Analyst

PAPER

View Article Online
View Journal | View Issue

Cite this: *Analyst*, 2023, **148**, 4099

Classification of formalin-fixed bladder cancer cells with laser tweezer Raman spectroscopy†

Nga Tsing Tang, ‡[a,b] Richard Robinson,[c,d] Richard D. Snook,[a,b] Mick Brown,[c] Noel Clarke[c,d,e] and Peter Gardner *[a,b]

Received 20th January 2023,
Accepted 19th June 2023

DOI: 10.1039/d3an00119a

rsc.li/analyst

Bladder cancer is a common cancer that is relatively hard to detect at an early stage because of its non-obvious symptoms. It is known that bladder cells can be found in urine samples which potentially could be used for early detection of bladder cancer. Raman spectroscopy is a powerful non-invasive tool for accessing biochemical information of cells. Combined with laser tweezers, to allow isolation of single cells, Raman spectroscopy has been used to characterise a number of bladder cells that might be found in a urine sample. Using principal component-canonical variates analysis (PC-CVA) and *k*-fold validation, the results shows that the invasive bladder cancer cells can be identified with accuracy greater than 87%. This demonstrates the potential of developing an early detection method that identifies the invasive bladder cancer cells in urine samples.

1. Introduction

Bladder cancer accounts for 3% of all cancer incidence and 3% mortality in the UK, with a higher incidence in men,[1] similar to the global figures of 3% and 2.1% respectively.[2]

Bladder cancer is usually hard to detect at early stage as symptoms are not usually obvious or specific. However, 80% of bladder cancer patients will present with Haematuria, blood in the urine,[3] which should be followed up by renal ultrasound, CT Urogram and/or cystoscopy. Whilst imaging has good sensitivity and specificity for diagnosis (CT Urography sensitivity of 79–93% and specificity of 83–99%), confirmation of bladder cancer requires invasive cystoscopy visualisation of the bladder wall and pathological validation *via* a transurethral resection of bladder tumour (TURBT).[4,5] Currently there are no non-invasive urinary bladder cancer diagnostic biomarkers with high and stable sensitivity available in clinical practice.[6] The gold standard for diagnosis of bladder cancer is reported to be cystoscopy, which is an invasive method that often needs to take the risks of infection, pain, and haematuria into account. Although cytology provides a non-invasive alternative diagnostic method, it also has limitations such as low sensitivity, and has high reliance on the analysis by pathologists.[7] It has also been reported that immunocytological and/or cytological approaches were not be able to be used for the analysis due to insufficient number of cells.[8] Fluorescence-Activated Cell Sorting (FACS) is another alternative. However, previous knowledge is required for targeting specific biomarkers for different types of cells and there are only four types of cytokeratins tested as potential urinary biomarkers.[9] Therefore, it would be very helpful if a non-invasive diagnostic method without the need for previous knowledge could be developed with high sensitivity and specificity for bladder cancer.

Urine is one of the most non-invasive samples taken for diagnostic reasons and it has been shown that it contains mostly urethral cells, transitional epithelial cells and occasionally, in men, prostate cells. By isolating these cells of interest, measurement is amenable to spectroscopic investigation. Indeed, Raman spectroscopy is a powerful analysis tool for biological materials which can be used for accessing the biochemistry of biological samples in a label-free manner. With the ability to work with aqueous samples, it can be coupled with other techniques to perform a wider range of biochemical analysis and in the work presented here Raman spectroscopy is coupled with laser tweezers, to form laser tweezers Raman spectroscopy (LTRS).[10]

There have been a few preliminary studies using Raman spectroscopy to classify or identify bladder cancer cells in mixed-cell population. Canetta *et al.* used modulated Raman

[a]*Department of Chemical Engineering and Analytical Science, School of Engineering, University of Manchester, Manchester, M13 9PL, UK.*
E-mail: peter.gardner@manchester.ac.uk
[b]*Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK*
[c]*Division of Cancer Sciences, University of Manchester, Manchester, M20 4GJ, UK*
[d]*Department of Urology, Salford Royal NHS Foundation Trust, Salford, M6 8HD, UK*
[e]*Department of Surgery, The Christie NHS Foundation Trust, Manchester, M20 4BX, UK*

†Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3an00119a

‡Current affiliation: The Open Innovation Hub for Antimicrobial Surfaces, Surface Science Research Centre, Department of Chemistry, University of Liverpool, Liverpool, L69 3BX, UK.

spectroscopy (MRS) and atomic force microscopy (AFM) to classify the human urothelial cells (SV-HUC-1) and the bladder cancer cell (MGH-U1) in urine sample and achieved high sensitivity and specificity at 83% and above.[11,12] Kerr *et al.* used Raman microscopy to classify two bladder cancer cell lines, which are defined as low grade (RT-112) and high grade (T24) cell lines. The classification results show high sensitivity and specificity at 90% and above for these two cell lines.[13]

Apart from standard Raman spectroscopy and microscopy, studies on the classification of urological cells had also been done with using LTRS. Harvey *et al.* on a mixed population of prostate and bladder cells in either water or an artificial urine environment[14] achieved sensitivity and specificity of 75% or above for the classifications of urological cells with different lengths of time of urine exposure up to 12 hours. Although there are fluctuations of these values with different lengths of urine exposure time, it still shows possibility of utilising the LTRS method for urological cancer cells detection from urine samples. Other studies by Casabella *et al.* demonstrated the use of a LTRS system with an automatic microfluidic device for single cell analysis in urological cell samples.[15] Similar techniques were also reported by Dochow *et al.* where LTRS was used to analyse erythrocytes, leukocytes, acute myeloid leukaemia cells (OCI-AML3), and breast tumour cells BT-20 and MCF-7 in microfluidic glass channels.[16] Two types of optical traps were used in the study: capillary based optical trap and microchip based optical trap, and *k*-fold cross validation of linear discriminant analysis was used for analysing the data. The results show that the accuracy on classifying these cells is improved in the microchip-based experiment when comparing to the capillary based experiment due to the choice of materials of the devices. The results of the microchip based optical trap method shows that the classification can achieve an accuracy of 86% or above. That study demonstrated the feasibility of Raman-activated cell sorting for classification of different types of single cells.[16] Also, a study by Schie *et al.* shows that rapid acquisition of mean spectra of eukaryotic cells is one of the possible solutions to achieve high throughput by significantly reducing the acquisition time down to a few seconds,[17] which can be combined with the Raman-activated cell sorting system.

In this work, the LTRS system was used for the classification of seven bladder cancer cell lines, including the two used by Kerr *et al.*,[13] in which single cell isolation and spectrum acquisition can be done at the same time. The eventual aim of this study is to determine the feasibility of developing a urological cancer diagnosis method for bladder cancer that can separate the invasive feature of the cell lines.

2. Materials and methods

2.1 Purpose built laser tweezers Raman spectroscopy system

The newly configured purpose built LTRS system which was used for all the measurements presented in this work is shown in Fig. 1.

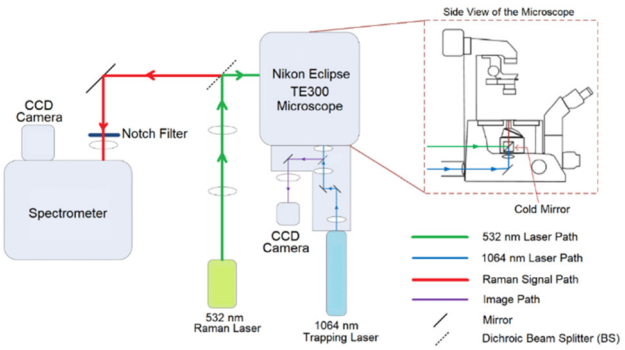

Fig. 1 Layout of the LTRS system used in the work presented.

The Raman system consists of a Horiba Scientific iHR-320 Imaging Spectrometer with focal length of 320 mm and *f*/4.1 aperture and a diffraction grating of 1200 lines per mm was coupled to a thermoelectrically cooled Horiba Syncerity charged coupled device (CCD) detector. The trapping part of the system consists of a Nikon Eclipse TE300 microscope equipped with a Plan Fluor 100× oil immersion objective which provides a high numerical aperture (NA) of 1.3 necessary for cell trapping.

A Laser Quantum's DPSS Ventus laser operating at 532 nm and capable of providing a maximum power of 110 mW at source was used for Raman excitation. A Laser Quantum's Diode Pumped Solid State (DPSS) Ventus laser operating at 1064 nm was used for cell trapping. The Raman laser power was set at 70 mW at source which was calculated to be reduced to around 41% at the sample by taking relative power measurements at source and objective. The reduction is caused by the overfilling of the objective and the complex optical structure of the system. The calculated approximate maximum Airy laser spot size for the Raman laser (532 nm) with the objective slightly overfilled is 0.50 μm which is significantly smaller than a cell. However, it had been proved that single point analyses in cells are representative for cell line classification with the fact that there is a certain degree of heterogeneity in a cell. Pavillon *et al.* suggested that a hybrid approach of rapid scanning across an area can provide a better picture of the overall information within a cell.[18] However, a study conducted by Harvey demonstrated that cell size, and therefore laser spot size coverage, has only little correlation to the classification outcome.[19] Also, similar study had also conducted by Kujdowicz *et al.* and Tang *et al.* demonstrated that despite variances across a single cell can be observed, these variances had insignificant effect on the overall outcome.[20,21] An integration time of 30 s was chosen so that high signal-to-noise ratio (SNR) can be obtained when acquiring Raman spectra. For the trapping laser, trapping power was set to 760 mW at source, which was then attenuated to ~127 mW at the sample.

2.2 Sample preparation

2.2.1 Cell line selection and sample preparation. Seven bladder cancer cell lines used in this work were cultured with

the corresponding standard culture media with a number of supplements. A basic supplement added in all media comprised 10 v/v % foetal bovine serum (FBS, F7524, Sigma-Aldrich, Poole, UK), 1 v/v % L-glutamine (G7513, Sigma-Aldrich, Poole, UK), and 1 v/v % penicillin/streptomycin (pen/strep, P4333, Sigma-Aldrich, Poole, UK).

Table 1 states the origins of the cell lines, the culture media used and additional supplements present for specific cell lines. These cell lines are all transitional cell carcinoma (TCC) as listed in the table.

2.2.2 Cell culture. Cells were cultured in standard T75 flasks and the incubator was maintained at an ambient condition of 37 °C and 5% CO_2. Standard culture media for each cell line were used as stated in Table 1.

2.2.3 Invasion assay. Bladder cell invasion through Matrigel (BD Matrigel™ Basement Membrane Matrix, BD Biosciences, NJ, USA) was assessed using the protocol of Hart *et al.*[23] 1×10^5 bladder cells were seeded into Matrigel coated polyethylene terephthalate (PET) cell culture inserts (8.0 μm pore size, growth area 0.3 cm^3, BD Falcon™, NJ, USA) above 5% FBS (PAA Laboratories, Pasching, Austria) in RPMI – 1640 (Sigma-Aldrich, Poole, UK), 10% fatty acid free bovine serum albumin (FAF BSA, Sigma-Aldrich, Poole, UK), 1% L-glutamine (Sigma-Aldrich, Poole, UK). After 18 hours incubation at 37 °C, 5% CO_2 inserts were washed in PBS and Matrigel removed prior to staining with crystal violet for 10 minutes. The number of cells stained with crystal violet present within a 10 mm × 10 mm graticule field were counted at 100× magnification.

2.2.4 Cytotoxicity assay. Bladder cells were seeded into 96 well plates at a concentration of 4×10^3 cells per well, followed by exposure to cisplatin in standard cell line growth media for 72 hours at 37 °C, 5% CO_2. Wells were fixed in 100 μl of trichloroacetic acid (TCA, 10% w/v dd H_2O) for 1 hour at 4 °C before drying at room temperature for 2 hours. Cell proliferation assessed by sulforhodamine B colorimetric (SRB) assay; 100 μl of SRB solution (0.4% w/v dd H_2O) was added to each well for 15 minutes at room temperature. Wells were then washed with acetic acid (1% w/v dd H_2O) before drying at room temperature for 2 hours. The SRB was then re-suspend in 100 μl Tris-hydrochloride (Tris-HCl, 1.5 M; pH 8.8) and immediately read on colorimetric plate reader at 490 nm absorbance (BioTeck, Winooski, WT, USA). Dose inhibitory response curves and IC_{50} values were determined using GraphPad Prism (GraphPad Software Inc., La Jolla, CA, USA).

Table 1 Summary of the origin of each bladder cell line, the culture media used for each cell line, and the supplements added

Name	Origin of cell line	Culture media
T24	ATCC collection – *via* the Translational Radiobiology Group, Paterson Institute for Cancer Research, University of Manchester Grade 3, primary tumour untreated TCC Female	RPMI-1640 (Sigma-Aldrich, Poole, UK)
J82	ATCC collection – *via* the Translational Radiobiology Group, Paterson Institute for Cancer Research, University of Manchester Grade 3, stage T3 Primary tumour treated TCC Male	EMEM (Sigma-Aldrich, Poole, UK)
5637	Obtained from Carcinogenesis group, Paterson Institute for Cancer Research, Cancer Research UK, Manchester Grade 2, primary tumour TCC Male	McCoy's 5A (Sigma-Aldrich, Poole, UK)
RT-112	ECACC collection Grade 2 papillary, stage T2 Primary tumour untreated TCC Female	EMEM (Sigma-Aldrich, Poole, UK) 1% Non-essential amino acid (NEAA, Sigma-Aldrich, Poole, UK)
UMUC-3	ATCC collection – *via* the Translational Radiobiology Group, Paterson Institute for Cancer Research, University of Manchester High grade cancer TCC Male	DMEM (Sigma-Aldrich, Poole, UK)
HT-1376	ATCC collection – *via* the Translational Radiobiology Group, Paterson Institute for Cancer Research, University of Manchester Grade 3 invasive, stage ≥ pT2 Primary tumour untreated TCC Female	DMEM (Sigma-Aldrich, Poole, UK)
T24-CDDPR	Developed by the Iwamura Group from the School of Medicine, Kitasato University[22] Cisplatin resistant cell line derived from T24 cell line Stepwise exposure of T24 cells to up to 40 μM of cisplatin	RPMI-1640 (Sigma-Aldrich, Poole, UK)

2.2.5 Raman experiments

2.2.5.1. Formalin fixation. When the cells reached 70–80% confluency, cells were trypsinised and washed with Dulbecco's phosphate-buffered saline (DPBS) twice before formalin fixation (10% formalin (neutral buffered, approximate 4% formaldehyde, Sigma-Aldrich)) and stored at 5 °C before Raman analysis.

2.2.5.2. Pre-measure preparation. Since cells measured in formalin will give a peak at ~1040 cm^{-1} in Raman spectra, which is a direct spectral contaminant from formalin,[24,25] formalin fixed samples have to be washed with DPBS before taking any Raman spectrum. On the day of Raman analysis, cells in formalin solution were washed with DPBS twice, resuspended in DPBS and measured within 24 hours. The reason for doing that is to get rid of the artefact that is brought by the formalin,[26] and also minimise the fluorescence background during the Raman spectra acquisition. Four biological replicates were measured for each cell line.

2.2.5.3. Biological replicates. Four biological replicates of each cell line were used, with each replicate derived from a different cell culture flask cultured under the same condition. Replicates (each contains 55 single-cell measurements) were measured on different days and times using the same instrument. During data analysis all the measured data were combined based on the type of cell line. Randomisation of the single cell spectra were also applied before any chemometric analysis. A schematic flow chart defining 'biological replicates' used in this work is shown in Fig. 2.

2.3 Raman cell spectra acquisition

During the Raman spectra acquisition, cells suspended in DPBS were placed into an open-top glass bottom dish. In order to trap a cell, agitation was applied until a cell falls into the optical trap. As mentioned in section 2.2.5 (iii), 55 cell spectra were measured from each set of replicates where the total measurement time is around 30 minutes per replicate. A new glass dish was used for each set of measurements to avoid any possible contamination. All replicates for each cell line were then combined into a bigger data set for further data processing and random sampling in later stages.

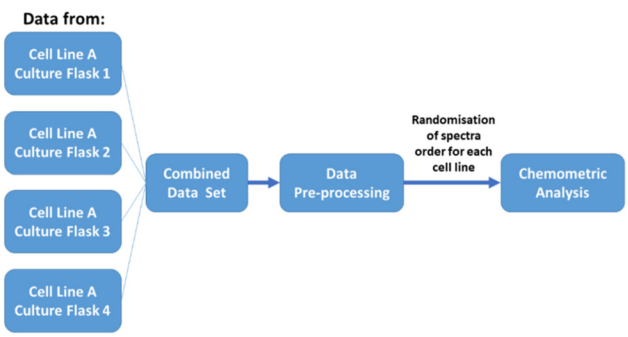

Fig. 2 Schematic flow chart explaining the source of a data set.

2.4 Data pre-processing for the Raman results

The acquired spectral data needs to be pre-processed before carrying out any chemometric analysis. Data pre-processing can be simply divided into five parts explained in detail in the following sections.

2.4.1 Background subtraction and select the suitable range for analysis. Spectra acquired were subjected to a background subtraction algorithm built-in the acquisition software, SynerJY (version 3.5.7.20). A background spectrum was first taken using an empty space within each individual sample, which contains the information of the glass bottom dish and DPBS. The algorithm will then perform a point-to-point direct subtraction between the cell spectrum and the background spectrum. This built-in function can compensate for wavelength dependent components of the system and dark charge build-up. Other preliminary data quality controls after this, such as range selection, cosmic ray removal and noise reduction, were done by using MATLAB 2017a. Spectra were then selected in a Raman shift range of 690–1750 cm^{-1} in MATLAB.

2.4.2 Removal of cosmic ray spikes. High energy radiation which produces sharp spikes on the Raman spectrum, known as cosmic rays, were removed by using an in-house written MATLAB algorithm. The algorithm works by calculating the first derivatives (gradients) of the data points, with any gradient greater than 2000 identified as cosmic ray spikes. Any spectrum with cosmic rays identified was removed from the data set and an average of 98% of spectra remaining after this step.

2.4.3 Noise reduction. A smoothing step was involved in the data processing by using a moving mean system. A data point x was replaced by a smoothed value x_s during the smoothing. This was done by taking the mean over the set of numbers of surrounding data points (window size). A window size of three was chosen to use in this work which is the smallest possible size available to prevent over smoothing.

2.4.4 Baseline correction. A broad fluorescent background is a commonly seen in a Raman spectrum as a broad wide baseline. Polynomial baseline subtraction and fitting was used to remove this baseline. For biological samples, a polynomial with order of 3–5 is typically used for the fitting and a 4^{th} order polynomial was used in this study.[27]

2.4.5 Vector normalisation. Intensities of spectra obtained are also dependent on the physical characteristics of the biological samples, such as cell sizes and thicknesses. Other influences, for instance, the position of the laser focus can also influence the results. Vector normalisation can be used to eliminate those effects by dividing the intensity by the spectral vector length at all wavenumbers. Note that a step of mean centring of the data is also involved and data at this point would be ready for the chemometric analysis.

2.5 Chemometrics analysis of the Raman results

Principal component-canonical variates analysis (PC-CVA) was used as the major chemometric analysis method as in the clirspec-summer-school: CLIRSPEC-Summer-School-2015 (v1.0)

package (https://doi.org/10.5281/zenodo.57398).[28] It is a quadratic discriminant analysis method and is suitable for multi-class problems.[29–31] Also, *k*-fold cross validation of PC-CVA was performed on the combined data set and results were presented as confusion matrices. As mentioned, 55 measurements were taken for each replicate per cell line, then all four replicates were combined to become a bigger data set (220 cells) for later data processing and analysis. After pre-processing the number of spectra per cell line was reduced. Considering the number of identities in the smallest data group (cell line), 70% of the spectra (120 spectra) were therefore taken from each cell line at each time to build 5 independent classifiers. This was done to avoid any bias brought by data size of each group.

Random Forest was also used as a complementary classification method in which 80% of the spectra from the full data set was used as the training set *i.e.* 20% of the spectra from the full data set was used as the test set. Random Forest was picked as the supervised machine learning method in this work, as a complementary analysis with the PC-CVA method, which is treated as a quick exploratory analysis on new studies. Although PC-CVA is a good first approach as an extension of PCA, machine learning algorithm will be required when translating the work to solve real clinical problem. Random Forest will allow this to be done. Random Forest is a robust technique that does not involve any transformation of the data into other forms of presentation such as scores plots, while PCA and CVA do some form of eigenvalue eigenvector decomposition. This may lead to extensive change in the loadings and hence the output if extra data is added into the classification. Another major consideration of using Random Forest is its ability to do both regression and classification tasks on large data sets with high accuracy in predicting outcome.

Independent tests were also performed where using data from one of the replicates as the test set and using the remaining data as the training test. 500 trees (the number determined by the out of bag error rate curves) were used for building the classifier in all cases described in this work.

Candidates used for building these classifiers were selected by undersampling method which will retain all features and reduce the chance of overfitting of some classes.[32,33]

3. Results

3.1 Characterisation of the bladder cell lines

Characterisation of the cell lines involved in this work was performed, which is used for the investigation of the separation pattern obtained from the Raman results.

3.1.1 Cell lines invasive potential. The bladder cell lines displayed significantly different invasive responses to 5% FBS across a Matrigel extracellular matrix (ECM) barrier. The T24 cell line was the most invasive cell line and significantly more invasive than J82, where the numbers of cells invaded were 1245.7 ± 78.9 and 873.3 ± 41.2 respectively (p = 0.0027). The cell lines 5637, RT-112, UMUC-3 and HT-1376 showed statistically similar but low levels of invasion of 53.3 ± 21.5, 150.4 ± 25, 190.4 ± 93.1, 134.6 ± 72.8 ($p > 0.05$) respectively. The corresponding results are plotted in Fig. 3.

3.1.2 Cell lines response to cisplatin. Cell line sensitivity to cisplatin (dose ranges 0.001–50 μg ml^{-1}) was assessed using a 72-hour SRB assay. The mean fold changes in cell proliferation for each drug concentration were normalised against the mean values for each cell line cultured in the presence of the drug vehicle alone. The calculated IC_{50} values for cisplatin were 0.046, 0.453, 0.729, 0.846, 0.917 and 0.974 μg ml^{-1} for the T24, J82, 5637, HT-1376, UMUC-3 and RT-112 cell lines respectively. The cell survival rates are potted against the cisplatin concentration in a log scale, which is shown in Fig. 4. Lines of best fit are also plotted for each case with fitting the data with sigmoid curves.

3.2 Results of the classification of the six bladder cancer cell lines

3.2.1 Classified with PC-CVA. The spectra of the six bladder cancer cell lines were first analysed with PC-CVA and the resultant scores plot and its corresponding loadings plot for the main separation are shown in Fig. 5(a) and (b) respect-

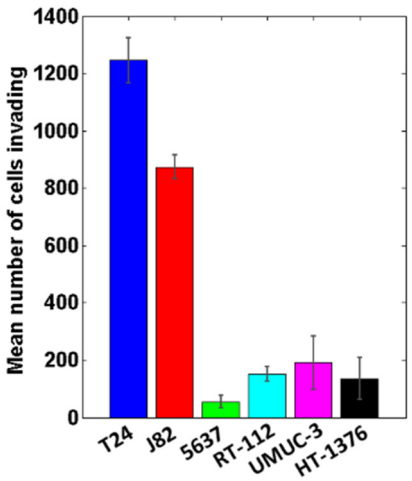

Fig. 3 The invasive ability of six bladder cancer cell lines towards 5% FBS.

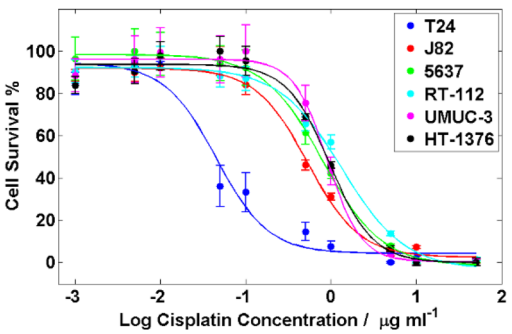

Fig. 4 Cisplatin dose inhibitory response curve (fitted with sigmoid curves) for cell proliferation of the six cell lines.

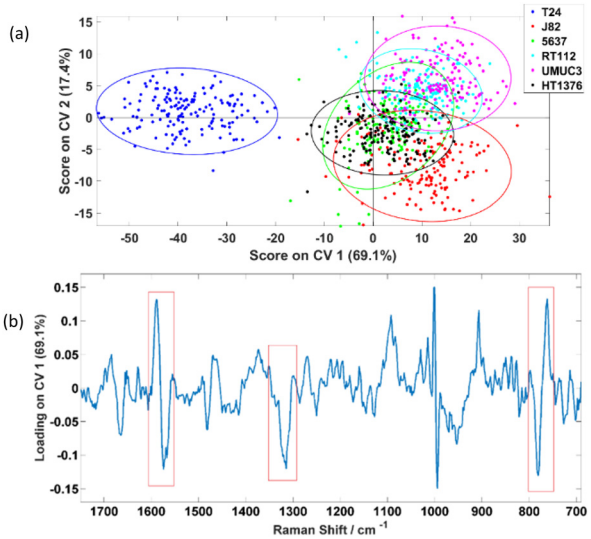

Fig. 5 PC-CVA (a) scores plot of CV 1 and CV 2 for classifying the six bladder cell lines (T24, J82, 5637, RT-112, UMUC-3, and HT-1376) with ellipse showing 95% confidence interval and (b) the corresponding CV 1 loadings plot.

ively. Example mean spectra can be found in the ESI, section 1.†

The PC-CVA scores plot for the first two CVs is shown in Fig. 5(a) which shows a distinctive separation between the T24 and the other five cell lines. Loadings on the corresponding CV 1 is shown Fig. 5(b), the amide II/lipid band at 1555–1600 cm^{-1}, amide III peak at 1315 cm^{-1} and the tryptophan ring breathing at 764 cm^{-1} (indicated in red boxes) are weighting dominantly for separating the invasive T24 and the other non-invasive cell lines. However, the phenylalanine peak at ~1000 cm^{-1} also shows up in the loading plot as significant features. Since the phenylalanine peak is very sharp, tiny differences in peak shape or peak shift will show up in loading plots. This inconsistency in phenylalanine peaks is constantly observed and had been reported by various studies. Casabella *et al.* concluded that one of the reasons of this observation is because of the photon flux fluctuation between two adjacent pixels on a detector, which is more apparent in sharp peak then board peaks.[27] Instead of unavoidable instrumentation reasons, Li-Chan *et al.* reported that this phenomenon is caused by the conformation and macro-environment within a cell.[34] Loadings plots on other CVs can be found in the ESI section 2.†

To further investigate the results obtained, 5-fold CVA cross-validation was applied to the data. Instead of just focusing on the first two CVs, this cross-validation method will consider all five CVs available for a 6-class problem *i.e.* classifying the six bladder cell lines in a 5-dimension manner during the analysis. Table 2 shows the classification results and it suggests that the six cell lines can be classified with average rates of correct classification range from 54.4% to 100.0%. Especially for the T24, the average correctly classified rate can be reached up to 100.0%. The resultant confusion matrices of the five folds are independently displayed in the ESI section 3.†

One of the observations is that the classifier was also able to classify J82 with 92.5% accuracy which suggests there are significant differences between this cell line and the other cell lines, yet this is not showing up in the CV 1 and CV 2 dimensions. This raises a question on what feature(s) of cells is causing the separation observed in the PC-CVA scores plots. Therefore, further investigation was performed, and the results can be found in section 3.3.

3.2.2 Classified with Random Forest. The data used in this case was for all four replicates combined available for the six cell lines. 160 spectra were used for each cell line, with 128 spectra for training (80%) and 32 spectra for testing (20%). The corresponding classification results for the six cell lines are shown in Table 3.

As shown in Table 3, the six bladder cell lines can be classified with high rates of correct classification ranging from 85.3% to 100.0%. Importantly the invasive T24 cell line was able to be classified with 100.0% accuracy. This is reasonably consistent with the classification results generated by *k*-fold PC-CVA, apart from the sensitivity of the HT-1376 cell line. The sensitivity of the HT-1376 cell line increase from 54.4% to 85.3% when classified by Random Forest.

3.2.3 Independent tests using Random Forest. Independent tests were carried out with Random Forest using one of the four replicates as the test set and combining the other three as the training set; four classifiers were built in total.

The average classification results for the 6-class problem are shown in Table 4 where the confusion matrix for each inde-

Table 2 A confusion matrix showing the average taken from the five folds cross validation results for the six bladder cell lines (T24, J82, 5637, RT-112, UMUC-3, and HT-1376)

		True condition					
		T24	J82	5637	RT-112	UMUC-3	HT-1376
Prediction	T24	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	J82	0.0%	92.5%	8.1%	1.3%	0.6%	14.4%
	5637	0.0%	2.5%	79.4%	8.8%	3.8%	28.8%
	RT-112	0.0%	1.3%	6.3%	81.3%	3.1%	0.6%
	UMUC-3	0.0%	0.6%	0.6%	7.5%	91.9%	1.9%
	HT-1376	0.0%	3.1%	5.6%	1.3%	0.6%	54.4%

Table 3 Confusion matrix for the classification results of the six bladder cell lines

		True condition					
		T24	J82	5637	RT-112	UMUC-3	HT-1376
Prediction	T24	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	J82	0.0%	87.1%	0.0%	2.9%	0.0%	11.8%
	5637	0.0%	3.2%	96.4%	5.7%	6.3%	0.0%
	RT-112	0.0%	3.2%	0.0%	85.7%	0.0%	2.9%
	UMUC-3	0.0%	0.0%	0.0%	5.7%	93.8%	0.0%
	HT-1376	0.0%	6.5%	3.6%	0.0%	0.0%	85.3%

Table 4 Confusion matrix showing the average results of the four independent tests for the 6-class problem

		True condition					
		T24	J82	5637	RT-112	UMUC-3	HT-1376
Prediction	T24	87.5%	1.0%	1.4%	3.1%	2.5%	7.7%
	J82	0.0%	59.1%	9.7%	7.3%	2.2%	24.3%
	5637	0.8%	10.7%	33.7%	31.3%	4.0%	38.3%
	RT-112	8.3%	5.0%	14.4%	37.4%	14.2%	7.8%
	UMUC-3	1.1%	0.4%	4.1%	18.2%	75.3%	5.3%
	HT1376	2.3%	23.8%	36.8%	2.8%	1.9%	16.6%

pendent test can be found in the ESI section 3.† The average results show that the sensitivities for the cell lines are not always high. Especially for 5637 and RT-112, they can only be classified with accuracies just above 30%. In a 6-class problem, decisions can be made if the trees' voting percentage greater than 17%, but this is not convincing enough if aiming to bring this into clinical translation. The aim of this project is to build a model that is able to detect invasive cancers as early as possible, therefore achievement of an average sensitivity of 88% for the invasive T24 cell lines are more important.

3.3 Results of the classification of the seven bladder cancer cell lines

3.3.1 Classified with PC-CVA. The cisplatin resistant cell line, T24-CDDPR, was included in for the classifications using PC-CVA making the total number of cell lines in the data seven. This is used to verify which phenotype is CV1 from Fig. 6(a) responsible for. The resultant scores plot for CV 1 and CV 2 is shown in Fig. 6(a), where the corresponding CV 1 (the separation direction of the cell lines) loading plot is shown in Fig. 6(b). Loadings plots on other CVs can be found in the ESI section 2.†

The PC-CVA scores plot for the classification of the seven cell lines is shown in Fig. 6(a), where it shows that the separation between the T24, T24-CDDPR and the other five cell lines is mainly on the CV 1 axis. Since the cisplatin sensitivity of the T24-CDDPR cell line should be a lot lower than the T24 cell line, separation on CV 1 is less likely to be due to the cisplatin resistance of the cell lines. It is more probable that there are other features that are dominant in both T24 and T24-CDDPR than the other cell lines.

5-Fold CVA was carried out using this combined set of data and the results are presented as average values taken form the five folds in Table 5, where the individual resultants are presented in the ESI section 4.†

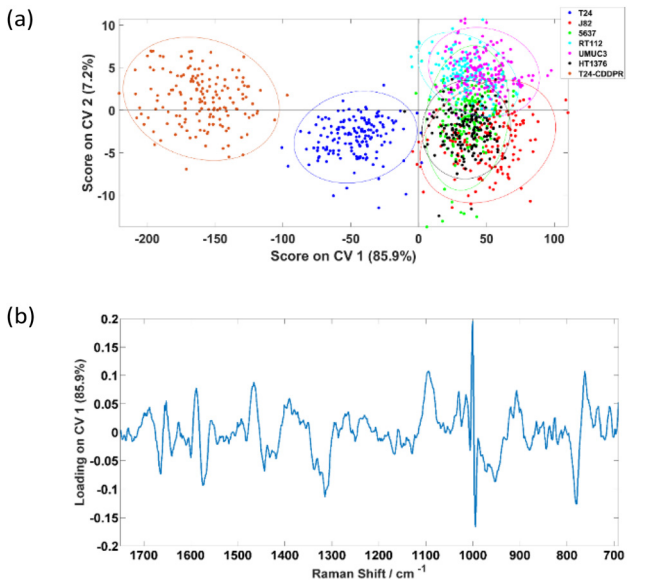

Fig. 6 PC-CVA (a) scores plot with CV 1 and CV 2 plotting against each other for the classification of the seven cell lines (T24, J82, 5637, RT-112, UMUC-3, HT-1376, and T24-CDDPR) with ellipse showing 95% confidence interval and (b) the corresponding CV1 loadings plot.

Table 5 shows that the invasive T24 and invasive resistant T24-CDDPR cell lines can be classified with very high accuracy of 98.1% and 100.0% respectively, hence LTRS can potentially be distinguishing the invasiveness of cell lines, and also able to identify the cisplatin sensitivities of the cell lines.

3.3.2 Classified with Random Forest. Similar sampling method was used with randomly selected 128 spectra (80%) as

Table 5 Confusion matric showing the average results of the five-fold cross validation of the seven cell lines (T24, J82, 5637, RT-112. UMUC-3. HT-1376, T24-CDDPR)

		True condition						
		T24	J82	5637	RT112	UMUC3	HT1376	T24-CDDPR
Prediction	T24	98.1%	0.6%	0.0%	0.0%	0.0%	0.0%	0.0%
	J82	1.3%	87.5%	10.0%	7.5%	1.3%	19.4%	0.0%
	5637	0.0%	5.0%	76.3%	6.9%	4.4%	27.5%	0.0%
	RT112	0.0%	1.3%	6.9%	76.9%	5.6%	0.6%	0.0%
	UMUC3	0.0%	1.3%	3.8%	8.8%	86.3%	2.5%	0.0%
	HT1376	0.0%	4.4%	3.1%	0.0%	2.5%	50.0%	0.0%
	T24-CDDPR	0.6%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%

Table 6 Confusion matrix of the classification results for seven bladder cell lines

		True condition						
		T24	J82	5637	RT-112	UMUC-3	HT-1376	T24-CDDPR
Prediction	T24	96.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	J82	0.0%	83.3%	0.0%	0.0%	0.0%	6.3%	0.0%
	5637	0.0%	5.6%	93.3%	3.1%	0.0%	3.1%	0.0%
	RT-112	0.0%	5.6%	0.0%	93.6%	0.0%	0.0%	0.0%
	UMUC-3	0.0%	0.0%	3.3%	3.1%	100.0%	3.1%	0.0%
	HT-1376	3.0%	5.6%	3.3%	0.0%	0.0%	87.5%	0.0%
	T24-CDDPR	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%

the training set and the remaining 32 spectra (20%) as the test set from the data. The classification results are shown as a confusion matrix in Table 6, and it shows that sensitivities of all the seven cell lines were achieved at 83.3% or above. For the cell lines of interest, the sensitivities of the T24 and cisplatin-resistant T24-CDDPR are 96.7% and 100.0% respectively. One interesting finding is that the T24 and T24-CDDPR cell lines had not been misclassified as each other in a 7-class problem. This again highlights the differences between the two cell lines.

3.3.3 Independent tests using Random Forest. Independent tests were again performed with the method descried previously and the resultant confusion matrix is shown in Table 7 and the corresponding results of each independent tests can be found in the ESI section 4.†

The cell lines of interest, T24 and T24-CDDPR, were classified with accuracy of 88.4% and 94.8% respectively. However, the correct classification rates of the remaining cell lines vary from 16.2–75.9% which is relatively low. In some cases, the accuracies go below 10%. This is believed to be caused by the complexity of the classifiers, in the fact that it is a seven-class problem.

4. Discussion

The aim of this work was to assess the ability of LTRS to distinguish different types of formalin-fixed bladder cell line with varied phenotypes. Although there are a range of publications reporting the use of Raman spectroscopy on the classifications between cancerous and non-cancerous samples, with some including classification on grades and stages of tumours, it had also been reported that the percentage agreement on grading and staging of a range of cancers between pathologists vary in different type of cancer, including bladder cancer.[35] From Fig. 5(a), it can be seen that the T24 cell line was signifi-

Table 7 Confusion matrix showing the average results of the four independent tests for the 7-class problem

		True condition						
		T24	J82	5637	RT-112	UMUC-3	HT-1376	T24-CDDPR
Prediction	T24	88.4%	1.6%	0.0%	5.0%	1.6%	5.1%	3.4%
	J82	0.7%	64.7%	7.8%	8.3%	2.0%	25.5%	0.0%
	5637	0.8%	8.9%	40.7%	28.9%	4.5%	38.3%	0.0%
	RT-112	7.7%	3.6%	14.2%	38.0%	13.1%	8.5%	1.2%
	UMUC-3	0.8%	0.6%	2.7%	17.0%	75.9%	6.4%	0.6%
	HT-1376	1.6%	20.7%	34.6%	2.8%	2.9%	16.2%	0.0%
	T24-CDDPR	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	94.8%

cantly different to the other five cell lines. It was first suspected this difference was related to the cells' invasiveness and/or cisplatin sensitivity. However, according to the characterisation results in Fig. 5(a), if separation was caused by the invasiveness of cells, J82 which is the second most invasive cell line following T24, should separate more towards the T24. Similar observations were acquired by Kujdowicz *et al.* that the two subtypes of T24 are highly distinguishable when analysing by Raman and FTIR spectroscopy.[20] One other possible reason of causing this separation pattern is that the classifier is picking up changes in the Raman signal related to cells' cisplatin sensitivity in which T24 is significantly more sensitive than other cell lines analysed. To elucidate this possibility further a cisplatin resistant T24 cell line, T24-CDDPR, was added into the analysis. There are significant changes in protein between the T24 and resistant T24-CDDPR cell lines according to the Agarose 2-DE and MS/MS analysis results by Taoka *et al.*,[22] and according to the study by Sun *et al.*,[36] the selection of cisplatin resistance in the T24 cell line enhanced proliferation, invasion and malignant behaviour. The introduction of cisplatin resistance in T24 and J82 leads to massive increase in MDR1 and HIF-1α and increase in aggressive behaviour. This is an indicative of these cells facing a more stem-like phenotype. Therefore it is reasonable to expect similar increase in aggressive invasion behaviour in the T24-CDDPR, even though it is a different cisplatin resistant T24 cell line. This concurs with our results show in Fig. 6(a) that T24-CDDPR is separated from both T24 and other cell lines. Although similar behaviours also observed in J82 and cisplatin resistant J82, T24 and cisplatin resistant T24 are still significantly more invasive then J82 cell lines according to the same study by Sun *et al.*[36]

However, results in Fig. 6(a) shows that the separation along CV 1 is unlikely to be caused by the cisplatin sensitivity of the cell lines as the T24-CDDPR does not cluster with the J82, 5637, RT-112, UMUC-3, and HT-1376. It separates in the same direction with the T24 as a different data cluster on the negative CV 1. This indicates that the T24-CDDPR is very likely to have the invasive property as the T24 as they both separate towards the same direction on the CV 1. The distinction observed between T24-CDDPR and T24 is suspected to be caused by an enriched characteristic in the T24-CDDPR caused during the cisplatin resistant development. Yang *et al.* showed that the T24 cell line, overexpressed the long noncoding RNA (lncRNA) ASAP1-IT1, similar to bladder cancer tissues, which plays a role in maintaining cell stemness. Although J82, 5643 and UMUC3 cell lines do express ASAP1-IT1 it is at a significantly lower level than T24, and overexpression of ASAP1-IT1 in T24 induces a stem like phenotype.[37] It is therefore possible that the separation of the T24 and the cisplatin resistant derivative T24-CDDPR is based upon their more stem like phenotype. Although the results acquired show that the invasiveness and drug resistance of these cells are extremely complex, in which a single component from chemometric analysis is not enough to explain the observations, clear separations between clusters of different type of cell lines were achieved. This is partly caused by the fact that there is lack of information on standardising the development of the drug-resistant T24 cell lines and corresponding study on their metabolisms, further analysis on these cell lines has to be done to confirm to test the hypothesis.

The ultimate aim of this study is to demonstrate the feasibility of using LTRS to identify different types of bladder cancer cells from urine samples. The samples used in this work presented were the formalin fixed cells, which is different from the cells that can be found in urine sample or in the *in vivo* condition. Despite this formalin fixation can preserve the biochemical information of cells and can significantly reduce the cells' stress response to photochemical oxidative damage.[38] This is very important when this study is aimed at performing preliminary test on whether the LTRS can be used for identifying different types of bladder cancer. Non-fixed cells and cells exposed in urine should then be used to mimic the real cells situation that can be found in urine samples in future research on this topic. However, this work with using formalin fixed cells will allow a starting point for demonstrating that LTRS can be used in real time capture and analysis of different phenotypic features of bladder cancer cells in a urine streamline.

5. Conclusions

This work presented has successfully demonstrated the use of LTRS system to classify different types of formalin fixed bladder cancer cells with high sensitivity when classifying with PC-CVA. Complementary classifications with using Random Forest were also performed and consistent results were obtained. However multiple replicates in Random Forest analysis will be needed to build a better model. As in an early stage of a pilot study of the development of using LTRS to detect bladder cancer cells in urine streamline, this work shows the positive outcomes on separating the cell line with invasive phenotype. It shows that phenotypic feature is more dominant than the internal variability within cell lines. Further work using a wider range of non-fixed urological cells and detailed study on the separation components is necessary if the technique is to be translated into clinical applications.

Author contributions

N. T. T. performed the bulk of the experimental work. R. R. performed the experimental work on the invasion assay and cytotoxicity assay, with experimental design and project supervision by N. W. C. and M. D. B., N. T. T. wrote the main body of the manuscript. P. G. supervised the project and handled logistics. N. T. T., R. D. S., M. D. B. and P. G. were equally involved in experimental design. M. D. B. was involved in the communication and logistic of obtaining the drug resistant T24-CDDPR cell line and contributed to the writing of the manuscript. R. D. S. was involved in experimental design and input into Raman instrument configuration. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We would like to thank the Genito Urinary Cancer Research Group for the donation of the cell lines for this study. Moreover, we would like to acknowledge Dr Kazumasa Matsumoto from the Kitasato University for the donation of the cisplatin resistant cell line T24-CDDPR.

References

1 Cancer Research UK, Prostate cancer incidence statistics, https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer/incidence#heading-Two, (accessed 3 April 2023).
2 H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal and F. Bray, *Ca-Cancer J. Clin.*, 2021, **71**, 209–249.
3 Cancer Research UK, Bladder cancer - Screening - Cancer Research UK, https://www.cancerresearchuk.org/about-cancer/bladder-cancer/getting-diagnosed/screening, (accessed 9 May 2017).
4 C. Z. Zhu, H. N. Ting, K. H. Ng and T. A. Ong, *J. Cancer*, 2019, **10**, 4038–4044.
5 S. J. Galgano, S. Rais-Bahrami, K. K. Porter and C. Burgan, *Diagnostics*, 2020, **10**, 1–16.
6 H. H. Lee and S. H. Kim, *Transl. Cancer Res.*, 2020, **9**, 6554–6564.
7 M. Matuszczak, A. Kiljańczyk and M. Salagierski, *Int. J. Mol. Sci.*, 2022, **23**, 8597.
8 B. J. Schmitz-Dräger, L. A. Tirsar, C. Schmitz-Dräger, J. Dörsam, Z. Mellan, E. Bismarck and T. Ebert, *World J. Urol.*, 2008, **26**, 31–37.
9 R. Malinaric, G. Mantica, L. Lo Monaco, F. Mariano, R. Leonardi, A. Simonato, A. Van der Merwe and C. Terrone, *Int. J. Environ. Res. Public Health*, 2022, **19**, 9648.
10 R. D. Snook, T. J. Harvey, E. Correia Faria and P. Gardner, *Integr. Biol.*, 2009, **1**, 43–52.
11 E. Canetta, M. Mazilu, A. C. De Luca, A. E. Carruthers, K. Dholakia, S. Neilson, H. Sargeant, T. Briscoe, C. S. Herrington and A. C. Riches, *J. Biomed. Opt.*, 2011, **16**, 037002.
12 E. Canetta, A. Riches, E. Borger, S. Herrington, K. Dholakia and A. K. Adya, *Acta Biomater.*, 2014, **10**, 2043–2055.
13 L. T. Kerr, A. Adams, S. O. Dea, K. Domijan, I. Cullen and B. M. Hennelly, *Photon. Solut. Better Health Care*, 2014, **9129**, 1–8.
14 T. J. Harvey, E. Correia Faria, A. Henderson, E. Gazi, A. D. Ward, N. W. Clarke, M. D. Brown, R. D. Snook and P. Gardner, *J. Biomed. Opt.*, 2008, **13**, 064004-1–064004-12.
15 S. Casabella, P. Scully, N. Goddard and P. Gardner, *Analyst*, 2016, **141**, 689–696.
16 S. Dochow, C. Krafft, U. Neugebauer, T. Bocklitz, T. Henkel, G. Mayer, J. Albert and J. Popp, *Lab Chip*, 2011, **11**, 1484–1490.
17 I. W. Schie, R. Kiselev, C. Krafft and J. Popp, *Analyst*, 2016, **141**, 6387–6395.
18 N. Pavillon and N. I. Smith, *J. Biomed. Opt.*, 2015, **20**, 016007.
19 T. J. Harvey, *The Development of Vibrational Spectroscopic Cytology for Prostate Cancer Diagnosis*, University of Manchester, 2008.
20 M. Kujdowicz, W. Placha, B. Mech, K. Chrabaszcz, K. Okoń and K. Malek, *Cancers*, 2021, **13**, 1–20.
21 N. T. Tang, R. D. Snook, M. D. Brown, B. A. Haines, A. Ridley, P. Gardner and J. L. Denbigh, *Molecules*, 2020, DOI: 10.3390/molecules25071652.
22 Y. Taoka, K. Matsumoto, K. Ohashi, S. Minamida, M. Hagiwara, S. Nagi, T. Saito, Y. Kodera and M. Iwamura, *Biomed. Res.*, 2015, **36**, 253–261.
23 C. A. Hart, M. Brown, S. Bagley, H. Sharrard and N. W. Clarke, *Br. J. Cancer*, 2005, **92**, 503–512.
24 Z. Huang, A. McWilliams, S. Lam, J. English, D. I. McLean, H. Lui and H. Zeng, *Int. J. Oncol.*, 2003, **23**, 649–655.
25 M. G. Shim and B. C. Wilson, *Photochem. Photobiol.*, 1996, **63**, 662–671.
26 F. Lyng, E. Gazi and P. Gardner, in *Biomedical Applications of Synchrotron Infrared Microspectroscopy: A Practical Approach*, The Royal Society of Chemistry, 2011, pp. 145–191.
27 S. Casabella, *Development of automated analysis and sorting of single cells using Laser Tweezers Raman Spectroscopy*, University of Manchester, 2015.
28 A. Henderson, CHI Toolbox, 2016.
29 A. M. C. Daviesa and T. Fearnb, *Spectrosc. Eur.*, 2008, **20**, 18–20.
30 L. Nørgaard, R. Bro, F. Westad and S. B. Engelsen, *J. Chemom.*, 2006, **20**, 425–435.
31 C. Peltier, M. Visalli and P. Schlich, *Food Qual. Prefer.*, 2015, **40**, 326–333.
32 G. E. A. P. A. Batista, R. C. Prati and M. C. Monard, *A Study of the Behavior of Several Methods for Balancing Machine Learning*, Training Data, 2004.
33 B. W. Yap, K. A. Rani, H. A. A. Rahman, S. Fong, Z. Khairudin and N. N. Abdullah, *An Application of Oversampling, Undersampling, Bagging and Boosting in Handling Imbalanced Datasets*, 2014, pp. 13–22.
34 E. Li-Chan, J. M. Chalmers and P. R. Griffiths, *Applications of Vibrational Spectroscopy in Food Science*, John Wiley & Sons, 2010.
35 E. Compérat, A. Oszwald, G. Wasinger, D. E. Hansel, R. Montironi, T. van der Kwast, J. A. Witjes and M. B. Amin, *World J. Urol.*, 2022, **40**, 915–927.
36 Y. Sun, Z. Guan, L. Liang, Y. Cheng, J. Zhou, J. Li and Y. Xu, *Oncol. Rep.*, 2016, **35**, 1549–1556.
37 L. Yang, Y. Xue, J. Liu, J. Zhuang, L. Shen, B. Shen, J. Yan and H. Guo, *Neoplasma*, 2017, **64**, 847–855.
38 T. J. Harvey, C. Hughes, A. D. Ward, E. Correia Faria, A. Henderson, N. W. Clarke, M. D. Brown, R. D. Snook and P. Gardner, *J. Biophotonics*, 2009, **2**, 47–69.