




Article

Classification of Healthy and Cancer Colon Cells Grown on Glass Coverslip by Means of Fourier Transform Infrared Spectroscopy and Multivariate Methods

Giuseppe Perna , Vito Capozzi  and Maria Lasalvia * 

Dipartimento di Medicina Clinica e Sperimentale, Università di Foggia, 71122 Foggia, Italy; giuseppe.perna@unifg.it (G.P.); vito.capozzi@unifg.it (V.C.)

* Correspondence: maria.lasalvia@unifg.it

Abstract: For several years, Fourier transform infrared (FTIR) microspectroscopy has been proving to be very promising for use in cytological diagnostics because of its capability of providing rapid and label-free biochemical information about cell samples. The adoption of FTIR as a clinical tool has been slowed because of the poor compatibility with cells deposited on glass slides, commonly used in clinical practice, because of the absorption of IR radiation by glassy materials in the 1000–1800 cm^{-1} spectral range. However, the possibility of also obtaining diagnostic information from the IR absorption spectra in the 2700–3700 cm^{-1} range (including few peaks related to vibrational modes in cell lipids and proteins) has recently emerged. In this work, we investigate the use of the FTIR technique in the 2700–3700 cm^{-1} range for diagnostic purposes about human colon cells grown on glass coverslips. In fact, using the principal components analysis (PCA) technique, we are able to discriminate FTIR spectra of healthy cells from those of cancerous ones, mainly due to the larger relative lipid content in the former compared to the latter. In addition, principal component analysis-linear discriminate analysis (PCA-LDA) and partial least square-discriminant analysis (PLS-DA) were used to build classification models for unknown FTIR spectra with optimal accuracy. These results support the promotion of the translation of the FTIR technique as a complementary diagnostic tool in cytological routine practice.

Keywords: colon cancer; FTIR spectroscopy; PCA; PCA-LDA; PLS-DA



Citation: Perna, G.; Capozzi, V.; Lasalvia, M. Classification of Healthy and Cancer Colon Cells Grown on Glass Coverslip by Means of Fourier Transform Infrared Spectroscopy and Multivariate Methods. *Photonics* **2023**, *10*, 481. <https://doi.org/10.3390/photonics10040481>

Received: 30 March 2023

Revised: 17 April 2023

Accepted: 20 April 2023

Published: 21 April 2023



Copyright: © 2023 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/>).

1. Introduction

Cytology tests concern the microscope examination of stained cells from bodily tissues or fluids to look for eventual abnormalities of morphologic characteristics, which allow a diagnosis to be made. Cytology tests have made considerable progress in the early diagnosis of some cancer pathologies; probably the most important example is the Papanicolaou test [1], used as a screening for the early detection of cervical cancer, which has significantly contributed to reducing mortality for this type of cancer. However, sometimes the cytological methods could be time-consuming [2], they could have a relatively low specificity [3] and they might be partly subjective, according to the skill and expertise of the cytopathologist performing the analysis [4]. Hence, they could result in decreasing the reliability of the diagnosis. Therefore, it is worth considering cytological methods that provide results with relatively high accuracy, independently on the operator's expertise, in order to improve the reliability of the diagnosis.

Fourier transform infrared (FTIR) microspectroscopy offers the right characteristics to be proposed as an additional cells analysis tool compared to the “gold standard” techniques, which are based on subjective clinical evaluation of morphological changes in cells. In fact, it provides label-free biochemical information about the cell samples according to the spectral energies absorbed by the functional groups inside several cell components (nucleic acids, lipids, proteins, etc.) [5–8]. Therefore, changes of the relative amount and

structure of cell components induced by the onset of pathology can be detected by FTIR technique. Recently, FTIR microspectroscopy has been used to distinguish cancerous and normal samples for different cell lines, including oral [9], liver [10], breast [11] and skin [8].

To fully exploit the information provided by the mid-infrared signal in the FTIR technique, it is required that the cellular samples are adherent to suitable substrates which do not absorb the IR radiation. Commonly used FTIR substrates are those made of CaF_2 and BaF_2 materials, which are transparent to IR radiation, or substrates that reflect the IR radiation, such as those consisting of thick glass slides over which a thin Ag/SnO_2 coating has been deposited. However, the use of such substrates is not widespread in clinical diagnostics due to their larger costs compared to those of standard glass slides and coverslips. The use of cheap standard glass substrates in the FTIR technique limits the analysis of IR signal to the so-called “high wavenumbers spectral range” ($2700\text{--}3700\text{ cm}^{-1}$) because of the absorption of glass-based materials below 2500 cm^{-1} . The spectral signals obtained from cell samples in the high wavenumber range are mainly based on the absorption due to stretching vibrational modes of CH_2 and CH_3 functional groups of lipid and protein cell components. Nonetheless, recent works demonstrated that the spectral information provided by the high wavenumber range is able to differentiate the spectra measured for different cellular samples [12,13]. In such differentiation, it is important to analyse the data by means of multivariate statistical techniques, primarily to evaluate the possibility of discrimination, by exploratory techniques such as principal component analysis (PCA) [14], and then also to estimate the accuracy of classifying FTIR spectra from cells of different classes by using classification techniques such as principal component analysis-linear discriminate analysis (PCA-LDA) [15] and partial least square-discriminant analysis (PLS-DA) [16].

In a previous paper, we successfully discriminated three different types of breast cells according to their different FTIR spectra [13]. In particular, we analysed the FTIR data by PCA method to highlight the possibility of FTIR technique to discriminate objects belonging to different classes, but we did not use any classification method to evaluate how accurately spectra belonging to different classes could be classified. In this work, we demonstrate that similar methodology can be used to discriminate healthy colon cells from cancer ones. In particular, we investigate cells from the FHC cell line as a model of healthy cells and cells from the CaCo-2 cell line as a model of cancerous cells. Both cell types were grown on glass coverslip substrates. In particular, although the FTIR spectra in the high wavenumber range of healthy colon cells are characterised by spectral features slightly different from those of cancerous colon cells, PCA is able to discriminate the spectra of the two cell samples according to the values of PC1 score. In addition, PCA-LDA and PLS-DA techniques were used to build classification models to correctly assign unknown spectra to the proper class. The achieved results point out that both the classification models were able to distinguish the different spectra types with high accuracy.

These results, obtained with cheap conventional glass coverslips, although they involve only two cell lines, represent a positive feasibility study for supporting the use of the FTIR spectroscopy, combined with multivariate analysis techniques, as a complementary diagnostic tool in cytology.

2. Materials and Methods

2.1. Cell Culture and Preparation

Foetal Human Colon (FHC) cells were purchased from ATCC (CRL-1831). It is a type of human cell line derived from normal foetal colon tissue. Therefore, it can be used to model healthy colon cells. The base medium for this cell line is DMEM F12 to which we have added 10 mM Hepes, 10 ng/mL cholera toxin, 5 µg/mL insulin, 5 µg/mL transferrin, 100 ng/mL hydrocortisone, 20 ng/mL EGF and Fetal Bovine Serum 10% final concentration.

Human colorectal adenocarcinoma (CaCo-2) cells were obtained from ATCC (Manassas, VA, USA). They consist of human colorectal adenocarcinoma epithelial cell lines isolated from colon tissue of a 72-year-old male. Therefore, they can be used to model

cancerous colon cells. CaCo-2 cells were grown in Dulbecco's Modified Eagle's medium (DMEM), supplemented with 4 mmol dm⁻³ L-glutamine, 1% penicillin/streptomycin, 10% fetal bovine serum (FBS) and 1% non-essential amino acids (NEAA) at 37 °C and 5% CO₂.

The cells to be measured by FTIR microspectroscopy were cultured on a glass coverslip; after that, a proper poly-lysine coating was deposited on the glass surface. The cultured cells, both FHC and CaCo-2, were fixed in paraformaldehyde 3.7% and stored inside a desiccator until FTIR spectra acquisition.

2.2. FTIR Measurements and Analysis

FTIR spectra were measured in the transmission mode by using a FTIR Microscope HYPERION 2000 (Bruker Optik GmbH, Ettlingen, Germany) where the IR radiation beam comes from a Vertex 70 Bruker interferometer (Bruker Optik GmbH). The IR signal was detected by a mercury cadmium telluride (MCT) device, cooled at liquid N₂ temperature. Each spectrum corresponds to the average signal of 64 scans in the 2700–3700 cm⁻¹ range, with a resolution of 4 cm⁻¹. This spectral range contains representative information about lipid and protein cell components. The IR radiation was focused with a 15X objective onto a few cells included in the sampling area of about 80 μm x 80 μm size. The background signal was measured from a coverslip area without any cell. The number of 50 cells were measured for each of the two cell types. The spectra were collected by means of the Opus 6.5 software (Bruker Optik). The spectra were normalised using standard normal variate (SNV) method, which decreases the spectrum baseline shifts related to scattering effects [17] and minimises the contribution of the absorption from cells having different thickness.

Exploratory data analysis was performed by means of PCA. Firstly, PCA was applied to the full dataset of 100 spectra in order to identify and remove any outlier spectra. In particular, outliers were detected by displaying the Hotelling's T² ellipse in the PC2 vs. PC1 score plot (Figure S1) because the first two PCs contain most of the variance (83%) of the dataset. After removing the outlier spectra, each of the two different groups of spectra were separated into a calibration set, comprising 25 cells of each cell type, and a test set, including the remaining cells of the spectral dataset in order to evaluate the ability of FTIR spectroscopy to properly classify the two types of spectra. The spectra of the calibration group were randomly selected by a random number generator. Classification models for discriminating samples from the two classes of each group were developed by using the PCA-LDA and PLS-DA techniques for the calibration set, whereas the test set is used to estimate the model classification accuracy. PCA, PCA-LDA and PLS-DA techniques were performed with the Unscrambler X CAMO software (version 10.4), whereas t-test analysis was performed by SigmaPlot software (version 12.5, Systat Software, San Jose, CA, USA).

3. Results and Discussion

The mean spectra of the normalised ones measured from the calibration set of healthy and cancerous cells are shown in Figure 1 as continuous lines, whereas the standard deviation signals are reported as dashed lines. Both types of spectra are similar to those reported for colon cells [18] and tissues [19]. They are characterised by several absorption features which can be attributed according to the published literature about cell lines and tissues [20]. The most resolved peaks have been labelled in Figure 1. In particular, the most intense and broad band in both spectra, located at about 3300 cm⁻¹, corresponds to the amide A band, which is related to the absorption from N-H stretching mode of proteins and nucleic acids. The other FTIR peaks in the investigated spectral range are due to the radiation absorption from functional groups related to lipid and protein components. Specifically, the two well resolved peaks at about 2850 cm⁻¹ and 2925 cm⁻¹ are due to symmetric and asymmetric stretching modes of the CH₂ groups of lipid components, respectively. Moreover, the peak at about 2960 cm⁻¹ is related to the asymmetric stretching mode of the CH₃ groups of both protein and lipid components. Besides the labelled peaks, other spectral features are evident in Figure 1, such as the broad shoulders in the

3000–3100 cm^{-1} and 3400–3550 cm^{-1} spectral ranges, which could be attributed to amide B and OH stretching modes, respectively [20].

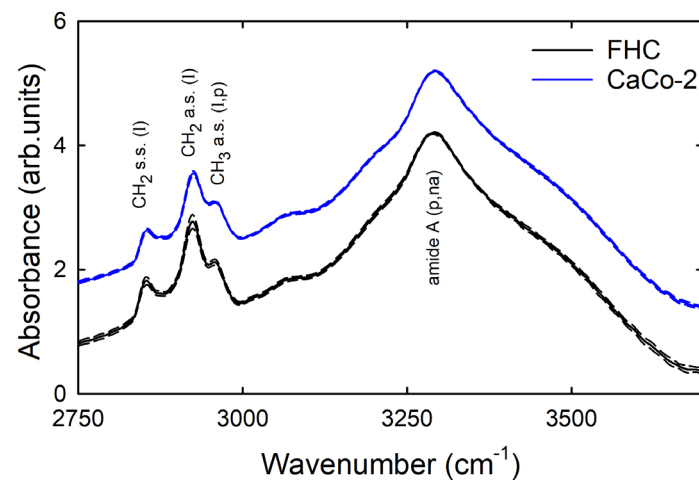


Figure 1. Mean FTIR spectra of healthy FHC (continuous black line) and cancerous CaCo-2 (continuous blue line) cells after SNV normalisation. Standard deviation spectra are also reported as dashed lines. The assignment of vibrational modes is also reported (na: nucleic acids; p: proteins; l: lipids; s.s.: symmetrical stretching; a.s.: asymmetrical stretching). The spectra are vertically shifted for clarity purpose.

The two mean spectra in Figure 1 are very similar to each other. However, it is interesting to study whether small and not clearly visible differences in the relative intensities of the spectral signals between the spectra of the two types of cells in the calibration set are able to discriminate the spectra measured on healthy cells from those measured on cancer cells. Therefore, we first analysed data from the calibration set by PCA, using a full cross-validation method. In fact, PCA score plots can show eventual similarities and differences between spectra from two classes, and PCA loading plots are able to highlight the spectral features to which the differences could be attributed [15].

Figure 2a shows the score plot of PC2 vs. PC1 for the calibration set. The first 3 principal components included around 90% of all the spectral variation found in the whole dataset. It is clearly visible that the group of healthy cells (blue circles) are well discriminated by the group of cancer ones (red circles), according to the PC1 score values. In particular, the cancerous cells are characterised by negative values of PC1 scores, whereas healthy cells have mainly positive values of PC1 scores. The mean and standard deviation values of the two distributions of PC1 values are reported at the bottom of Figure 2a. The t-test analysis clearly points out that the distribution of PC1 score values for healthy cells is significantly different from that of the corresponding values for cancer cells. Such discrimination properties of the PCA technique suggest the possibility of classification of spectra of unknown nature by means of proper multivariate techniques. Before addressing this issue, it is interesting to investigate which absorption peak in the measured range is mainly responsible for the discrimination between the two types of spectra. This can be achieved by analysing the loading plot of PC1 and comparing it to the difference signal of the average spectra shown in Figure 1. The loading 1 and the difference plots are displayed in Figure 2b; they are very similar to each other, and the most resolved and intense peaks at about 2850 cm^{-1} and 2920 cm^{-1} suggest that the relative content of lipid components is mainly responsible for the discrimination possibilities between the two types of cells. In particular, as positive values of spectral peaks in the loading plot indicate that samples with positive scores have a larger relative content of the chemical components corresponding to those peaks, Figure 2 suggests that the relative lipid content is larger in healthy cells than in cancerous ones.

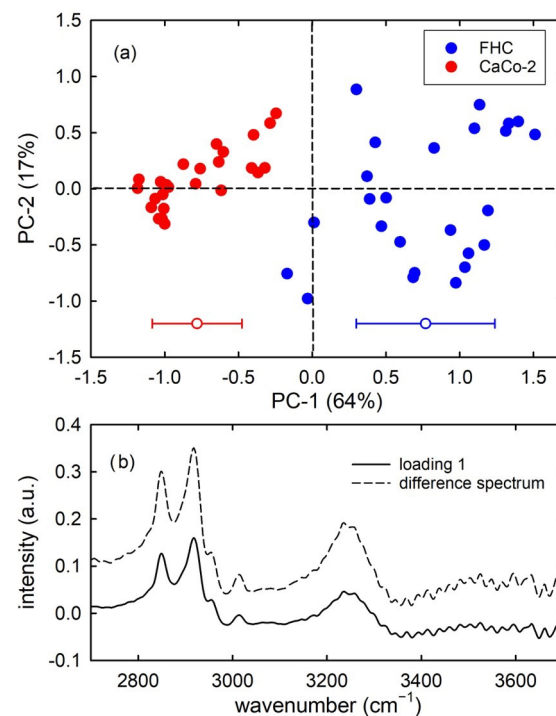


Figure 2. PC2 vs. PC1 score plot (a) for the healthy FHC cells (blue dots) and cancerous CaCo-2 cells (red dots) FTIR spectra. The mean and standard deviation of the PC1 score values are shown on the bottom side of (a). Results of the t-test point out that the two distributions of score values are significantly different ($p < 0.001$). Difference of mean spectrum values (healthy-cancer signal) and loading 1 spectrum are reported in (b).

These results are in good agreement with those reported by L. Dong et al. who found larger intensities of lipid related peaks in FTIR spectra of normal colon tissues, with respect to cancerous ones [21]. Additionally, B. Brozek-Pluska et al. deduced from Raman spectra that the lipid/protein ratio value is larger in noncancerous colon tissue than in a cancerous one [22,23]. Instead, C. Scalfi-Happ et al. investigated the fraction of lipid bodies inside colon cells by means of Raman spectroscopy and they found that such a fraction is significantly larger in malignant than in non-malignant cells [24]. Overall, these results highlight that the relative lipid content in colon cells is significantly different for malignant and non-malignant cells, and the corresponding spectral peaks could be used as a spectral biomarker for malignancy. Therefore, classification techniques can be applied to the test sets of our colon cells to evaluate the discrimination performance.

First of all, we considered PCA-LDA as a classification technique and evaluated its capability to correctly attribute unknown spectra to the proper class. In particular, the scores of the first three PCs of the spectra from the calibration set were used as input parameters to develop an LDA model that will be successively used to diagnose the class of unknown samples (spectra from test set). The obtained PCA-LDA model correctly classified all 50 spectra, as can be deduced from Figure 3, where the blue filled circles correspond to the spectra from healthy cells and the red filled circles to those from cancerous ones. Briefly, a discriminant score value is calculated for each object (corresponding to a cell spectrum) relating to its membership to both classes and the object is attributed to the class for which its score value is the largest one [15]. As the values of discriminant scores are negative, in Figure 3, each object is attributed to the class for which the score value is close to zero.

After this step of building the PCA-LDA model, we estimated its classification performance by using spectra from the test set. In fact, this is a crucial point to evaluate the possibility of translating the investigated method to the clinical setting for diagnostic purposes. The projection of the objects from the test set on the classification plot is displayed in Figure 3 by means of blue and red hollow circles corresponding to the spectra classified

as healthy and cancer, respectively. All the test set spectra were assigned by the PCA-LDA model to the proper class, as can be deduced from the proximity of the representative points of the test set to the corresponding representative points of the calibration set. Therefore, the PCA-LDA model classified the test set spectra with excellent accuracy.

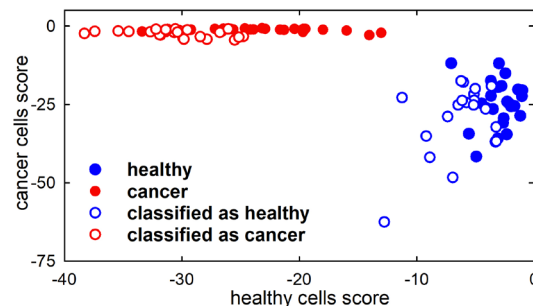


Figure 3. Discrimination plot from linear discriminant analysis model, developed using three principal components. The scatter plot shows the discrimination of healthy FHC cell spectra (filled blue circles) and cancerous CaCo-2 cell spectra (filled red circles). The projections of the test samples on the PCA-LDA model are shown as hollow circles.

Other authors correctly classified healthy colon and cancer samples by means of vibrational spectroscopy and the PCA-LDA method. More than 20 years ago, A. Salman et al. performed an FTIR microspectroscopy investigation of thin colon tissue specimens in parallel with normal histopathological analysis; preliminary results of LDA showed that normal and malignant tissues could be identified with about 89% accuracy [25]. Also M. Khanmohammadi et al., by using PCA-LDA classified FTIR spectra from colon specimen into cancer and normal tissue classes with an accuracy of 95.8% and sensitivity and specificity was 100% and 93.1%, respectively [26]. We remark that these works are effective for highlighting the possibility of discriminating at the histological level samples of healthy and cancerous colons using PCA-LDA of FTIR spectra, but the aim of our work mainly concerns the discrimination at a cytological level. As for colon single cells, M. Liu et al. used Raman spectroscopy to examine the colon cancer cells with a single base mutation in KRAS gene segment; they successfully discriminated different cells with KRAS wild type from cells with KRAS mutant type by PCA-LDA with 81% accuracy [27]. The PCA-LDA method has also been used to discriminate normal and cancerous vibrational spectra of samples different from colon specimen, such as oral tissue [28,29], breast [30] and brain [31] cells. In addition to vibrational spectra, histopathological images from two classes of colon cancer were also recently discriminated by the LDA method with about 90% accuracy [32]. Therefore, our results about the performance of the PCA-LDA method applied to FTIR spectra from different colon cells correspond to those reported by other authors.

Additionally, a PLS model was developed for FTIR spectra of the calibration set. It clearly separates healthy and cancerous samples, as visible in the score plot of Figure 4a (filled circles), mainly for factor 1 values. The loading 1 plot provides information about which variables are mainly responsible for the above separation; it can be deduced in Figure 4b that the most important variables in the separation of factor 1 score values were those corresponding to the peaks at about 2850 cm^{-1} and 2920 cm^{-1} and, to a lesser extent, at about 3250 cm^{-1} . The loading plot in Figure 4b is similar to that of the difference between the average spectra of healthy and cancerous cell spectra reported in Figure 2b, as expected, because the spectral peaks centred at the above wavenumber values are mainly responsible for the difference between the two average FTIR spectra.

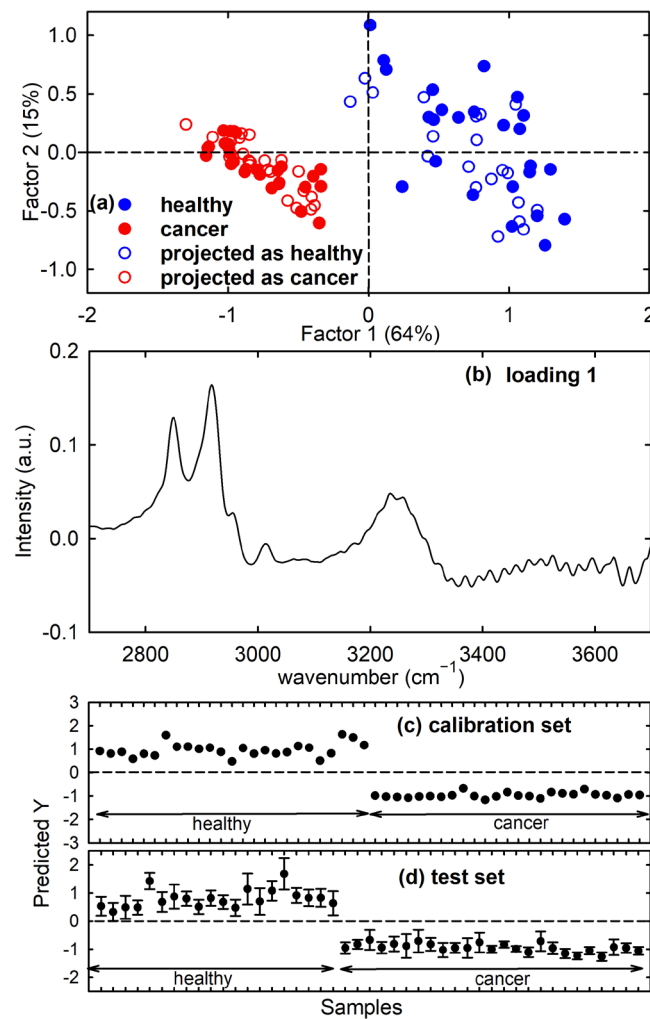


Figure 4. Score plot (factor 2 vs. factor 1) of the developed PLS-DA model showing the calibration (filled circles) and projected test set (hollow circles) samples for the healthy FHC cell (blue circles) and cancerous CaCo-2 cell (red circles) spectra in (a). The root mean square error for cross-validation (RMSECV) with three factors is 0.30. Loading 1 plot for the PLS model in (b). Predicted Y values with respect to the reference values (1: healthy FHC cells; −1: cancer CaCo-2 cells) after PLS regression for the spectra of the calibration set and predicted Y values, with corresponding deviations from the PLS-DA model for the spectra of the test set, are reported in (c) and (d), respectively, with black dots.

In the PLS method, a Y matrix consisting of n rows (where n is the number of samples) is introduced and each row is a categorical variable specifying the type of sample. For the investigated calibration set, n = 50 and the reference values of the categorical variable are encoded as −1 and +1, corresponding to cancer and healthy type, respectively, and zero value is the class boundary. After PLS regression, predicted Y values are obtained for the samples of the calibration set; these values are close to the reference values, as displayed in Figure 4c. The root mean square error from cross-validation (RMSECV) with three factors is 0.30. Indeed, we considered the first three factors of the PLS model because, with such three components, the explained variance achieved a value larger than 90% of the total variance. The performance of the developed PLS model in the attribution of unknown spectra has been estimated by using the spectra from the test set and including three factors in the prediction step. All the unknown spectra were correctly assigned to the proper class, as visible in Figure 4d, where the predicted values with deviation (uncertainty) are reported. Indeed, the predicted Y values of the samples of the test set are similar to the corresponding Y values of the samples used to make the calibration model in Figure 4c. The estimated

deviations around the prediction value do not include the zero value; hence, it can be deduced that any sample belongs to any of the two classes and no misclassification was observed. Therefore, all the test set spectra were assigned to the right class, thus producing maximum values of the performance parameters. These results can also be visualised by showing the projected scores of the test samples on the factor 2 vs. factor 1 score plot, shown as hollow circles in Figure 4a.

Recently, B. Brozek-Pluska evaluated the biochemical characteristics of normal and cancerous human cell lines by means of Raman spectroscopy; she found that PLS-DA yielded effective cell classification based on vibrational features, reporting sensitivity and specificity values of 100% [33]. Additionally, W. Liu et al. reported large sensitivity and specificity values (98.9% and 97%, respectively) by a PLS-DA model for the discrimination of Raman spectra obtained by subtracting cytoplasm from nucleus spectra in cells from normal and cancerous colon human tissue [34]. Furthermore, M.S. Bergholt et al. PLS-DA modelled *in vivo* tissue Raman spectra and they yielded a diagnostic accuracy of 88.8% (sensitivity 93.9% and specificity 88.3%) for colorectal cancer detection [35]. While many works have been published about Raman spectra of human colorectal cells and tissue samples analysed by PLS-DA technique, we have not found any works related to the analysis by PLS-DA of FTIR spectra of cells and tissues from the colon. However, such PLS-DA analyses of FTIR spectra have been reported for normal and cancerous breast cells [36], as well as for different types of melanoma cell lines [8]; in both cases, large percentages of correct prediction for the status of the cells have been obtained.

Therefore, both PCA-LDA and PLS-DA techniques successfully discriminated FTIR spectra of healthy colon cells from cancerous ones by only considering spectral signals in the high wavenumber range. It could be argued that the obtained result might depend on the particular, random choice of the calibration and test sets. Hence, we performed two further random selections of the calibration and of the test set within the initial dataset of FTIR spectra in order to evaluate possible variations in the classification accuracy. The analysis performed did not show drastic differences with respect to what was previously discussed (Figures S2–S4).

4. Conclusions

The obtained results point out that the FTIR spectra measured in the high wavenumber range and the PCA analysis of such spectra are able to discriminate healthy colon cells from cancerous ones, according to biochemical differences mainly related to the relative content of lipid components, which is larger in former cells than in the latter ones. FTIR microspectroscopy provides objective biochemical information about cytological samples, whereas the optical inspection of cell samples with a microscope is based on subjective evaluation of morphological features by a cytologist. Hence, the FTIR technique offers the proper characteristics to be proposed as a suitable cell analysis tool in cytological diagnostics practice because reliable information can be obtained from cytological samples deposited on cheap glass slides commonly used in medical practice.

In order to estimate the prediction capability of this vibrational technique, we first developed PCA-LDA and PLS-DA algorithms by using a subset of each dataset as a calibration set; then, the performance of these models was estimated according to the classifications obtained for a test set, which did not include any of the spectra used for the calibration set. Our results showed that both models have good and comparable prediction capability for a triplicate random choice of calibration and test sets extracted from the original dataset.

Despite the excellent achieved results, our investigation has several limitations that must be overcome in order to allow a possible adoption of multivariate classification of FTIR spectra for diagnostic purposes. First of all, the number of measured cells in the dataset is relatively small and it should be increased to obtain a more reliable accuracy of classification than what we reported in this study. Hence, our results about cell lines should be considered as preliminary data. More interestingly, they should be confirmed by using

different types of both healthy and cancerous colon cells, as well as ex vivo cells extracted from patients and tissue samples from biopsies, in order to evaluate the effectiveness of using this vibrational and multivariate method in cytological diagnosis. Nonetheless, the obtained results are promising because they involve cellular samples deposited on very cheap glass slides, which are widely used in biomedical practice.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/photonics10040481/s1>, Figure S1: PCA of the original dataset to detect outliers; Figure S2: PCA of two randomly chosen spectra to constitute two different calibration sets; Figure S3: PCA-LDA analysis of two randomly chosen spectra to constitute two different calibration and test sets; Figure S4: PLS-DA analysis of two randomly chosen spectra to constitute two different calibration and test sets.

Author Contributions: Conceptualization, V.C. and M.L.; methodology, G.P.; formal analysis, G.P.; investigation, G.P. and M.L.; data curation, G.P.; writing—original draft preparation, M.L.; writing—review and editing, V.C.; supervision, M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data obtained in this work will be available from the corresponding author, M.L., upon reasonable request.

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

References

1. Swailes, A.L.; Hossler, C.E.; Kesterson, J.P. Pathway to the Papanicolaou smear: The development of cervical cytology in twentieth-century America and implications in the present day. *Gynecol. Oncol.* **2019**, *154*, 3–7. [CrossRef] [PubMed]
2. Thakur, N.; Alam, M.R.; Abdul-Ghaffar, J.; Chong, Y. Recent application of artificial intelligence in non-gynecological cancer cytopathology: A systematic review. *Cancers* **2022**, *14*, 3529. [CrossRef] [PubMed]
3. O’Flynn, H.; Ryan, N.A.J.; Narine, N.; Shelton, D.; Rana, D.; Crosbie, E.J. Diagnostic accuracy of cytology for the detection of endometrial cancer in urine and vaginal samples. *Nat. Commun.* **2021**, *12*, 952. [CrossRef] [PubMed]
4. Wright, T.C.; Stoler, M.H., Jr.; Behrens, C.M.; Sharma, A.; Sharma, K.; Apple, R. Interlaboratory variation in the performance of liquid-based cytology: Insights from the ATHENA trial. *Int. J. Cancer* **2014**, *134*, 1835–1843. [CrossRef]
5. Beć, K.B.; Grabska, J.; Huck, C.W. Biomolecular and bioanalytical applications of infrared spectroscopy—A review. *Anal. Chim. Acta.* **2020**, *1133*, 150–177. [CrossRef]
6. Baker, M.J.; Trevisan, J.; Bassan, P.; Bhargava, R.; Butler, H.J.; Dorling, K.M.; Fielden, P.R.; Fogarty, S.W.; Fullwood, N.J.; Heys, K.A.; et al. Using Fourier transform IR spectroscopy to analyze biological materials. *Nat. Protoc.* **2014**, *9*, 1771–1791. [CrossRef]
7. Errico, S.; Moggio, M.; Diano, N.; Portaccio, M.; Lepore, M. Different experimental approaches for Fourier-transform infrared spectroscopy applications in biology and biotechnology: A selected choice of representative results. *Biotechnol. Appl. Biochem.* **2022**, 1–25. [CrossRef]
8. Shakya, B.R.; Teppo, H.R.; Rieppo, L. Discrimination of melanoma cell lines with Fourier Transform Infrared (FTIR) spectroscopy. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *254*, 119665. [CrossRef]
9. Byrne, H.J.; Behl, I.; Calado, G.; Ibrahim, O.; Toner, M.; Galvin, S.; Healy, C.M.; Flint, S.; Lyng, F.M. Biomedical applications of vibrational spectroscopy: Oral cancer diagnostics. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *252*, 119470. [CrossRef]
10. Sheng, D.; Xu, F.; Yu, Q.; Fang, T.; Xia, J.; Li, S.; Wang, X.A. study of structural differences between liver cancer cells and normal liver cells using FTIR spectroscopy. *J. Mol. Struct.* **2015**, *1099*, 18–23. [CrossRef]
11. Santos, I.P.; Martins, C.B.; Batista de Carvalho, L.A.E.; Marques, M.P.M.; Batista de Carvalho, A.L.M. Who’s who? Discrimination of human breast cancer cell lines by Raman and FTIR microspectroscopy. *Cancers* **2022**, *14*, 452. [CrossRef]
12. Rutter, A.V.; Crees, J.; Wright, H.; Raseta, M.; van Pittius, D.G.; Roach, P.; Sulé-Suso, J. Identification of a glass substrate to study cells using Fourier transform infrared spectroscopy: Are we closer to spectral pathology? *Appl. Spectrosc.* **2020**, *74*, 178–186. [CrossRef]
13. Lasalvia, M.; Capozzi, V.; Perna, G. Discrimination of different breast cell lines on glass substrate by means of Fourier transform infrared spectroscopy. *Sensors* **2021**, *21*, 6992. [CrossRef]
14. Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20150202. [CrossRef]
15. Varmuza, K.; Filzmoser, P. *Introduction to Multivariate Statistical Analysis in Chemometrics*; CRC Press: Boca Raton, FL, USA, 2009.

16. Lee, L.C.; Liong, C.Y.; Jemain, A.A. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: A review of contemporary practice strategies and knowledge gaps. *Analyst* **2018**, *143*, 3526–3539. [\[CrossRef\]](#)
17. Zeaiter, M.; Rutledge, D. Preprocessing methods. In *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*; Brown, S.D., Tauler, R., Walczak, B., Eds.; Elsevier: Amsterdam, The Netherlands, 2009; Volume 3, pp. 121–231.
18. Inan Genç, A.; Gök, S.; Banerjee, S.; Severcan, F. Valdecocixib recovers the lipid composition, order and dynamics in colon cancer cell lines independent of COX-2 expression: An ATR-FTIR spectroscopy study. *Appl. Spectrosc.* **2017**, *71*, 105–117. [\[CrossRef\]](#)
19. Song, C.L.; Kazarian, S.G. Micro ATR-FTIR spectroscopic imaging of colon biopsies with a large area Ge crystal. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2020**, *228*, 117695. [\[CrossRef\]](#)
20. Talari, A.C.S.; Martinez, M.A.G.; Movasaghi, Z.; Rehman, S.; Rehman, I.U. Advances in Fourier transform infrared (FTIR) spectroscopy of biological tissues. *Appl. Spectr.* **2017**, *52*, 456–506. [\[CrossRef\]](#)
21. Dong, L.; Sun, X.; Chao, Z.; Zhang, S.; Zheng, J.; Gurung, R.; Du, J.; Shi, J.; Xu, Y.; Zhang, Y.; et al. Evaluation of FTIR spectroscopy as diagnostic tool for colorectal cancer using spectral analysis. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2014**, *122*, 288–294. [\[CrossRef\]](#)
22. Brozek-Pluska, B.; Dzik, A.; Abramczyk, H. Virtual spectral histopathology of colon cancer-biomedical applications of Raman spectroscopy and imaging. *J. Mol. Liq.* **2020**, *303*, 112676. [\[CrossRef\]](#)
23. Brozek-Pluska, B.; Musial, J.; Kordek, R.; Abramczyk, H. Analysis of human colon by Raman spectroscopy and imaging-elucidation of biochemical changes in carcinogenesis. *Int. J. Mol. Sci.* **2019**, *20*, 3398. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Scalfi-Happ, C.; Udart, M.; Hauser, C.; Rück, A. Investigation of lipid bodies in a colon carcinoma cell line by confocal Raman microscopy. *Med. Laser Appl.* **2011**, *26*, 152–157. [\[CrossRef\]](#)
25. Salman, A.; Argov, S.; Ramesh, J.; Goldstein, J.; Sinelnikov, I.; Guterman, H.; Mordechai, S. FT-IR microscopic characterization of normal and malignant human colonic tissues. *Cell Mol. Biol.* **2001**, *47*, 159–166.
26. Khanmohammadi, M.; Bagheri Garmarudi, A.; Samani, S.; Ghasemi, K.; Ashuri, A. Application of linear discriminant analysis and Attenuated Total Reflectance Fourier Transform Infrared microspectroscopy for diagnosis of colon cancer. *Pathol Oncol Res.* **2011**, *17*, 435–441. [\[CrossRef\]](#)
27. Liu, M.; Liu, X.; Huang, Z.; Tang, X.; Lin, X.; Xu, Y.; Chen, G.; Kwok, H.F.; Lin, Y.; Feng, S. Rapid discrimination of colon cancer cells with single base mutation in KRAS gene segment using laser tweezers Raman spectroscopy. *J. Biophotonics* **2019**, *12*, e201800332. [\[CrossRef\]](#)
28. Jeng, M.J.; Sharma, M.; Sharma, L.; Chao, T.Y.; Huang, S.F.; Chang, L.B.; Wu, S.L.; Chow, L. Raman spectroscopy analysis for optical diagnosis of oral cancer detection. *J. Clin. Med.* **2019**, *8*, 1313. [\[CrossRef\]](#)
29. Sharma, M.; Jeng, M.J.; Young, C.K.; Huang, S.F.; Chang, L.B. Developing an algorithm for discriminating oral cancerous and normal tissues using Raman spectroscopy. *J. Pers. Med.* **2021**, *11*, 1165. [\[CrossRef\]](#)
30. Lasalvia, M.; Capozzi, V.; Perna, G. A comparison of PCA-LDA and PLS-DA techniques for classification of vibrational spectra. *Appl. Sci.* **2022**, *12*, 5345. [\[CrossRef\]](#)
31. Iturrioz-Rodríguez, N.; De Pasquale, D.; Fiaschi, P.; Ciofani, G. Discrimination of glioma patient-derived cells from healthy astrocytes by exploiting Raman spectroscopy. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2022**, *269*, 120773. [\[CrossRef\]](#)
32. Hage Chehade, A.; Abdallah, N.; Marion, J.M.; Oueidat, M.; Chauvet, P. Lung and colon cancer classification using medical imaging: A feature engineering approach. *Phys. Eng. Sci. Med.* **2022**, *45*, 729–746. [\[CrossRef\]](#)
33. Brozek-Pluska, B. Statistics assisted analysis of Raman spectra and imaging of human colon cell lines—Label free, spectroscopic diagnostics of colorectal cancer. *J. Mol. Struct.* **2020**, *1218*, 128524. [\[CrossRef\]](#)
34. Liu, W.; Wang, H.; Du, J.; Jing, C. Raman microspectroscopy of nucleus and cytoplasm for human colon cancer diagnosis. *Biosens. Bioelectron.* **2017**, *97*, 70–74. [\[CrossRef\]](#)
35. Bergholt, M.S.; Zheng, W.; Lin, K.; Wang, J.; Xu, H.; Ren, J.L.; Ho, K.Y.; Teh, M.; Yeoh, K.G.; Huang, Z. Characterizing variability of in vivo Raman spectroscopic properties of different anatomical sites of normal colorectal tissue towards cancer diagnosis at colonoscopy. *Anal. Chem.* **2015**, *87*, 960–966. [\[CrossRef\]](#)
36. Verdonck, M.; Denayer, A.; Delvaux, B.; Garaud, S.; De Wind, R.; Desmedt, C.; Sotiriou CWillard-Gallo, K.; Goormaghtigh, E. Characterization of human breast cancer tissues by infrared imaging. *Analyst* **2016**, *141*, 606–619. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.