheless, the best sensitivity was obtained by PLSDA$_J$ in test I.B (88.89%). Globally, the use of first derivative did not improve the performance of the classifiers except for SVM whose specificity was significantly improved with the first derivative (from 68.75% to 81.25%). The slight decrease in PLSDAs' performances can be justified by the fact that PLS framework is tailored mainly for highly-correlated signals. Whereas, by using the first derivative, a significant part of this correlation is omitted. This also explains why these classifiers tended to select larger models more often than with raw spectra.

4.2. Test II: with One-Way ANOVA

In this test, the λ_p cutoff was fixed to 0.05 for both II.A and II.B. The total set of variables (spectrum length) was of $N = 933$ wavenumbers. Unsurprisingly, no significant variable was selected when applying the One-Way ANOVA test to PREOP database (test II.A), whence the empty case in top left of figure 5. This result confirms one more time the poor performance found in test I.A with the same database. However, a set of 20 variables was selected among the $N = 933$ in test II.B. In terms of classification score, performance has considerably improved for all algorithms. The new proposed methods performed better than PLSDA$_C$ that was poorly specific (68.4%). The best performance was achieved by PLSDA$_{BM}$ with a specificity a sensitivity and an accuracy reaching 83.3%, 79% and 81% respectively. Unlike the previous tests, the behavior of PLSDAs was more stable in terms of LVs selection: i) the second extracted exponent was dominantly selected in the first rank by the proposed methods and ii) PLSDA$_C$ was constantly selecting the two first extracted components ($L = 2$). These results show that PREOP data can be potentially used in BC screening especially that the underlying samples were noninvasively collected.

In the case of INTRAOP database, 455 and 431 variables were selected in tests II.A and II.B respectively. No significant change was noticed in classification scores of test II and the different algorithms behaved globally the same as in test I. The best result was equally achieved by PLSDA$_J$ and PLSDA$_{BM}$ (Se = 77.7%, Sp = 87.5% and Acc = 82.35%).

Test II indicates that useful information can be confined in a limited set of wavenumbers. Significant variables spotted in INTRAOP database largely exceeded the ones in PREOP database. This observation is coherent with the expectations supposing that bladder wash is biochemically richer than voided urine. Preprocessing ANOVA test was not very relevant in the case of INTRAOP database since no significant improvement was noticed in classification scores in test II. The best result could be practically obtained with raw full spectra (test I.A with PLSDA$_B$ and PLSDA$_J$). Contrariwise, test II was very relevant in the case of PREOP database. The first derivative helped to rule

out variables detrimental to classification and resulted in a performance quasi-
equivalent to the one of INTRAOP database. Hence, according to this test, PREOP dataset is also eligible for BC diagnostic.

5. Conclusion and perspectives

In this paper, a preliminary study of BC detection via MIR spectra was carried out on two different categories of collected samples, bladder wash (INTRAOP) and voided urine (PREOP). A new system based on FEWS technology was used for spectrum acquisition. Three PLSDA classifiers, namely PLSDA$_B$, PLSDA$_J$ and PLSDA$_{BM}$, with new LV selection strategy and decision rules were developed and evaluated in a carefully designed batch of tests. Across the different tests, INTRAOP spectra constituted an informative dataset for the different classifiers, mainly for our proposed methods that outperformed the classical ones in all the proposed tests. Though the PREOP dataset was less informative and consequently more difficult to exploit, a promising result was obtained after two preprocessing steps: smoothed first derivative and One-Way ANOVA test. An asset of PREOP dataset is the noninvasive collection of samples, which makes it more attractive to use especially for the follow-up. Based on the formerly described experimental outcome, we can corroborate that, by embedding the proposed statistical methods in the new machine, we are able to provide a new minimally invasive medical device that is very promising in terms of fast and automatic BC detection. Moreover, the low price of the used biosensors (0.05 euros) helps reduce the screening and follow-up costs.

Although very promising, results reported in this study are still preliminary given the small used cohorts. All these tests should be reproduced with larger cohorts in order to ensure about the robustness of the derived conclusions. Typically, in the case of PREOP dataset, where the proposed methods were less efficient compared to the INTRAOP case, using larger cohort may improve their robustness. In this case, a completely noninvasive medical device may be provided by analyzing only the spectra recorded from voided urine (avoiding

the use of bladder wash). In addition, the selection of significant variables in test II will be achieved more robustly paving the ground to the next step that aims to biochemically interpret and identify the significant selected variables. The proposed methods may also be extended to the characterization of BC in order to check if they are able to discriminate the different stages of the tumor. Other tumors can be investigated such as Nonalcoholic Steatohepatitis (NASH) (Bensaid et al., 2016). Finally, it would be very interesting to broaden the comparison and include more sophisticated approaches dedicated to extract hidden relevant information by mining such as deep learning based on recent convolutional neural networks.

References

Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley Interdiscip. Rev. Comput. Stat.*, *2*, 97–106. doi:10.1002/wics.51.

Albert, J.-D., Monbet, V., Jolivet-Gougeon, A., Fatih, N., Le Corvec, M., Seck, M., Charpentier, F., Coiffier, G., Boussard-Pledel, C., Bureau, B., Guggenbuhl, P., & Loréal, O. (2016). A novel method for a fast diagnosis of septic arthritis using mid infrared and deported spectroscopy. *Jt. Bone Spine*, *83*, 318–323. doi:10.1016/j.jbspin.2015.05.009.

American Cancer Society (2015). Cancer Facts & Figures 2015. *Cancer Facts Fig. 2015*, (pp. 1–9). doi:10.1097/01.NNR.0000289503.22414.79. arXiv:NIHMS150003.

Anne, M.-L., Le Lan, C., Monbet, V., Boussard-Plédel, C., Ropert, M., Sire, O., Pouchard, M., Jard, C., Lucas, J., Adam, J. L., Brissot, P., Bureau, B., & Loréal, O. (2009). Fiber evanescent wave spectroscopy using the mid-infrared provides useful fingerprints for metabolic profiling in humans. *J Biomed Opt*, *14*. doi:10.1117/1.3253319.

Badalament, R. A., Hermansen, D. K., Kimmel, M., Gay, H., Herr, H. W., Fair, W. R., Whitmore, W. F., & Melamed, M. R. (1987). The sensitivity of bladder wash flow cytometry, bladder wash cytology, and voided cytology in the detection of bladder carcinoma. *Cancer*, *60*, 1423–1427.

Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: linear models. PLS-DA. *Anal. Methods*, *5*, 3790–3798. doi:10.1039/C3AY40582F.

Barman, I., Dingari, N. C., Singh, G. P., Kumar, R., Lang, S., & Nabi, G. (2012). Selective sampling using confocal Raman spectroscopy provides enhanced specificity for urinary bladder cancer diagnosis. *Anal. Bioanal. Chem.*, *404*, 3091–3099. doi:10.1007/s00216-012-6424-6.

Bensaid, S., Kachenoura, A., Costet, N., Ledinghen, V. D., Vergniol, J., Laine, F., Turlin, B., Tariel, H., & Senhadji, L. (2016). Early diagnosis of NAFLD-NASH transition using mid infrared spectroscopy. In *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2016, Orlando, FL, USA, August 16-20, 2016* (pp. 3602–3605). doi:10.1109/EMBC.2016.7591507.

Bensalah, K., Fleureau, J., Rolland, D., Lavastre, O., Rioux-Leclercq, N., Guillé, F., Patard, J.-J., Senhadji, L., & de Crevoisier, R. (2010). Raman Spectroscopy: A Novel Experimental Approach to Evaluating Renal Tumours. *Eur. Urol.*, *58*, 602–608. doi:10.1016/j.eururo.2010.06.002.

Bird, B., Romeo, M. J., Diem, M., Bedrossian, K., Laver, N., & Naber, S. (2008). Cytology by Infrared Micro-Spectroscopy: Automatic Distinction of Cell Types in Urinary Cytology. *Vib. Spectrosc.*, *48*, 101–106. doi:10.1016/j.vibspec.2008.03.006.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. (1st ed.). Springer-Verlag New York, Inc.

Bureau, B., Boussard-Plédel, C., Adam, J. L., & Lucas, J. (2005). Infrared optical fiber as evanescent wave bio-sensors. In I. Gannot (Ed.), *Biomed. Opt. 2005* (pp. 1–8). International Society for Optics and Photonics. doi:10.1117/12.587620.

Couapel, J.-P., Senhadji, L., Rioux-Leclercq, N., Verhoest, G., Lavastre, O., de Crevoisier, R., & Bensalah, K. (2013). Optical spectroscopy techniques can accurately distinguish benign and malignant renal tumours. *BJU Int.*, *111*, 865–871. doi:10.1111/j.1464-410X.2012.11369.x.

De Jong, B. W. D., Schut, T. C. B., Maquelin, K., Van Der Kwast, T., Bangma, C. H., Kok, D. J., & Puppels, G. J. (2006). Discrimination between nontumor bladder tissue and tumor by Raman spectroscopy. *Anal. Chem.*, *78*, 7761–7769. doi:10.1021/ac061417b.

De Luca, A., Dholakia, K., & Mazilu, M. (2015). Modulated Raman Spectroscopy for Enhanced Cancer Diagnosis at the Cellular Level. *Sensors*, *15*, 13680–13704. doi:10.3390/s150613680.

Draga, R. O. P., Grimbergen, M. C. M., Vijverberg, P. L. M., van Swol, C. F. P., Jonges, T. G. N., Kummer, J. A., & Ruud Bosch, J. L. H. (2010). In vivo bladder cancer diagnosis by high-volume Raman spectroscopy. *Anal. Chem.*, *82*, 5993–5999. doi:10.1021/ac100448p.

Fargeas, A., Kachenoura, A., Acosta, O., Albera, L., Drean, G., & De Crevoisier, R. (2013). Feature extraction and classification for rectal bleeding in prostate cancer radiotherapy: A PCA based method. *IRBM*, *34*, 296–299. doi:10.1016/j.irbm.2013.07.009.

Fleureau, J., Bensalah, K., Rolland, D., Lavastre, O., Rioux-Leclercq, N., Guillé, F., Patard, J.-J., de Crevoisier, R., & Senhadji, L. (2011). Characterization of renal tumours based on Raman spectra classification. *Expert Syst. Appl.*, *38*, 14301–14306. doi:10.1016/j.eswa.2011.05.092.

Goodison, S., Rosser, C. J., & Urquidi, V. (2013). Bladder cancer detection and monitoring: Assessment of urine- and blood-based marker tests. *Mol. Diagnosis Ther.*, *17*, 71–84. doi:10.1007/s40291-013-0023-x.

Hocdé, S., Loréal, O., Sire, O., Boussard-Plédel, C., Bureau, B., Turlin, B., Keirsse, J., Leroyer, P., & Lucas, J. (2004). Metabolic imaging of tissues by infrared fiber-optic spectroscopy: An efficient tool for medical diagnosis. *J. Biomed. Opt*, *9*, 404–407. doi:10.1117/1.1646415.

Houizot, P., Anne, M.-L., Boussard-Plédel, C., Loréal, O., Tariel, H., Lucas, J., & Bureau, B. (2014). Shaping of Looped Miniaturized Chalcogenide Fiber Sensing Heads for Mid-Infrared Sensing. *Sensors*, *14*, 17905–17914. doi:10.3390/s141017905.

Hughes, C., Iqbal-Wahid, J., Brown, M., Shanks, J. H., Eustace, A., Denley, H., Hoskin, P. J., West, C., Clarke, N. W., & Gardner, P. (2013). FTIR microspectroscopy of selected rare diverse sub-variants of carcinoma of the urinary bladder. *J. Biophotonics*, *6*, 73–87. doi:10.1002/jbio.201200126.

Issaq, H. J., Nativ, O., Waybright, T., Luke, B., Veenstra, T. D., Issaq, E. J., Kravstov, A., & Mullerad, M. (2008). Detection of Bladder Cancer in Human Urine by Metabolomic Profiling Using High Performance Liquid Chromatography/Mass Spectrometry. *J. Urol.*, *179*, 2422–2426. doi:10.1016/j.juro.2008.01.084.

Jolliffe, I. T. (1982). A Note on the Use of Principal Components in Regression. *J. R. Stat. Soc. Ser. C (Applied Stat.)*, *31*, 300–303. doi:10.2307/2348005.

Kamat, A. M., Hegarty, P. K., Gee, J. R., Clark, P. E., Svatek, R. S., Hegarty, N., Shariat, S. F., Xylinas, E., Schmitz-Dräger, B. J., Lotan, Y., Jenkins, L. C., Droller, M., Van Rhijn, B. W., & Karakiewicz, P. I. (2013). ICUD-EAU international consultation on bladder cancer 2012: Screening, diagnosis, and molecular markers. *Eur. Urol.*, *63*, 4–15. doi:10.1016/j.eururo.2012.09.057.

Keirsse, J., Boussard-Plédel, C., Loréal, O., Sire, O., Bureau, B., Leroyer, P., Turlin, B., & Lucas, J. (2003). IR optical fiber sensor for biomedical applications. *Vib. Spectrosc.*, *32*, 23–32. doi:10.1016/S0924-2031(03)00044-4.

Keirsse, J., Bureau, B., Boussard-Pledel, C., Leroyer, P., Ropert, M., Dupont, V., Anne, M. L., Ribault, C., Sire, O., Loreal, O., & Adam, J. L. (2004). Chalcogenide glass fibers used for in situ infrared spectroscopy in biology and medicine. In B. Culshaw, A. G. Mignani, & R. Riesenberg (Eds.), *Photonics Eur.* (pp. 61–68). International Society for Optics and Photonics. doi:10.1117/12.545430.

Le Corvec, M., Allain, C., Lardjane, S., Cavey, T., Turlin, B., Fautrel, A., Begriche, K., Monbet, V., Fromenty, B., Leroyer, P., Guggenbuhl, P., Ropert, M., Sire, O., & Loréal, O. (2016). Mid-infrared fibre evanescent wave spectroscopy of serum allows fingerprinting of the hepatic metabolic status in mice. *Analyst*, *141*, 6259–6269. doi:10.1039/C6AN00136J.

Le Corvec, M., Charpentier, F., Kachenoura, A., Bensaid, S., Henno, S., Bardou-Jacquet, E., Turlin, B., Monbet, V., Senhadji, L., Loréal, O., Sire, O., Betagne, J. F., Tariel, H., & Lainé, F. (2016). Fast and Non-Invasive Medical Diagnostic Using Mid Infrared Sensor: The AMNIFIR Project. *IRBM*, *37*, 116–123. doi:10.1016/j.irbm.2016.03.003.

Li, S., Li, L., Zeng, Q., Zhang, Y., Guo, Z., Liu, Z., Jin, M., Su, C., Lin, L., Xu, J., & Liu, S. (2015). Characterization and noninvasive diagnosis of bladder cancer with serum surface enhanced Raman spectroscopy and genetic algorithms. *Sci. Rep.*, *5*, 9582. doi:10.1038/srep09582.

Liu, J.-J., Droller, M. J., & Liao, J. C. (2012). New Optical Imaging Technologies for Bladder Cancer: Considerations and Perspectives. *J. Urol.*, *188*, 361–368. doi:10.1016/j.juro.2012.03.127.

Lucas, J., Bureau, B., Boussard-Plédel, C., Kierse, J., Anne, M.-L., Lucas, P., & Riley, M. (2004). Infrared evanescent wave bio-sensors. In *17th Annu. Meet.*

IEEE Lasers Electro-Optics Soc. LEOS (pp. 823–824). volume 2. doi:10.1109/LEOS.2004.1363494.

Moonen, P. M. J., Peelen, P., Kiemeney, L. A. L. M., Boon, M. E., Schalken, J. A., & Witjes, J. A. (2006). Quantitative Cytology on Bladder Wash versus Voided Urine: A Comparison of Results. *Eur. Urol.*, *49*, 1044–1050. doi:10.1016/j.eururo.2006.01.029.

Moreira, J. M., Ohlsson, G., Gromov, P., Simon, R., Sauter, G., Celis, J. E., & Gromova, I. (2010). Bladder cancer-associated protein, a potential prognostic biomarker in human bladder cancer. *Mol Cell Proteomics*, *9*, 161–177. doi:M900294-MCP200[pii]\r10.1074/mcp.M900294-MCP200.

Ollesch, J., Heinze, M., Heise, H. M., Behrens, T., Brüning, T., & Gerwert, K. (2014). It's in your blood: Spectral biomarker candidates for urinary bladder cancer from automated FTIR spectroscopy. *J. Biophotonics*, *7*, 210–221. doi:10.1002/jbio.201300163.

Palmer, S., Sokolovski, S. G., Rafailov, E., & Nabi, G. (2013). Technologic developments in the field of photonics for the detection of urinary bladder cancer. *Clin. Genitourin. Cancer*, *11*, 390–396. doi:10.1016/j.clgc.2013.04.016.

Pezzei, C., Brunner, A., Bonn, G. K., & Huck, C. W. (2013). Fourier transform infrared imaging analysis in discrimination studies of bladder cancer. *Analyst*, *138*, 5719–25. doi:10.1039/c3an01101a.

Schmitz-Dräger, B. J., Droller, M., Lokeshwar, V. B., Lotan, Y., Hudson, M. A., van Rhijn, B. W., Marberger, M. J., Fradet, Y., Hemstreet, G. P., Malmstrom, P.-U., Ogawa, O., Karakiewicz, P. I., & Shariat, S. F. (2015). Molecular Markers for Bladder Cancer Screening, Early Diagnosis, and Surveillance: The WHO/ICUD Consensus. *Urol. Int.*, *94*, 1–24. doi:10.1159/000369357.

Shariat, S. F., Karam, J. a., Lotan, Y., & Karakiewizc, P. I. (2008). Critical

evaluation of urinary markers for bladder cancer detection and monitoring. *Rev. Urol.*, *10*, 120–135.

Sjöström, M., Wold, S., & Söderström, B. (1986). PLS discriminant plots. In E. S. GELSEMA, & L. N. KANALS (Eds.), *Pattern Recognit. Pract.* (pp. 461–470). Amsterdam: Elsevier. doi:10.1016/B978-0-444-87877-9.50042-X.

Smith, Z. L., & Guzzo, T. J. (2013). Urinary markers for bladder cancer. *F1000Prime Rep.*, *5*, 21. doi:10.12703/P5-21.

Szymáska, E., Saccenti, E., Smilde, A. K., & Westerhuis, J. A. (2012). Double-check: Validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics*, *8*, 3–16. doi:10.1007/s11306-011-0330-3.

Tariel, H., & Charpentier, F. (2015). *WO Patent No. 2015/110767*. WIPO/PCT.

Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-tieulent, J., & Jemal, A. (2015). Global Cancer Statistics, 2012. *CA a cancer J. Clin.*, *65*, 87–108. doi:10.3322/caac.21262.