

## Article

# Rapid Detection of Recurrent Non-Muscle Invasive Bladder Cancer in Urine Using ATR-FTIR Technology

Abdullah I. El-Falouji <sup>1,\*</sup>, Dalia M. Sabri <sup>1</sup>, Naira M. Lotfi <sup>1</sup>, Doaa M. Medany <sup>1</sup>, Samar A. Mohamed <sup>1</sup>, Mai Alaa-eldin <sup>1</sup>, Amr Mounir Selim <sup>2</sup>, Asmaa A. El Leithy <sup>3</sup> , Haitham Kalil <sup>4,5</sup> , Ahmed El-Tobgy <sup>6</sup> and Ahmed Mohamed <sup>7</sup> 

<sup>1</sup> Institute of Biotechnology for Postgraduate Studies and Research (IBPR), Suez Canal University, Ismailia 41522, Egypt

<sup>2</sup> Surgical Oncology Department, National Cancer Institute, Cairo University, Cairo 11796, Egypt

<sup>3</sup> College of Biotechnology, Misr University for Science and Technology, Giza 12562, Egypt

<sup>4</sup> Chemistry Department, Faculty of Science, Suez Canal University, Ismailia 41522, Egypt

<sup>5</sup> Chemistry Department, Cleveland State University, Cleveland, OH 44115, USA

<sup>6</sup> Department of Urology, Faculty of Medicine, Suez Canal University, Ismailia 41522, Egypt

<sup>7</sup> Division of Bioinformatics and Colonial Foundation Healthy Ageing Centre, Walter and Eliza Hall Institute of Medical Research, Melbourne, VIC 3052, Australia

\* Correspondence: falouji@hotmail.com or abdullah\_ismail@suez.edu.eg; Tel.: +20-111-599-0299



**Citation:** El-Falouji, A.I.; Sabri, D.M.; Lotfi, N.M.; Medany, D.M.; Mohamed, S.A.; Alaa-eldin, M.; Selim, A.M.; El Leithy, A.A.; Kalil, H.; El-Tobgy, A.; et al. Rapid Detection of Recurrent Non-Muscle Invasive Bladder Cancer in Urine Using ATR-FTIR Technology. *Molecules* **2022**, *27*, 8890. <https://doi.org/10.3390/molecules27248890>

Academic Editors: Angela Staicu, Adriana Smarandache, Tatiana Tozar, Andra Dinache and Maria Paula Marques

Received: 30 September 2022

Accepted: 9 December 2022

Published: 14 December 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Non-muscle Invasive Bladder Cancer (NMIBC) accounts for 80% of all bladder cancers. Although it is mostly low-grade tumors, its high recurrence rate necessitates three-times-monthly follow-ups and cystoscopy examinations to detect and prevent its progression. A rapid liquid biopsy-based assay is needed to improve detection and reduce complications from invasive cystoscopy. Here, we present a rapid spectroscopic method to detect the recurrence of NMIBC in urine. Urine samples from previously-diagnosed NMIBC patients ( $n = 62$ ) were collected during their follow-up visits before cystoscopy examination. Cystoscopy results were recorded (41 cancer-free and 21 recurrence) and attenuated total reflection Fourier transform infrared (ATR-FTIR) spectra were acquired from urine samples using direct application. Spectral processing and normalization were optimized using parameter grid searching. We assessed their technical variability through multivariate analysis and principal component analysis (PCA). We assessed 35 machine learning models on a training set (70%), and the performance was evaluated on a held-out test set (30%). A Regularized Random Forests (RRF) model achieved a 0.92 area under the receiver operating characteristic (AUROC) with 86% sensitivity and 77% specificity. In conclusion, our spectroscopic liquid biopsy approach provides a promising technique for the early identification of NMIBC with a less invasive examination.

**Keywords:** liquid biopsy; bladder cancer; FTIR; machine learning

## 1. Introduction

Early, non-invasive, and conclusive disease identification relates to several favorable outcomes. These include reducing the strain on healthcare expenditures that are already stretched tight since the COVID-19 pandemic and employing therapeutic procedures that reduce morbidity and death at an early stage. There are indications that some routinely-employed clinical tests are unsuitable or misleading. In addition, they typically employ inadequate single illness markers when multiple factors are at play. The growing subject of metabolomics includes a technique known as metabolic fingerprinting, which is a form of high throughput and is the universal analysis used to differentiate samples swiftly and correctly. The differentiation is based on the change in a selective area as a result of different statuses of disease or changes in the biological environment. Using infrared or Raman spectroscopy, metabolic fingerprinting has been presented as a potential technique and a powerful tool for disease diagnostics [1].

Approximately 70–75% of urothelial carcinoma cases are classified as non-muscle-invasive bladder cancer (NMIBC). In NMIBC, the tumor is not invading the muscularis propria and is confined to mucosa and submucosa layers [2,3]. NMIBC is characterized by a high recurrence rate of 60–70% while, in 10–20% of patients, the tumor will develop a progression to muscle invasive carcinoma (MIBC).

Due to the high recurrence rate, a follow-up using a cystoscopy examination is required frequently according to European Association of Urology (EAU) Guidelines [4]. However, cystoscopy could give an indefinite diagnosis in patients with active inflammation, indwelling catheter, or abnormal appearance of the bladder mucosa. Besides, it is an invasive, painful, and uncomfortable procedure with the risk of causing urinary tract infections in 10% of the patients [5,6]. Therefore, a non-invasive liquid biopsy marker for bladder carcinoma diagnosis is needed.

There is great potential for using urine-based tumor markers as a non-invasive, affordable method for diagnosing and monitoring the progression of tumors, replacing the use of cystoscopy in tumor follow-up [7]. Urine samples have distinct advantages such as non-invasive sampling, and the explicit correlation between some proteins in urinary proteome and the development of diseases [8–10].

Biochemical profiling of clinical samples using attenuated total reflectance–Fourier transform infrared (ATR-FTIR) spectroscopy is an easy analytical technique that does not require sample preparation [11,12]. Different spectrum regions correlate with different biochemical moieties, and the investigation of such regions can potentially reveal altered biochemical molecule classes. More importantly, IR fingerprint regions can represent a unique snapshot of the complex biochemical profiles of clinical samples. Such fingerprints, when combined with non-linear machine learning models, can identify discriminant features that are associated with health and disease [13,14]. Spectral profiling of biofluids such as blood, urine and saliva is an extremely promising source of liquid biopsy-based assays [15,16].

The ATR-FTIR spectroscopy-based liquid biopsy test, a vibrational spectroscopy techniques was used in a recent clinical trial on suspected brain cancer patients, alongside clinical assessment in primary care to help achieve earlier cancer detection and diagnosis [17]; this technique gives more accurate classification for early-stage tumors than other liquid biopsy approaches that are based on tumor genetic material such as circulating DNA [18]. Any changes in FTIR peaks could be a consequence of the induced significant alterations in molecular characteristics as well as arrangement, and structure and dynamics in tissues, membranes, and cells. These changes could help differentiate between diseased and non-diseased patients [12].

FTIR spectroscopy is a promising tool to diagnose a wide range of diseases such as Alzheimer's [19], prostate cancer [20], bladder cancer [21], and viruses such as COVID-19 [22].

Therefore, ATR-FTIR spectroscopy has the potential to become a useful diagnostic tool that is deliverable and feasible in the healthcare system [23,24].

Here, we have examined the performance of a rapid spectroscopy-based liquid biopsy test in relation to tumor recurrence in non-muscle invasive bladder cancer patients. The spectral data coupled with machine learning algorithms have been used to differentiate between this recurrent cancer and non-recurrence (control) patients to identify an FTIR fingerprint for NMIBC recurrence.

## 2. Results

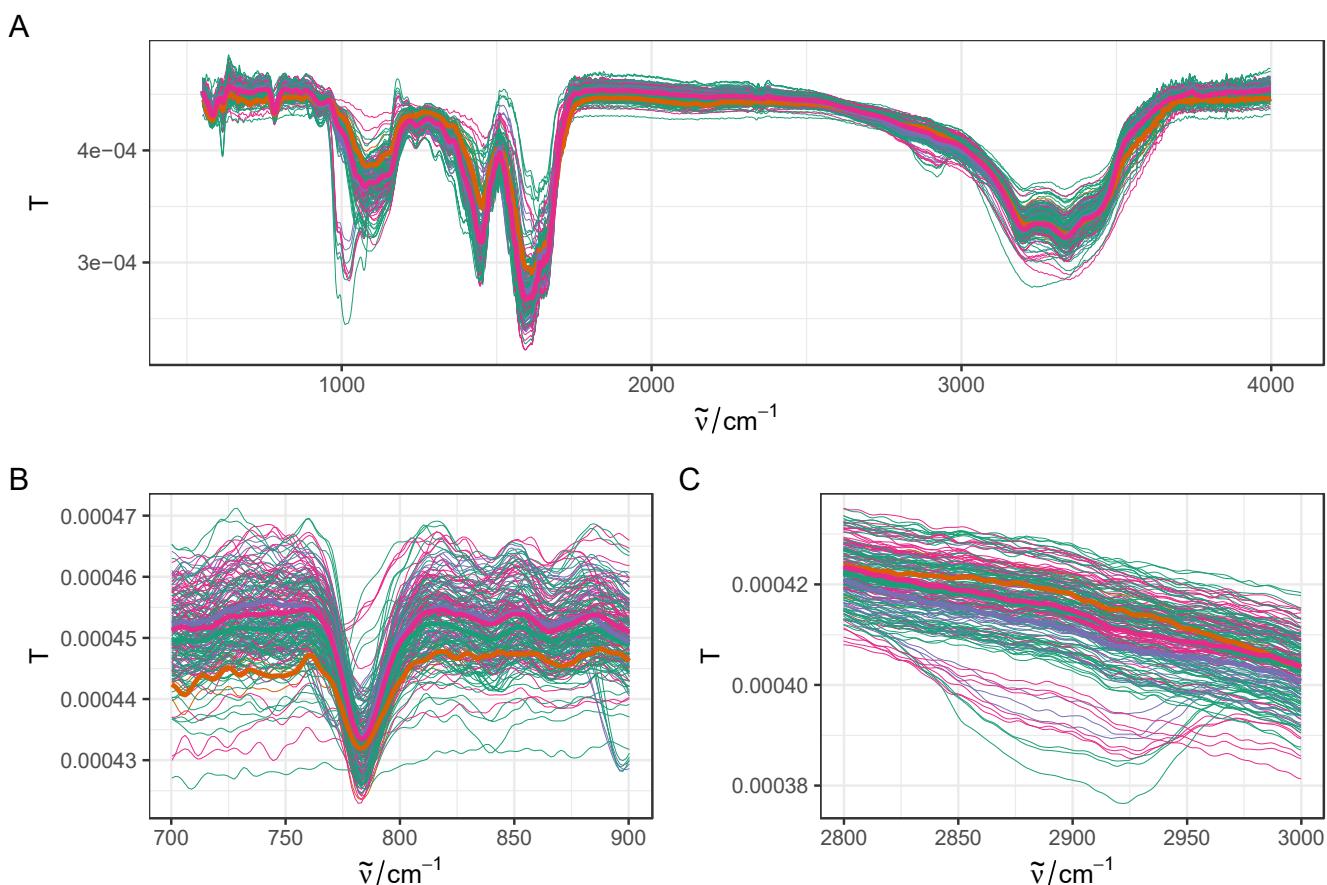
### 2.1. Overview FTIR Spectra Dataset

Urine samples from previously diagnosed NMIBC patients were collected during their first follow-up visit prior to cystoscopy examination. The pathological results of cystoscopy were recorded for a total of 62 individuals (Table 1). Since the aim of this study was to detect NMIBC recurrence, we further collapsed the patient samples into two groups, recurrence ( $n = 21$ ) and NMIBC-free, which includes all other cystoscopy results ( $n = 41$ ). Spectroscopic FTIR measurements were recorded at least in triplicate, and exported from

Bruker software (Opus 6.5) after baseline correction, as described in the Methods section. Empty and corrupted spectra files were discarded, resulting in a total of 187 spectra that were included in the downstream analysis. For initial data exploration, raw spectra were first vector-normalized and different regions were visualized as shown in Figure 1.

**Table 1.** Cystoscopy results for the cohort.

Cystoscopy Result	Number of Patients
Free	31
Hyperplasia	1
Inflammation	9
Recurrence	21



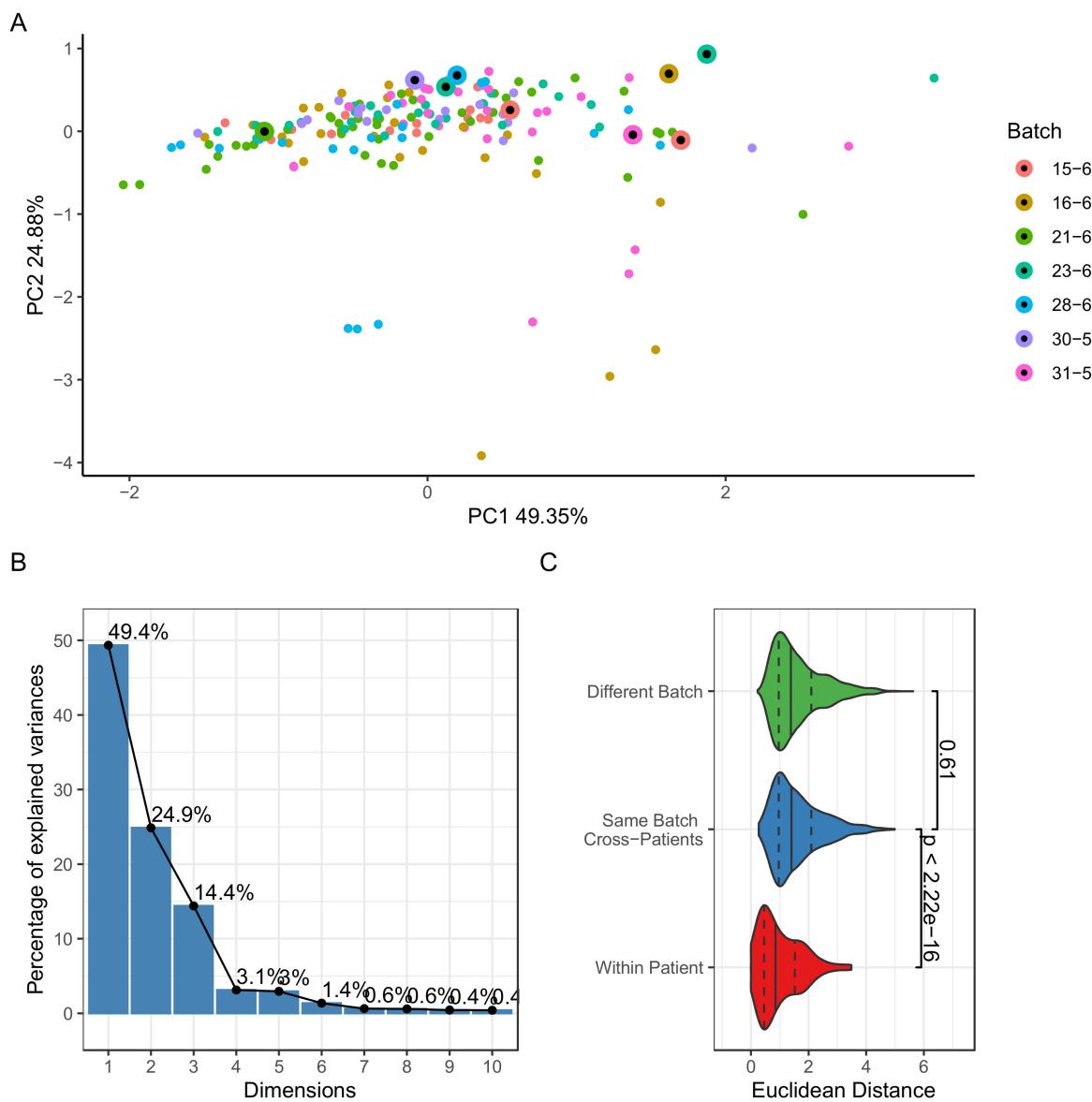
**Figure 1.** Overview of FTIR dataset (baseline corrected and vector normalized), with full range (A), fingerprint region (B), and hydrocarbon C-H stretching region (C). Bold lines represent average spectra in each group.

## 2.2. Assessment of Technical Variability and Batch Effects

Spectra were acquired on seven different days, which may introduce technical variability batch effects, affecting the downstream analysis. Several studies and tools have investigated batch effects in FTIR, which can arise from atmospheric changes such as humidity and temperature, operator variability, or instrument calibration [13,25]. We addressed the batch effects by randomizing the acquisition with respect to the NMIBC status. Additionally, one of the samples was designated as a quality control (QC) sample and measured repeatedly in all batches.

We examined the PCA plots and colored the samples by batch (Figure 2A,B). The plots revealed a high variance and explained the first principal component (49.53%); however, they did not show any remarkable separation between samples from different batches.

QC samples (Figure 2A, black dot markers) were not clustered with samples from several batches. Taken together, the results did not indicate any significant batch effects, despite the high variability within the dataset.



**Figure 2.** Assessment of technical variability and batch effects. (A) Principal component analysis showing the first two components colored by batch. Quality control samples are marked with black dots. (B) Scree plot showing variance explained by top 10 principal components. (C) Violin plots showing the distribution of Euclidean distances between spectra from the same patient, across patients and batches. The vertical lines represent the first, second, and third quartiles. Wilcoxon Rank sum  $p$  values are shown between groups.

Next, we investigated whether the technical variability was less than the biological variability by comparing spectra acquired from replicates and across different samples and batches. The distribution of pairwise Euclidean distance was used to evaluate how close the spectra were. Figure 2C confirmed that the distances between the replicate spectra were significantly lower than for spectra from patients ((Wilcoxon-rank sum test,  $p < 2.22 \times 10^{-16}$ ). We concluded that the technical variability was lower than the biological signal pursued in this analysis.

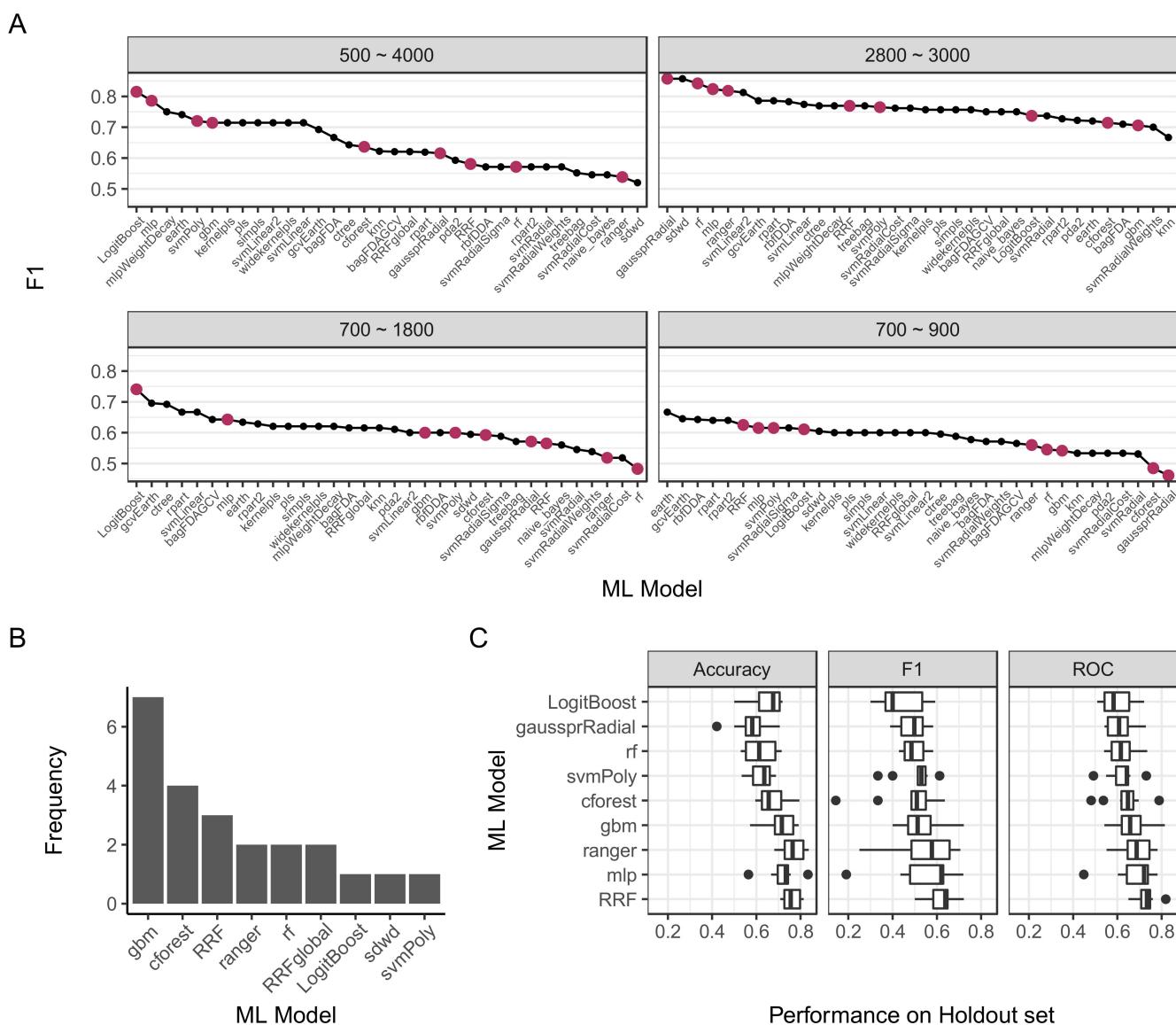
### 2.3. Preprocessing Parameters and Model Selection

To create a classifier for NMIBC status using FTIR spectra, we created a two-class response variable, *recurrence*, to denote NMIBC diagnosis by Cystoscopy ( $n = 21$ ), and *free* to include all other diagnoses ( $n = 41$ ). Using recurrence as the positive response class, we then assessed 35 different machine learning models from the *caret* R package. Since feature selection and spectral preprocessing can significantly alter the performance of the assessed models, we sought optimal processing parameters for each model. Through a hyperparameter grid search, we evaluated 25,200 combinations, optimizing spectrum range, normalization, binning, [26] filter derivative, and window size (Table 2). Notably, the selected four spectrum ranges correspond to biological regions of interest, full spectrum  $500\text{--}4000\text{ cm}^{-1}$ , bacterial fingerprint region  $700\text{--}900\text{ cm}^{-1}$ , extended fingerprint region  $700\text{--}1800\text{ cm}^{-1}$ , and hydrocarbon C-H stretching region  $2800\text{--}3000\text{ cm}^{-1}$ .

**Table 2.** Preprocessing parameters for spectra.

Preprocessing Parameter	Variants
Spectrum Range	$500\text{--}4000\text{ cm}^{-1}$
	$700\text{--}900\text{ cm}^{-1}$
	$700\text{--}1800\text{ cm}^{-1}$
	$2800\text{--}3000\text{ cm}^{-1}$
Savitzky-Golay Derivative	0, 1, 2
Normalization	No normalization
	Amide $1500\text{--}1700\text{ cm}^{-1}$
	Urea $1400\text{--}1500\text{ cm}^{-1}$
Savitzky-Golay Window Size	5, 7, 9, 13
Bin size	1, 2, 3, 5, 10

Different numbers of patients in the NMIBC-free ( $n = 41$ ) and recurrence ( $n = 21$ ) groups created a class imbalance that can affect model training. We addressed the class imbalance using two techniques. First, we used ROC as a performance metric to optimize the classifier models. Second, we utilized the synthetic minority over-sampling technique (SMOTE) [27] to impute new data points for the minority class, recurrence in our case. The preprocessing combinations for each model were assessed on the holdout test set and the top performing combination was plotted in Figure 3A. Based on performance on the test set, we selected five models for further validation, *gaussprRadial*, *rf*, *LogitBoost*, *mlp*, and *svmPoly*.



**Figure 3.** Machine learning model selection and validation. (A) Top performance F1 score for each model against different spectrum ranges. Maroon dots represent models selected for further validations. (B) Models with high accuracy in both training set, as estimated by ROC > 0.8 in cross-validation, and in the test set (F1 score > 0.5). Frequency indicates the number of processing parameter combinations in a specific model achieved by these criteria. (C) Performance scores of selected models on test sets. Dataset splitting and model fitting were repeated 10 times, and performance was evaluated with three metrics.

Because of the large number of combinations that were evaluated relative to the dataset size, dataset splitting into training and test sets may incorrectly inflate performance in the holdout set. In such cases, high test set performance correlated with a poor training set performance, which indicates poor model robustness (Table S1). To address this issue, we introduced two additional criteria: (1) requiring good performance in both training and test set, and (2) the model should give similar high performance in similar pre-processing conditions. Consequently, we filtered models with training ROC > 0.8 and F1 test performance > 0.5, and selected four additional models that frequently matched these criteria (Figure 3B), gbm, cforest, RRF, and ranger. The final list of selected models and processing parameters is shown in Table 3.

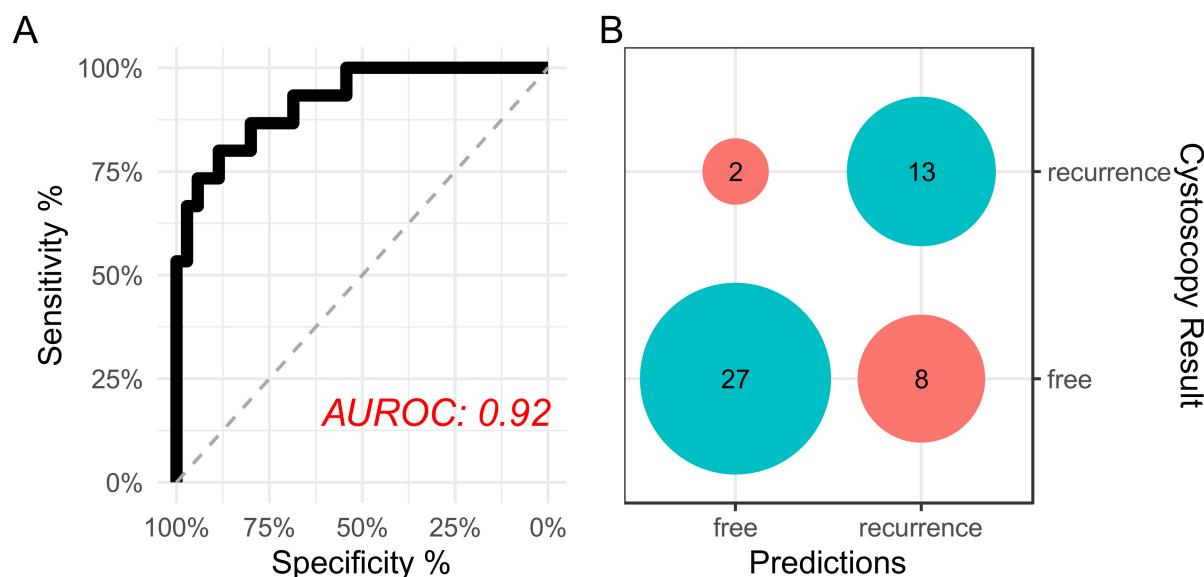
**Table 3.** Models selected for further validation with their processing parameters.

Model	CV ROC	Test Accuracy	F1 Score	Spectrumrange	SG Derivative	SG Window	Bin Size	Normalization Peak
cforest	0.83	0.74	0.62	2800~3000	1	13	1	urea
gbm	0.81	0.79	0.67	2800~3000	1	7	1	none
ranger	0.82	0.77	0.64	2800~3000	1	13	1	urea
RRF	0.82	0.81	0.71	2800~3000	1	13	1	urea
gaussprRadial	0.46	0.92	0.86	2800~3000	0	9	1	none
LogitBoost	0.53	0.88	0.81	500~4000	1	9	3	amide
mlp	0.67	0.85	0.82	2800~3000	0	7	2	amide
rf	0.42	0.92	0.84	2800~3000	0	5	2	none
svmPoly	0.7	0.85	0.76	2800~3000	2	7	10	amide

CV: cross-validation; SG: Savitzky-Golay filter; model abbreviations are provided in Table S3.

#### 2.4. Model Validation and Tuning

To assess the robustness of the model against dataset splitting, we repeated model training and testing 10 times for each of the selected models, where we reconstructed different training and test sets for each iteration by setting different random number generator seeds. We subsequently assessed the performance of the models using three metrics, accuracy, F1, and AUROC. Based on the results shown in Figure 3C, we concluded that the regularized random forest (RRF) model was the best-performing model. To investigate RRF model tuning, we investigated the constructed models (Table S2), which indicated 2 as the optimal number of randomly selected predictors, a regularization value of 1.000, and an importance coefficient of 0.0. The final selected model achieved an AUROC of 92% with 86% sensitivity and 77% specificity (Figure 4).

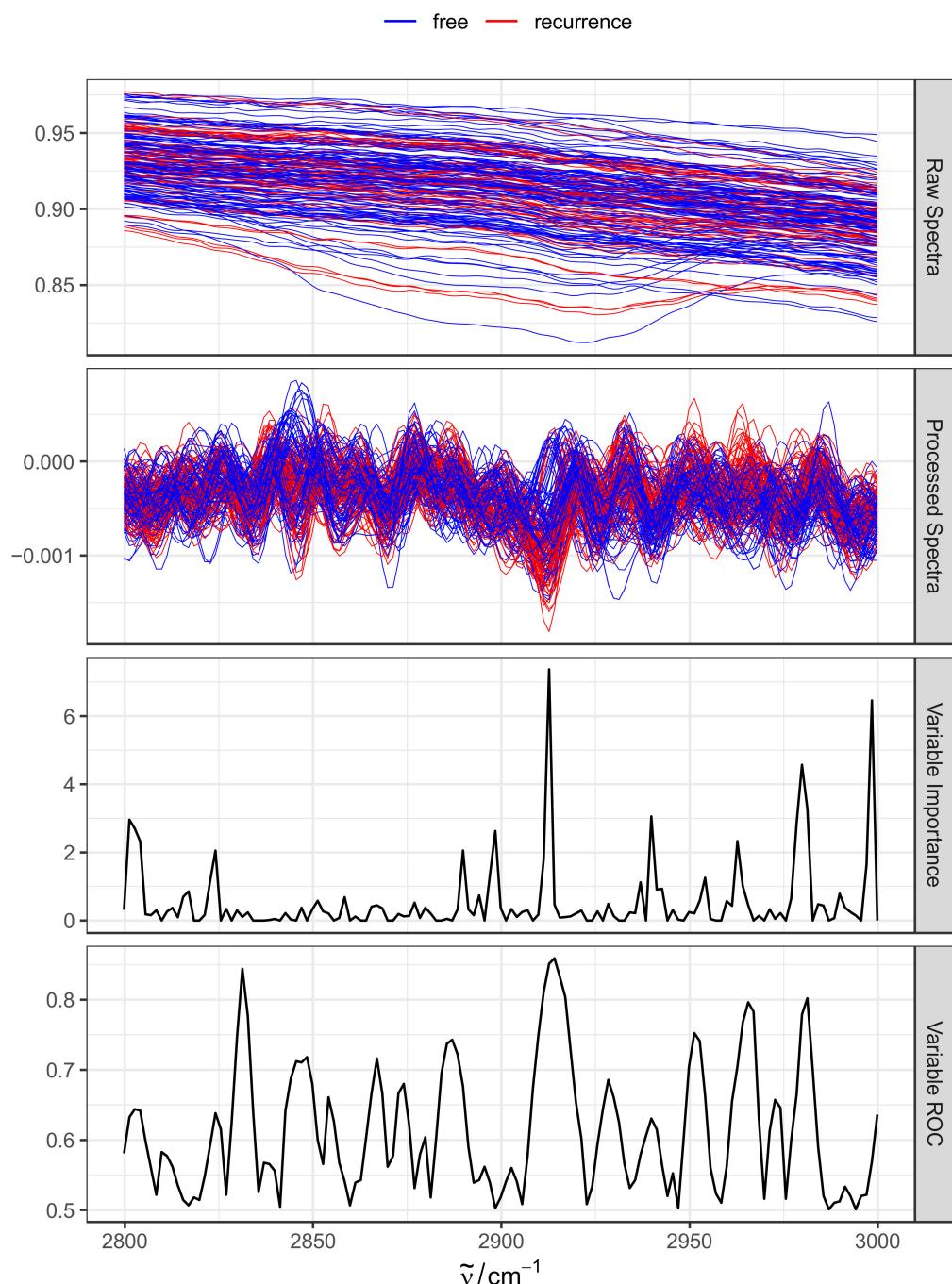


**Figure 4.** Final model performance. (A) Receiver operating characteristic curve showing the area under the curve. The curve was constructed using class probabilities obtained from the final RRF model against the holdout test set; (B) confusion matrix of the performance of the binary classifier.

#### 2.5. Extraction of Feature Importance

Random forest models enable estimating variable importance by shuffling predictor values and measuring the performance using out-of-bag samples [28,29]. This technique not only allows for the calculation of per-feature importance scores, but also ROC estimates. We exploited this feature to calculate importance and ROC estimates, and to visualize

them alongside spectra. As shown in Figure 5, variable importance and predictive ROC spikes around  $2912\text{ cm}^{-1}$  were characteristic of lipid CH<sub>2</sub> asymmetrical stretching. An additional spike is also observed around  $2980\text{ cm}^{-1}$ , reflecting the stretching vibrations of methyl hydrocarbon chains. An investigation of processed spectra (Figure 5, second top panel) clearly reveals spectral differences between recurrence (red) and NMIBC-free (blue) samples. Taken together, the variable importance results suggest an altered lipid profile in urine in recurrent NMIBC.



**Figure 5.** Investigation of wavenumber ranges contributing to model performance. Raw and processed spectra for the selected spectrum range (top two panels) are shown, colored by NMIBC status. Variable importance scores (third panel) are directly obtained from the model using out-of-bag estimation strategy. Variable ROC scores are obtained via similar strategy but evaluating estimates against the test set.

### 3. Discussion

Non-muscle invasive bladder cancer (NMIBC) is low-grade urothelial carcinoma that is commonly treated by transurethral resection (TURBT). While not life threatening, the high risk of recurrence and progression for muscle-invasive bladder cancer necessitates regular patient follow-ups to detect and prevent progression. However, the gold standard for NMIBC detection still relies solely on cystoscopy, an invasive procedure and source of major discomfort for patients.

It has been estimated that urine markers would need to have a sensitivity between 90–95% to replace cystoscopy [30,31]. Several cytological assays have been reported previously, with sensitivities ranging from 7–93% [32]. Other multivariate biomarkers have been suggested based on gene expression studies [33,34]; however, none were able to achieve clinical utility [4].

In this study, we attempted to establish a rapid, easy-to-administer, and non-invasive assay for the detection of NMIBC in follow-up patients. The simple assay relies on FTIR spectroscopic measurement of urine samples with no sample preparation required. Using cystoscopy examination and pathological investigation of biopsies as the ground truth, we used machine learning to establish an accurate classifier from acquired spectra. The achieved sensitivity was 86%. Although our study did not reach the sensitivity sought after for clinical application, this is a pilot study and performance can be strongly affected by outliers and inter-individual variations. Nevertheless, these promising results present a great potential for improvement with a larger cohort size.

Exploring the machine learning models revealed a strong predictor signal in the spectrum range 2800–3000 cm<sup>-1</sup>, a characteristic region for stretching vibrations of lipid hydrocarbon chains. Altered lipid metabolism has previously been generally implicated in cancer [35,36], and specifically in NMIBC [37]. Spectroscopic evidence of changes in the serum levels of biochemical molecules including lipids has also been reported [38].

To the best of our knowledge, this study is the first report on this change in the FTIR spectrum that ranges from 2800 to 3000 cm<sup>-1</sup>. Because of this explicit fingerprint, this spectroscopic technique can be used as a diagnostic tool for the early detection of recurrent NMIBC. Unlike other tools and traditional techniques in the literature that have used complex procedures, our study proposed a direct and simple approach to detect NMIBC patients' recurrence. Interestingly, only a few studies have been published addressing bladder cancer in terms of using fluids, such as urine and bladder washing water. When we compared and contrasted these studies to our FTIR technique, we found that only one of them dealt with bladder wash [21] and solely for the bladder cancer detection marker, not recurrent NMIBC. In contrast, their fingerprint FTIR area had a different specific range than what we have found here. Gok et al. listed FTIR with a broad range from 1500 to 800 cm<sup>-1</sup> as a fingerprinting tool for bladder cancer in general, where spectroscopic evidence of changes in serum levels of biochemical molecules, including lipids, has also been reported.

### 4. Materials and Methods

#### 4.1. Patients and Samples Collection

A total of 62 patients were recruited after Transurethral Resection of Bladder Tumor (TURBT) of non-muscle invasive bladder cancer (NMIBC) and included in this prospective study. All patients were subjected to diagnosis and treatment at the National Cancer Institute (NCI), Cairo, Egypt.

Urine samples for cytology were obtained and abdominal ultrasound and cystoscopy were performed at three-month-intervals after Transurethral Resection of Bladder Tumor (TURBT) according to European Association of Urology (EAU) guidelines [4]. Cystoscopy was considered as the standard method for recurrences diagnosis. In the case of positive urine cytology, CT program and random bladder biopsies were performed. The voided urine specimens were divided into two aliquots; one was prepared for cytopathological examination and the other was stored at -80 °C for the FTIR assay.

The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of the National Cancer Institute (NCI), approval number (2111-502-015). All participants provided a written informed consent which was signed by each patient.

#### 4.2. ATR-FTIR Measurement

A BRUKER ALPHA FT-IR spectrometer, equipped with a ZnSe crystal attenuated total refraction (ATR) accessory, was used for recording the infrared spectra. A deuterated triglycine sulfate (DTGS) detector was used for measurements. The background interferogram was recorded with a clean ZnSe surface. After shaking the urine container, 5  $\mu\text{L}$  of the sample was pipetted onto the ZnSe crystal surface. Water absorption hinders the appearance of many components in the spectrum; therefore, the sample was dried for 15 min using a gentle stream of  $\text{N}_2$  gas prior to data acquisition to remove excess water. All samples were measured by collecting and averaging 28 scans for a final resolution of  $4 \text{ cm}^{-1}$ .

#### 4.3. Spectral Data Pre-Processing

Each urine sample was loaded and acquired at least three times. Spectral data were baseline corrected using vendor software and exported as SPC files. The files were then imported into R statistical environment (v4.2) for analysis using hyperSpec package v0.100.0 [39]. Spectral derivative transformations and Savitzky–Golay filtering [26] were performed using Signal R package (v0.7-7) [40], with a polynomial order of 3 and a window size ranging from 5 to 13. Raw, first or second order derivative transformations were performed as part of hyperparameter optimization. Transformed spectra were either not normalized, normalized to Amide-I band maximum in the  $1500\text{--}1700 \text{ cm}^{-1}$  range, or to Urea band maximum in the  $1400\text{--}1500 \text{ cm}^{-1}$ . Outliers were identified and removed through the *pcout* method from *mvoutlier* R package (v2.1.1). Wavenumber ranges  $700\text{--}900 \text{ cm}^{-1}$ ,  $700\text{--}1800 \text{ cm}^{-1}$ , and  $2800\text{--}3000 \text{ cm}^{-1}$  were extracted where relevant.

#### 4.4. Machine Learning

We relied on the extensive functionality implemented in caret R package (v6.0-92) to perform dataset partitioning, model training, and testing. A comprehensive guide on the implemented models and configurable parameters is available in Kuhn [41]. Following outlier removal, processed spectra were randomly split into training and testing sets in a 70:30 ratio. Replicate spectra from the sample urine samples were kept together to prevent data leaking. Thirty-five machine learning models were evaluated as detailed in Table S3. Two-class (free vs recurrence) variable was used for training and prediction. Model training was performed on the dataset using a 10-fold repeated cross-validation. To account for replicate spectra, we used a stratified K-fold sampling strategy, ensuring that replicates were always contained together in single fold. To address the class imbalance, ROC measure was used as a cost function instead of accuracy. Additionally, synthetic minority over-sampling technique (SMOTE) [27] was used to create new data points for the minority class. Trained models were then evaluated on the holdout testing set using F1 and ROC scores where appropriate.

### 5. Conclusions

This study presented a rapid and easy spectroscopic assay to detect NMIBC recurrence from urine using FTIR and sophisticated machine learning modeling. We examined the performance of 35 machine learning models using a training set (70%) and a held-out test set (30%). With 86% sensitivity and 77% specificity, a regularized random forest model achieved an area under the receiver operating characteristic (AUROC) of 0.92. This technology can be further utilized to describe the transition to the defined application of ATR-FTIR spectroscopy using urine samples for the detection of recurrence in NMIBC patients and the subsequent impact on clinical sectors. This can be done by simply using the spectroscopy of

urine instead of undergoing the complications associated with using cystoscopy, which is usually uncomfortable and sometimes painful. Therefore, this technological advancement leads to a prospective clinical validation study that is conducted in the population that is the focus of this article, which presents exploratory findings that confirm the fingerprint correlation with NMIBC. In addition, this finding can be used as an alternative diagnostic tool to detect NMIB cancer at an earlier stage, or it might completely replace cystoscopy with the use of a urine test in the clinic. This allows for early intervention and appears as a step forward in both the advancement of technology, and the improvement of the clinical course of treatment for patients who suffer from this cancer or in the diagnosis of this disease. We hope to further improve the current assay in a larger cohort to assess its potential in NMIBC surveillance.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/molecules27248890/s1>, Table S1: Performance of hyperparameter combinations, Table S2: Model tuning of regularized random forests, Table S3: Machine learning models abbreviations.

**Author Contributions:** Conceptualization, A.I.E.-F., A.E.-T. and A.M.; methodology, A.I.E.-F., A.M.S. and A.M.; software, A.M.; validation, A.I.E.-F., D.M.M., A.M.S., A.A.E.L. and A.M.; formal analysis, A.I.E.-F. and A.M.; investigation, A.I.E.-F., H.K. and A.M.; resources, A.I.E.-F., A.M.S. and A.A.E.L.; data curation, A.I.E.-F., D.M.M., M.A.-e., A.M.S., A.A.E.L. and A.M.; writing—original draft preparation, A.I.E.-F., D.M.S., N.M.L., D.M.M., S.A.M., M.A.-e., A.M.S., A.A.E.L., H.K., A.E.-T. and A.M.; writing—review and editing, A.I.E.-F., D.M.S., N.M.L., D.M.M., S.A.M., M.A.-e., A.M.S., A.A.E.L., H.K., A.E.-T. and A.M.; visualization, A.I.E.-F., D.M.S. and A.M.; supervision, A.I.E.-F.; project administration, A.I.E.-F.; funding acquisition, A.I.E.-F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Rally Projects Fund, Suez Canal University, grant number 3/2020.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Raw and processed FTIR spectra are available upon request.

**Acknowledgments:** The authors would like to thank the Dean of IBPR institute, the Vice President for Post-graduate Studies and Research and the Scientific Review Committee of Rally Projects for their scientific input and logistic support in this project.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. David, D.J. The application of atomic absorption to chemical analysis. A review. *Analyst* **1960**, *85*, 779–791. [[CrossRef](#)]
2. Cassell, A.; Yunusa, B.; Jalloh, M.; Mbodji, M.M.; Diallo, A.; Ndoye, M.; Diallo, Y.; Labou, I.; Niang, L.; Gueye, S.M. Non-muscle invasive bladder cancer: A review of the current trend in Africa. *World J. Oncol.* **2019**, *10*, 123–131. [[CrossRef](#)]
3. Oosterlinck, W.; Lobel, B.; Jakse, G.; Malmström, P.-U.; Stöckle, M.; Sternberg, C.; EAU Working Group on Oncological Urology. Guidelines on bladder cancer. *Eur. Urol.* **2002**, *41*, 105–112. [[CrossRef](#)]
4. Babjuk, M.; Böhle, A.; Burger, M.; Capoun, O.; Cohen, D.; Compérat, E.M.; Hernández, V.; Kaasinen, E.; Palou, J.; Rouprêt, M.; et al. EAU Guidelines on Non-Muscle-invasive Urothelial Carcinoma of the Bladder: Update 2016. *Eur. Urol.* **2017**, *71*, 447–461. [[CrossRef](#)]
5. Almallah, Y.; Rennie, C.; Stone, J.; Lancashire, M. Urinary tract infection and patient satisfaction after flexible cystoscopy and urodynamic evaluation. *Urology* **2000**, *56*, 37–39. [[CrossRef](#)]
6. Van Der Aa, M.N.; Steyerberg, E.W.; Sen, E.F.; Zwarthoff, E.C.; Kirkels, W.J.; Van Der Kwast, T.H.; Essink-Bot, M.L. Patients' perceived burden of cystoscopic and urinary surveillance of bladder cancer: A randomized comparison. *BJU Int.* **2008**, *101*, 1106–1110. [[CrossRef](#)]
7. Lotan, Y.; Roehrborn, C.G. Cost-effectiveness of a modified care protocol substituting bladder tumor markers for cystoscopy for the followup of patients with transitional cell carcinoma of the bladder: A decision analytical approach. *J. Urol.* **2002**, *167*, 75–79. [[CrossRef](#)]
8. Locatelli, F.; Canaud, B.; Eckardt, K.-U.; Stenvinkel, P.; Wanner, C.; Zoccali, C. The importance of diabetic nephropathy in current nephrological practice. *Nephrol. Dial. Transplant.* **2003**, *18*, 1716–1725. [[CrossRef](#)]

9. Sun, W.; Li, F.; Wu, S.; Wang, X.; Zheng, D.; Wang, J.; Gao, Y. Human urine proteome analysis by three separation approaches. *Proteomics* **2005**, *5*, 4994–5001. [CrossRef]
10. Shao, C.; Li, M.; Li, X.; Wei, L.; Zhu, L.; Yang, F.; Jia, L.; Mu, Y.; Wang, J.; Guo, Z. A tool for biomarker discovery in the urinary proteome: A manually curated human and animal urine protein biomarker database. *Mol. Cell. Proteom.* **2011**, *10*, 1–8. [CrossRef]
11. Shaw, R.A.; Low-Ying, S.; Man, A.; Liu, K.-Z.; Mansfield, C.; Rileg, C.B.; Vijarnsorn, M. Infrared spectroscopy of biofluids in clinical chemistry and medical diagnostics. In *Biomedical Vibrational Spectroscopy*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2008; pp. 79–103.
12. Sala, A.; Anderson, D.J.; Brennan, P.M.; Butler, H.J.; Cameron, J.M.; Jenkinson, M.D.; Rinaldi, C.; Theakstone, A.G.; Baker, M.J. Biofluid diagnostics by FTIR spectroscopy: A platform technology for cancer detection. *Cancer Lett.* **2020**, *477*, 122–130. [CrossRef]
13. Wijesinghe, H.G.; Hare, D.J.; Mohamed, A.; Shah, A.K.; Harris, P.N.; Hill, M.M. Detecting antimicrobial resistance in Escherichia coli using benchtop attenuated total reflectance-Fourier transform infrared spectroscopy and machine learning. *Analyst* **2021**, *146*, 6211–6219. [CrossRef]
14. Korb, E.; Bağcioğlu, M.; Garner-Spitzer, E.; Wiedermann, U.; Ehling-Schulz, M.; Schabussova, I. Machine learning-empowered FTIR spectroscopy serum analysis stratifies healthy, allergic, and SIT-treated mice and humans. *Biomolecules* **2020**, *10*, 1058. [CrossRef]
15. Sitnikova, V.E.; Kotkova, M.A.; Nosenko, T.N.; Kotkova, T.N.; Martynova, D.M.; Uspenskaya, M.V. Breast cancer detection by ATR-FTIR spectroscopy of blood serum and multivariate data-analysis. *Talanta* **2020**, *214*, 120857. [CrossRef]
16. Yang, X.; Ou, Q.; Yang, W.; Shi, Y.; Liu, G. Diagnosis of liver cancer by FTIR spectra of serum. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *263*, 120181. [CrossRef]
17. Gray, E.; Cameron, J.M.; Butler, H.J.; Jenkinson, M.D.; Hegarty, M.G.; Palmer, D.S.; Brennan, P.M.; Baker, M.J. Early economic evaluation to guide the development of a spectroscopic liquid biopsy for the detection of brain cancer. *Int. J. Technol. Assess. Health Care* **2021**, *37*, 1–6. [CrossRef]
18. Theakstone, A.G.; Brennan, P.M.; Jenkinson, M.D.; Mills, S.J.; Syed, K.; Rinaldi, C.; Xu, Y.; Goodacre, R.; Butler, H.J.; Palmer, D.S. Rapid spectroscopic liquid biopsy for the universal detection of brain tumours. *Cancers* **2021**, *13*, 3851. [CrossRef]
19. Mordechai, S.; Shufan, E.; Katz, B.P.; Salman, A. Early diagnosis of Alzheimer’s disease using infrared spectroscopy of isolated blood samples followed by multivariate analyses. *Analyst* **2017**, *142*, 1276–1284. [CrossRef]
20. Yap, X.-L.; Wood, B.; Ong, T.-A.; Lim, J.; Goh, B.-H.; Lee, W.-L. Detection of prostate cancer via IR spectroscopic analysis of urinary extracellular vesicles: A pilot study. *Membranes* **2021**, *11*, 591. [CrossRef]
21. Gok, S.; Aydin, O.Z.; Sural, Y.S.; Zorlu, F.; Bayol, U.; Sevencan, F. Bladder cancer diagnosis from bladder wash by Fourier transform infrared spectroscopy as a novel test for tumor recurrence. *J. Biophotonics* **2016**, *9*, 967–975. [CrossRef]
22. Barauna, V.G.; Singh, M.N.; Barbosa, L.L.; Marcarini, W.D.; Vassallo, P.F.; Mill, J.G.; Ribeiro-Rodrigues, R.; Campos, L.C.; Warnke, P.H.; Martin, F.L. Ultrarapid on-site detection of SARS-CoV-2 infection using simple ATR-FTIR spectroscopy and an analysis algorithm: High sensitivity and specificity. *Anal. Chem.* **2021**, *93*, 2950–2958. [CrossRef]
23. Paraskevaidi, M.; Martin-Hirsch, P.L.; Martin, F.L. ATR-FTIR spectroscopy tools for medical diagnosis and disease investigation. In *Nanotechnology Characterization Tools for Biosensing and Medical Diagnosis*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 163–211.
24. Finlayson, D.; Rinaldi, C.; Baker, M.J. Is infrared spectroscopy ready for the clinic? *Anal. Chem.* **2019**, *91*, 12117–12128. [CrossRef]
25. Grant-Peters, M.; Rich-Griffin, C.; Grant-Peters, J.E.; Cinque, G.; Dendrou, C.A. Photizo: An open-source library for cross-sample analysis of FTIR spectroscopy data. *Bioinformatics* **2022**, *38*, 3490–3492. [CrossRef]
26. Savitzky, A.; Golay, M.J. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]
27. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
28. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
29. Archer, K.J.; Kimes, R.V. Empirical characterization of random forest variable importance measures. *Comput. Stat. Data Anal.* **2008**, *52*, 2249–2260. [CrossRef]
30. Soria, F.; Krabbe, L.M.; Todenhöfer, T.; Dobruch, J.; Mitra, A.P.; Inman, B.A.; Gust, K.M.; Lotan, Y.; Shariat, S.F. Molecular markers in bladder cancer. *World J. Urol.* **2019**, *37*, 31–40. [CrossRef]
31. Yossepowitch, O.; Herr, H.W.; Donat, S.M. Use of urinary biomarkers for bladder cancer surveillance: Patient perspectives. *J. Urol.* **2007**, *177*, 1277–1282. [CrossRef]
32. Kavalieris, L.; O’Sullivan, P.; Frampton, C.; Guilford, P.; Darling, D.; Jacobson, E.; Suttie, J.; Raman, J.D.; Shariat, S.F.; Lotan, Y. Performance Characteristics of a Multigene Urine Biomarker Test for Monitoring for Recurrent Urothelial Carcinoma in a Multicenter Study. *J. Urol.* **2017**, *197*, 1419–1426. [CrossRef]
33. Pan, C.C. The value of molecular markers in classification and prediction of progression in non-muscle-invasive bladder cancer. *Transl. Androl. Urol.* **2018**, *7*, 736–739. [CrossRef]
34. Dyrskjøt, L.; Zieger, K.; Kruhøffer, M.; Thykjaer, T.; Jensen, J.L.; Primdahl, H.; Aziz, N.; Marcussen, N.; Møller, K.; Orntoft, T.F. A molecular signature in superficial bladder carcinoma predicts clinical outcome. *Clin. Cancer Res.* **2005**, *11*, 4029–4036. [CrossRef]
35. Menendez, J.A.; Lupu, R. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nat. Rev. Cancer* **2007**, *7*, 763–777. [CrossRef]
36. Molendijk, J.; Robinson, H.; Djuric, Z.; Hill, M.M. Lipid mechanisms in hallmarks of cancer. *Mol. Omics* **2020**, *16*, 6–18. [CrossRef]

37. Sahu, D.; Lotan, Y.; Wittmann, B.; Neri, B.; Hansel, D.E. Metabolomics analysis reveals distinct profiles of nonmuscle-invasive and muscle-invasive bladder cancer. *Cancer Med.* **2017**, *6*, 2106–2120. [[CrossRef](#)]
38. Chen, S.; Zhu, S.; Cui, X.; Xu, W.; Kong, C.; Zhang, Z.; Qian, W. Identifying non-muscle-invasive and muscle-invasive bladder cancer based on blood serum surface-enhanced Raman spectroscopy. *Biomed. Opt. Express* **2019**, *10*, 3533–3544. [[CrossRef](#)]
39. Beleites, C. *Package ‘hyperSpec version 0.100.0.’*; R Foundation for Statistical Computing: Vienna, Austria, 2021.
40. Ligges, U.; Short, T.; Kienzle, P.; Schnackenberg, S.; Billinghamurst, D.; Borchers, H.-W.; Carezia, A.; Dupuis, P.; Eaton, J.W.; Farhi, E. *Package ‘signal version 0.7-7.’*; R Foundation for Statistical Computing: Vienna, Austria, 2021.
41. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **2008**, *28*, 1–26. [[CrossRef](#)]