optics

MDPI

Article

Advanced Raman Spectroscopy Based on Transfer Learning by Using a Convolutional Neural Network for Personalized Colorectal Cancer Diagnosis

Dimitris Kalatzis [1], Ellas Spyratou [1,2], Maria Karnachoriti [2,3], Maria Anthi Kouri [1,2,4], Spyros Orfanoudakis [5], Nektarios Koufopoulos [6], Abraham Pouliakis [6], Nikolaos Danias [7], Ioannis Seimenis [8], Athanassios G. Kontos [3] and Efstathios P. Efstathopoulos [1,*]

1 2nd Department of Radiology, Medical School, National and Kapodistrian University of Athens, 12462 Athens, Greece
2 Department of Medicine, Democritus University of Thrace, 68100 Alexandroupolis, Greece
3 Physics Department, School of Applied Mathematical and Physical Sciences, National Technical University of Athens, Iroon Politechniou 9, Zografou, 15780 Athens, Greece
4 Medical Physics Program, University of Massachusetts Lowell, 265 Riverside St, Lowell, MA 01854, USA
5 Alpha Information Technology S.A., Software & System Development, 39 Dimokratias Avenue, 68131 Alexandroupolis, Greece
6 2nd Department of Pathology, School of Medicine, Attikon University Hospital, National and Kapodistrian University of Athens, 12462 Athens, Greece
7 4th Department of Surgery, School of Medicine, Attikon University Hospital, University of Athens, 1 Rimini Street, 12462 Athens, Greece
8 Medical School, National and Kapodistrian University of Athens, 75 Mikras Assias Str., 11527 Athens, Greece
* Correspondence: stathise@med.uoa.gr

check for updates

Citation: Kalatzis, D.; Spyratou, E.; Karnachoriti, M.; Kouri, M.A.; Orfanoudakis, S.; Koufopoulos, N.; Pouliakis, A.; Danias, N.; Seimenis, I.; Kontos, A.G.; et al. Advanced Raman Spectroscopy Based on Transfer Learning by Using a Convolutional Neural Network for Personalized Colorectal Cancer Diagnosis. *Optics* **2023**, *4*, 310–320. https://doi.org/10.3390/opt4020022

Academic Editor: Christoph Gerhard

Received: 28 February 2023
Revised: 4 April 2023
Accepted: 25 April 2023
Published: 27 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/).

Abstract: Advanced Raman spectroscopy (RS) systems have gained new interest in the field of medicine as an emerging tool for in vivo tissue discrimination. The coupling of RS with artificial intelligence (AI) algorithms has given a boost to RS to analyze spectral data in real time with high specificity and sensitivity. However, limitations are still encountered due to the large amount of clinical data which are required for the pre-training process of AI algorithms. In this study, human healthy and cancerous colon specimens were surgically resected from different sites of the ascending colon and analyzed by RS. Two transfer learning models, the one-dimensional convolutional neural network (1D-CNN) and the 1D–ResNet transfer learning (1D-ResNet) network, were developed and evaluated using a Raman open database for the pre-training process which consisted of spectra of pathogen bacteria. According to the results, both models achieved high accuracy of 88% for healthy/cancerous tissue discrimination by overcoming the limitation of the collection of a large number of spectra for the pre-training process. This gives a boost to RS as an adjuvant tool for real-time biopsy and surgery guidance.

Keywords: Raman spectroscopy; colorectal cancer; tissue discrimination; convolutional neural network; transfer learning

1. Introduction

Colorectal cancer is the third most common cancer diagnosed in Europe, according to the latest incidence data provided by the World Cancer Research Fund International [1]. Early diagnosis has proven to be key to reducing cancer-related mortality. In cases where surgical operations are required for cancer treatment, accurate discrimination between healthy and cancerous tissues is critical for the postoperative care of the patient. Conventional imaging techniques that are currently used for cell discrimination are limited in sensitivity and resolution; both are required to diagnose cancer families or smaller cancer nodules. Biopsy, i.e., tissue excision for histological examination, remains the gold standard

for diagnostic purposes by offering a high rate of diagnostic sensitivity and selectivity. However, waiting for the results of biopsy testing can be up to a few days after the surgical operation, due to there not being a direct procedure for the technique. In addition, biopsies also suffer from low selective spatial coverage, which is necessary. Thus, in cases where remnants of malignant tissue still exist in the patient, a second surgical operation or/and radiotherapy is required. As a result, diagnosis and surgical treatment can be a multiple-step procedure instead of a one-step procedure, a crucial fact for the outcome of the surgical operation.

Raman spectroscopy (RS) is a biophotonic technique with high detection sensitivity and specificity. Any biochemical change in tissue constitution can be reflected in its spectral fingerprint [2,3]. Numerous studies have been published on the use of RS as an effective tool for tissue discrimination [4–6]. The evolution of fiber-optic probe design, the development of detector technology, and the use of artificial intelligence (AI) for spectral data analysis have enabled access to Raman spectroscopy in clinical praxis [7].

Advanced Raman systems can be used for guided biopsy and surgery endoscopically for the detection of the oral cavity, the bladder, the gastrointestinal tract, the colon, the rectum, or superficially on skin. Raman-spectroscopy-based systems for diagnosis and surgery guidance have demonstrated sensitivities and specificities in the range of 73–100% and 66–100%, respectively [8,9]. For the detection of colorectal cancer, ex vivo Raman measurements on colon tissue have reached 89% specificity for non-malignant/malignant group classification [10]. Endoscopic multi-fiber Raman probes have been used in vivo on colon tissue for the detection of adenomatous polyps/hyperplastic polyps with sensitivity and specificity reaching 91% and 83%, respectively [11].

In particular, AI tools have given a boost for automating in real-time Raman spectra data processing for tissue classification with high sensitivity and specificity [12,13]. Deep learning AI algorithms such as the faster region-based convolutional neural network (faster-RCNN) have achieved high sensitivity of 97.4% and lower specificity of 60.3% in the histopathological screening of colorectal cancer [14]. A one-dimensional convolutional neural network (1D-CNN) was trained with raw Raman data in order to classify either prostate cancer or healthy tissue and it achieved over 93% accuracy [15]. Higher accuracy was reached by using a one-dimensional residual convolutional neural network (1D-ResNet) architecture. Under this scope, over 20,000 Raman data of colorectal cancer tissue were used and the deep learning (DL) algorithm achieved up to 98.5% accuracy in cancer detection [16].

On the other hand, conventional AI algorithms need a large-scale database of Raman spectra for training [17]. Especially in medical research, the collection of a large amount of clinical data is a challenging task [18]. An approach to overcome this barrier is transfer learning (TL) [19]. Recently, a public dataset of well-characterized minerals was used for TL methods to improve the classification of the Raman spectra of pesticides by 6% [20], as well as the identification of Raman spectra of organic compounds by 4.1% [21]. Through this method, a large database of similar data to the available data is used to create a pre-trained model and then this knowledge is applied to solve the new problem. In this study, we used a large public Raman database [22] to create the pre-trained models. The gained knowledge was applied to our Raman data from colorectal cancer to predict whether they are healthy or cancerous tissue. Subsequently, to investigate if TL increased the accuracy of our classification task, we compared the two models, 1D–CNN, and 1D–ResNet, with and without transfer learning.

2. Materials and Methods

2.1. Raman Spectra Collection

Human colon specimens were collected from 12 patients who underwent open surgery. The specimens were extracted after written informed consent was obtained from the patients and the approval of the Ethics Committee of the School of Medicine of Attikon University Hospital was gained. Cancerous and healthy tissues were surgically resected

from different sites of the ascending colon. The colon tissues were categorized by standard histopathological evaluation based on hematoxylin and eosin sections [23]; nowadays, microscopy-based histopathological examination is the gold standard for colorectal cancer diagnosis, as well as staging and grading. All the specimens were cut into slices of $5 \times 5 \times 0.5$ mm with a hand-made microtome and kept fresh in a non-toxic-based fixative (Z7) used for tissue preservation before they were measured [24].

Micro-Raman measurements were conducted by a Renishaw Invia spectrometer at 785 nm laser excitation which induces low autofluorescence background signals. The spectrometer was equipped with a 1200 grooves/mm diffraction grating for analyzing the Raman signal and a thermoelectrically cooled CCD detector. The measurements were conducted with a $\times 50$ microscope objective and NA = 0.5 numerical aperture and a laser power density of 20 mW/μm^2. The spectra were recorded in the 500 to 3300 cm^{-1} frequency range.

2.2. Raman Data and Pre-Processing

For this study, a total of 248 samples of 12 patients/sets were used. Table 1 summarizes the number of spectra of healthy and cancerous data per set, the total spectra per class, as well as the cancer stage and grade of each patient. The labeling of the data was performed based on the visible boundaries between healthy and cancerous tissue.

Table 1. Data set of the Raman spectra.

Patients/Sets	Healthy	Cancerous	Stage	Grade
set1	3	5	ypT3N2a	G2
set2	18	10	pT1N0	G1
set3	8	10	pT4bN1	G2
set4	13	11	pT3N0	G2
set5	9	10	pT3pN0	G1
set6	7	5	pT3N0	G1
set7	13	10	pT3N1M1	G2
set8	13	10	pT3N1c	G2
set9	10	9	pT2N0	G2
set10	14	10	pT3N0	G2
set11	16	15	pT2N0	G2
set12	10	9	pT3N0Mx	G2
Total Spectra	134	114		

Regarding the pre-processing of the data, three procedures were applied to the dataset. Typically, the intensity of the background is substantially higher than the faint Raman's signals.

Before the use of Raman spectra for training and testing, it is highly important to apply suitable background correction in order to improve the performance of our classification models [25]. Thus, the initial procedure includes background correction by using the sensitive non-linear iterative peak-clipping (SNIP) algorithm [26].

The following procedure is scaling which is performed via a min–max normalization method, in which the intensity values of each Raman spectrum were scaled within the range [0, 1], where 0 is the lower value and 1 is the higher value of the spectrum intensity. For the final procedure, only the region of 800–1800 cm^{-1} was chosen as input in our deep learning models. Thus, each sample consists of 1000 one-dimensional intensity values where each value corresponds to one wavenumber. The same pre-processing methods were used with and without transfer learning models in order to achieve accurate comparison.

2.3. Pre-Training Dataset and Classification Models

Deep learning for medicine applications requires the collection of a plethora of training data, which in turn renders it a challenging task. Particularly in the case of clinical data

collection, there are plenty of limitations i.e., the limited number of patients, as well as privacy and security regulations.

To overcome the aforementioned difficulties, we used a public dataset which is needed for the pre-training phase, as introduced by Ho and colleagues [22]. This Raman spectra dataset consists of 60,000 spectra of pathogen bacteria, 2000 spectra per class, and 30 classes.

Cross-validation is a widely used method for assessing the performance of models. In our study, we employed a leave-p-out cross-validation approach to evaluate the performance of our models [27]. This approach involves leaving out two patients at a time and then training the model on the remaining data. We then evaluated the model's performance on the held-out patients. (Figure 1). This method can be considered a form of external validation, as the held-out patients were not used during the model training. Due to the scarcity of Raman spectra data from human tissue, we could not obtain additional external datasets for validation. To address this issue, we used a leave-p-out approach, which is a well–established method for evaluating model performance in a small dataset. We would like to point out that our approach provides a robust estimate of the model's performance since we calculated the average performance over the six folds.

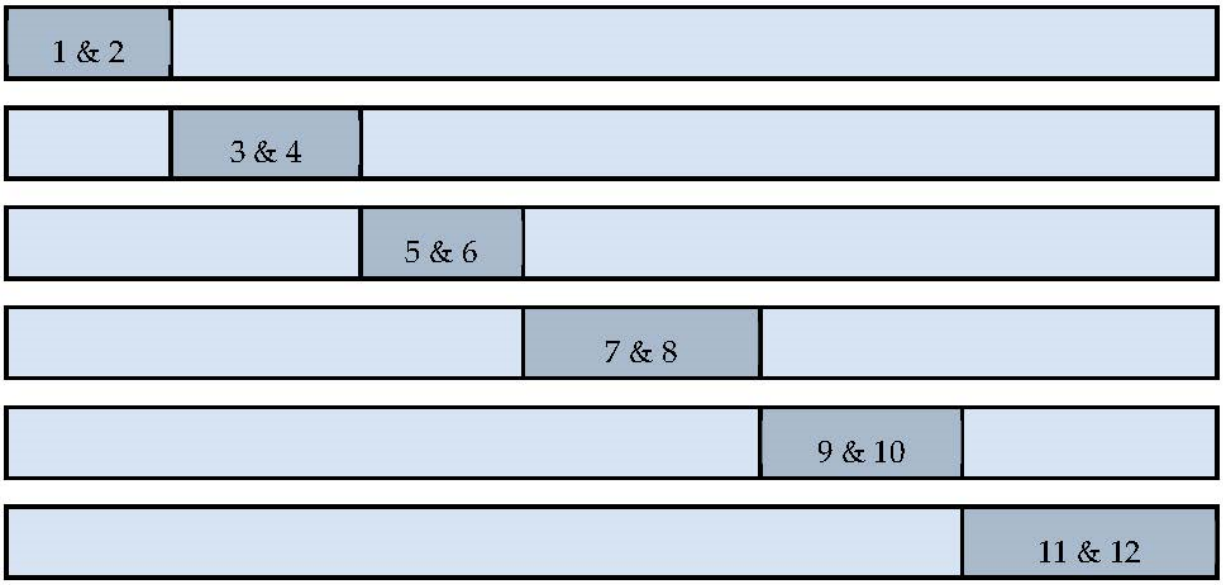

Figure 1. Leave-p-out cross-validation scheme.

2.3.1. Transfer Learning

Transfer learning is a useful tool when a small amount of data are available to train a model. Thus, by transferring learned features from one task to a similar one, it is possible to achieve higher performance.

Specifically, three steps took place in order to develop the classification models. First, the models were pre-trained by using the public Raman dataset [22], then the layers with extracted features were saved, and in the final step, freezing and training occurred (Figure 2). In this final step, our clinical datasets were used (Table 1) to re-train the pre-trained models with lower learning rates where all the layers are frozen, except the last two layers (dense classifier). A comparison between two models (1D–CNN and 1D–ResNet) with and without transfer learning into 248 spectra was performed in order to evaluate their classification task (healthy or cancerous).

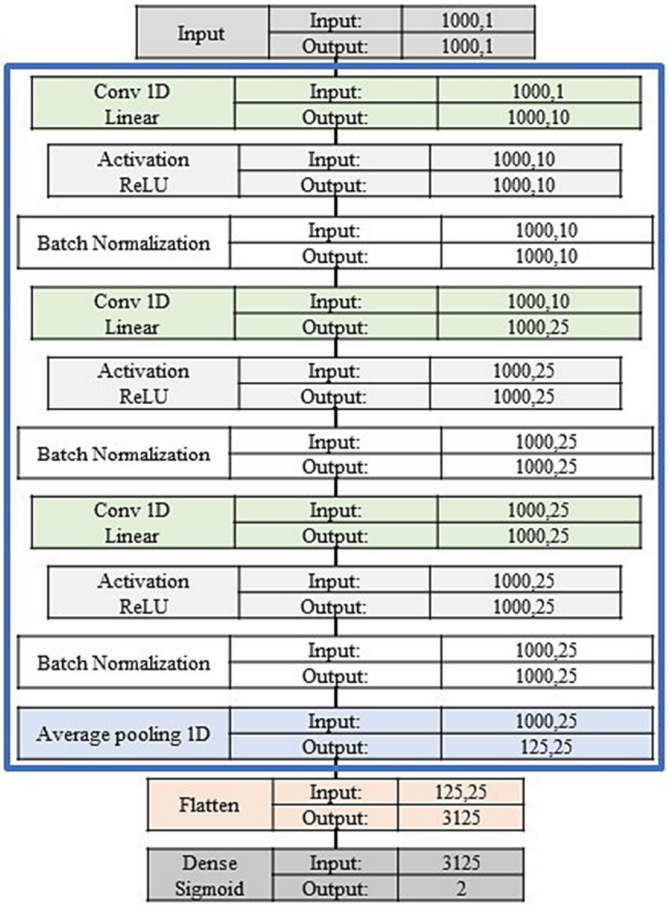

Figure 2. 1D-CNN model, the layers boxed in blue were used for transfer learning.

Convolutional Neural Network (CNN)

LeCun et al. were first to introduce the application of a CNN in the recognition of handwritten digits [28]. Since then, many researchers have relied on the use of a CNN in two-dimensional data in order to achieve image analysis. Therefore, its high accuracy and high processing speed have rendered the CNN as the primary tool in various computer vision tasks [29,30]. In our study, a similar architecture model consisting of one-dimensional spectra data was designed.

More specifically, we designed a 1D-CNN four-layer network which includes an input layer and four hidden layers, a convolutional layer of three convolutions with ReLU activation functions, a pooling layer, a flattened layer, and one dense layer with a sigmoid activation function.

Initially, the samples from the input layer are fed into the convolution layer, and subsequently, in the convolution layer, a filter or feature detector is applied to the input samples to generate an activation map using the ReLU activation function. Then, the output from the convolution layer passes into the pooling layer to reduce the size of the activation map. Finally, the pooled activation map is passed into a flattened layer to convert the data into a one-dimensional array for inputting them into the dense layer. In the dense layer, the weights are applied to the sigmoid activation function.

Residual Network (ResNet)

A residual network is a specific type of neural network that was introduced by He et al., for image recognition [31]. The architecture of the ResNet can avoid the vanishing gradient problem through a direct connection which skips some layers in between. These connections are known as skip connections or shortcut connections and constitute the core of residual blocks.

Herein, we developed a ResNet that consists of an input layer and 28 hidden layers, a convolutional layer followed by 6 residual layers, and each residual layer contains four convolutional layers with the ReLU activation function. The initial convolution layer has 64 convolutional filters, while the remaining 6 layers have 100 filters. Then, the pooling layer and the flattened layer follow, as well as the dense layer with s sigmoid activation function. Thus, the total depth of the network is 28 layers.

Python 3 programming language and TensorFlow libraries were used to write the code for the implementation of Deep Learning models. A graphic processing unit such as NVIDIA RTX3070Ti was chosen to handle the training of Deep Learning models.

3. Results and Discussion

3.1. Pre-Processing Data

Different approaches are referred to in the literature on the Raman spectra range. In particular, Zhang et al. [32], reported that the range between 300 cm^{-1} and 1800 cm^{-1} contains a plethora of information about the cellular components. Lin et al. [11], in a pilot study with Raman spectroscopy in colorectal tissues, showed that the fingerprint range 800–1800 cm^{-1} and the high-wavenumber 2800–3600 cm^{-1} provided valuable information for the assessment of colorectal carcinogenesis.

Figure 3 shows the mean value and the standard deviation of the Raman spectrum from healthy and cancerous colorectal tissues in the range of 500–3200 cm^{-1}. Raman spectral profiles of normal and cancerous tissues are very similar with bands observed at the same frequencies. The key distinction between them is the different normalized intensity of certain modes, indicating subtle but significant differences in molecular composition [33]. The most significant changes observed in the Raman spectra of cancerous tissues are the decrease in the intensity of the 854 and 2852 cm^{-1} bands and the increase for the 1005, 1341m, and 2937 cm^{-1} ones. Figure 3 depicts the two spectra ranging from 800–1800 cm^{-1} and 2200–3200 cm^{-1}. In this particular study, only the range 800–1800 cm^{-1} was selected for the training and test dataset creation. The reason behind this choice was the exported results from the following procedure. First, we created two different datasets containing 248 samples each. The spectra range for the first and second datasets was 800–1800 cm^{-1}, and 2200–3200 cm^{-1}, respectively. Then, each dataset was used to train the 1D-CNN and 1D-ResNet in order to evaluate which one of the two datasets gives the best results in a classification task.

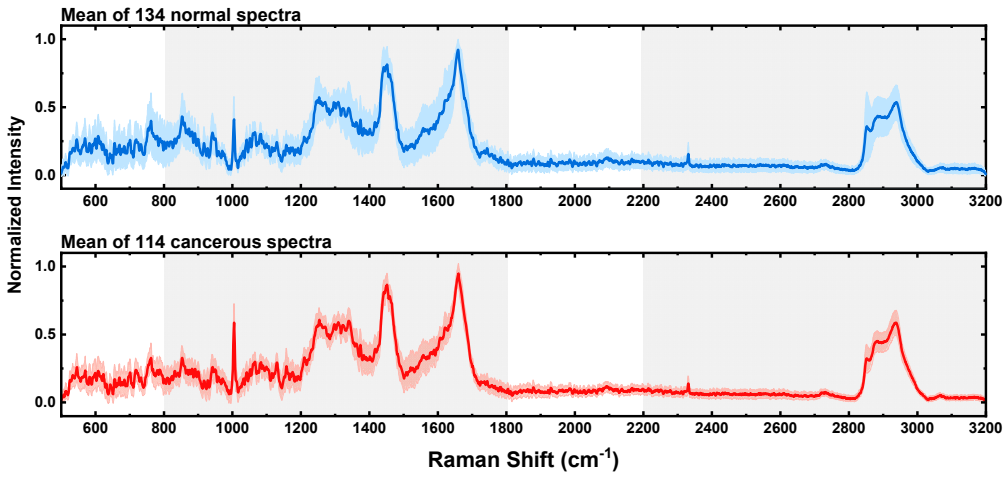

Figure 3. The average and standard deviation (SD) Raman spectra of normal (blue solid line with light blue space) and cancerous (red solid line with light red space) colorectal tissues (normal (n = 134) and cancer (n = 114)). The grey shaded areas on the graphs mark the spectral regions that were used to train the 1D−CNN.

Table 2 shows the rates of 1D-CNN and 1D-ResNet between the different datasets. Confusion matrices are represented in Figure 4. The 800–1800 cm^{-1} range of 1D−CNN exhibits 7.2% higher accuracy than the 2200–3200 cm^{-1} range (Table 2). Comparing the two ranges, we observed that the recall and precision of 800–1800 cm^{-1} were higher by 10.5% and 6.1%, respectively. In addition, similar results were noticed in 1D-ResNet between the two different ranges.

Table 2. Results of 1D-CNN in the two different datasets from colorectal cancer tissues.

Model	Wavenumber (cm^{-1})	Accuracy	Recall	Precision	f1_Score
1D-CNN	800–1800	0.834	0.859	0.796	0.827
1D-CNN	2200–3200	0.762	0.754	0.735	0.744
1D-ResNet	800–1800	0.850	0.859	0.823	0.841
1D-ResNet	2200–3200	0.814	0.850	0.769	0.808

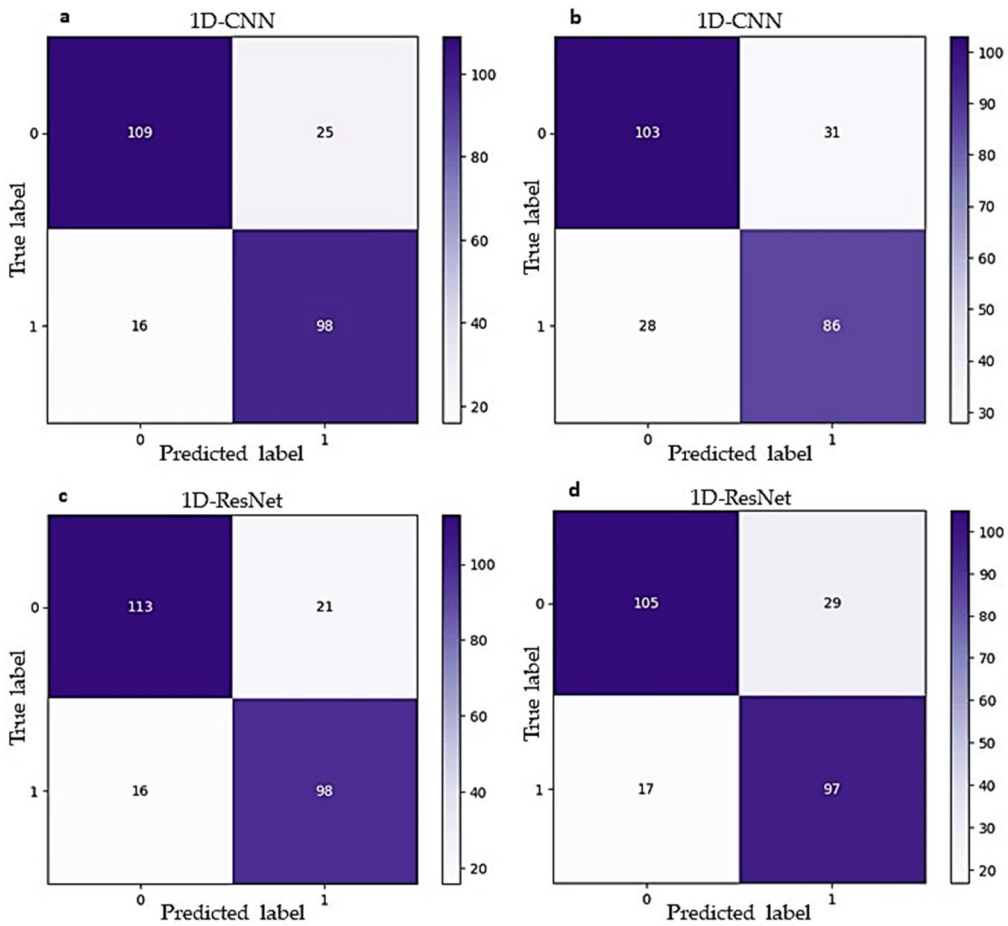

Figure 4. Confusion matrix of 1D-CNN: (**a**) 800–1800 cm^{-1} and (**b**) 2200–3200 cm^{-1} and 1D-ResNet: (**c**) 800–1800 cm^{-1} and (**d**) 2200–3200 cm^{-1}. The labels 0 and 1 correspond to healthy and cancerous tissue, respectively.

3.2. Transfer Learning vs. Non-Transfer Learning

In the final phase of the experiments, two deep learning models (1D-CNN and 1D-ResNet) with and without transfer learning were trained with Raman data in the range of 800–1800 cm^{-1}. As Table 3 illustrates, the accuracy of 1D-CNN with transfer learning was 5.3% higher than that of 1D-CNN without transfer learning. A lower percentage of the order of 2% was noticed in 1D-ResNet without transfer learning compared to the 1D-RestNet with transfer learning.

Table 3. Results of transfer learning and non-transfer learning models in the range 800–1800 cm^{-1}.

Model	Accuracy	Recall	Precision	f1_Score
1D-CNN	0.834	0.859	0.796	0.827
1D-CNN transfer	0.887	0.885	0.870	0.878
1D-ResNet	0.850	0.859	0.823	0.841
1D-ResNet transfer	0.870	0.833	0.879	0.855

A noteworthy observation is that the recall of 1D-CNN with the transfer learning model is quite high. In particular, from the total 114 cancerous samples, only 13 could be classified as healthy (Figure 5a). The recall rate is highly important because a false-negative estimation from the deep learning model in the clinical application can be fatal for the patient. As Table 3 shows, the 1D-CNN with transfer learning presents 88.7% accuracy as well as 88.5% recall, which both are higher than the rates of all other models.

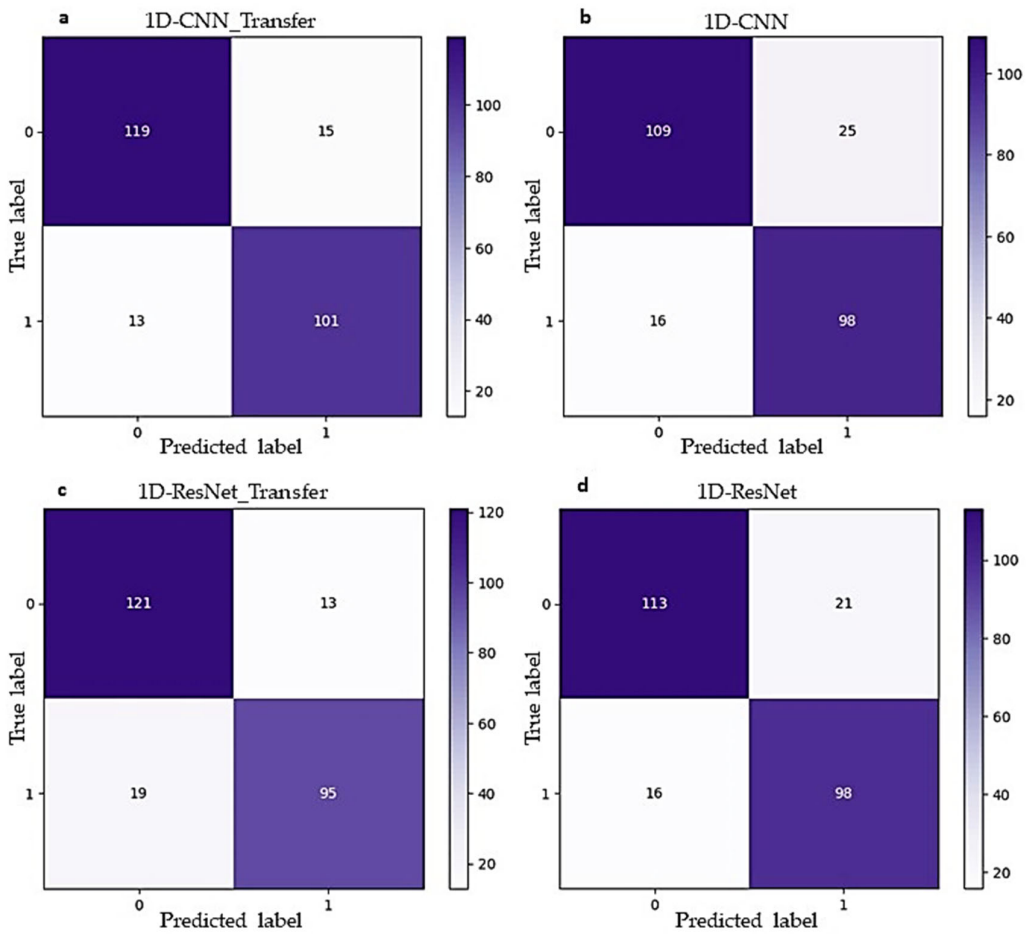

Figure 5. Confusion matrix: (**a**) 1D-CNN_transfer, (**b**) 1D-CNN, (**c**) 1D-ResNet_transfer, and (**d**) 1D-ResNer. The labels 0 and 1 correspond to healthy and cancerous tissue, respectively.

Furthermore, especially in 1D-CNN and 1D-ResNet with transfer learning the precision is 87% and 87.9%, respectively. The 1D-CNN with transfer learning gave 15 false-positive answers while the model without transfer learning gave 25 false-positive answers.

The box plot depicts the distribution of accuracies for four different models (Figure 6). Each box represents the range of accuracies for a single model, from the 25th percentile (bottom of the box) to the 75th percentile (top of the box). The horizontal orange line inside the box represents the median accuracy. The whiskers extend from the box to show the

range of accuracies, with any outliers shown as circles outside the whiskers. Figure 6 shows that the ResNet models generally have higher median accuracies than the 1D-CNN models and that adding transfer learning generally improves accuracy for both types of models.

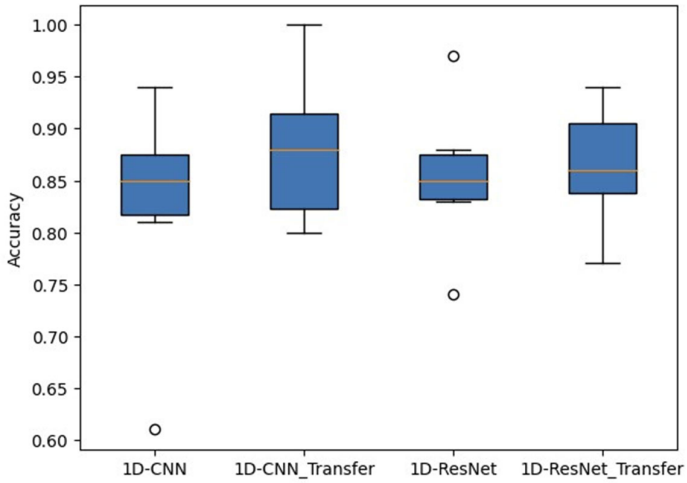

Figure 6. Box plot of the four models' accuracies resulting from the six folds of the leave-p-out validation (the white circles represent the outliers).

The Pearson correlation (Figure 7) exhibits that there is a strong positive correlation between the accuracies of the 1D-CNN and 1D-CNN with transfer learning, as well as between the accuracies of the ResNet and ResNet with transfer learning models.

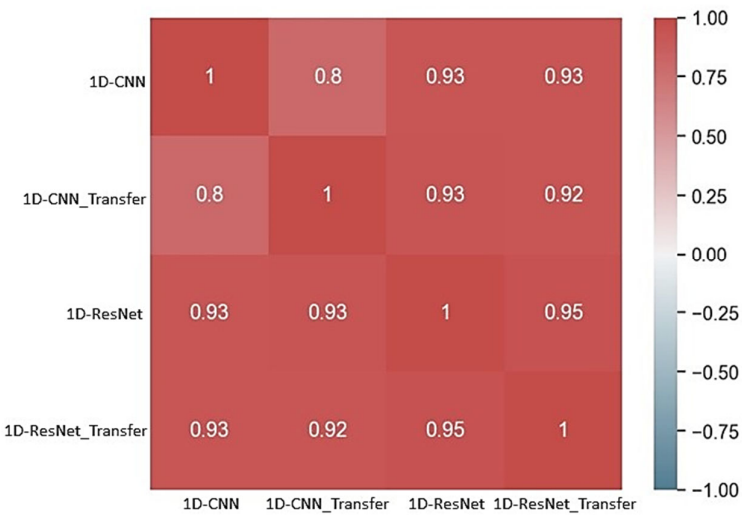

Figure 7. Pearson correlation matrix of the four models.

However, there is only a weak positive correlation between the accuracies of the 1D-CNN and ResNet models, and adding transfer learning does not appear to significantly increase this correlation. The strongest positive correlation is between the ResNet and ResNet with transfer learning models.

4. Conclusions

In summary, we developed and evaluated two transfer learning models using a Raman open database for the pre-training process. Comparing the transfer learning models with those without transfer learning, we proved that they leverage the knowledge they have acquired, resulting in a significant increase in prediction accuracy. In particular, the 1D-CNN transfer learning and 1D-ResNet transfer learning models achieve 5.3% and 2%

higher accuracy, respectively, in contrast to the non-transfer-learning models. In addition, the transfer learning method helped us to solve the severe limitations and difficulties we encountered in collecting a large amount of clinical data. Furthermore, the high recall values, especially in the 1D-CNN transfer learning model, which reached 88.5%, show that the model correctly identifies the majority of malignant tumors, minimizing false negative predictions. Thus, we provide models that can be a useful tool in the hands of surgeons, helping them to determine in real time, during the surgery, the healthy margins of the cancerous tumor.

Author Contributions: Conceptualization, D.K., E.S. and E.P.E.; writing—original draft preparation, D.K. and E.S.; methodology, D.K., E.S., A.G.K., M.A.K., M.K., S.O., A.P., N.K. and N.D.; software, D.K.; data curation, D.K., M.K., M.A.K., S.O., A.P., N.K., I.S. and A.G.K.; writing—review and editing, E.S., E.P.E., N.D., I.S., A.G.K. and E.P.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research was co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship, and Innovation, under the call RESEARCH–CREATE–INNOVATE (project code: T2EDK-01223). Title: Development of an advanced portable biophotonic system for the personalized spectroscopic discrimination of cancer margins/tissues (acronym: BIOPHASMA).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and was approved by the Scientific Council of the Bioethics and Ethics committee (present: Professor A. Psyrri. Consultant 1st class NHS A. Siatelis. Consultant 2nd class NHS, P. Economopoulou. Assistant Professor, N. Siafakas. University Education in nursing M. Tsirouda) of University General Hospital "Attikon", Administration of the 2nd Sanitary district of Piraeus and Aegean Islands, Hellenic Republic. The submission of the research protocol with title: "Development of advanced portable biophotonic system for the personalized spectroscopic discrimination of cancer margins/tissues-BIOPHASMA" (BAKTIN, EBΔ583/27-10-2021) was unanimously approved at the 10th (4 November 2021) meeting of the Bioethics and Ethics committee.

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

Data Availability Statement: Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the corresponding author upon reasonable request.

Acknowledgments: The authors thank the Institute of Nanoscience and Nanotechnology, NCSR Demokritos, Greece, directed by Polycarpos Falaras for the measurements that were carried out in their Raman laboratory.

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

References

1. Colorectal Cancer Statistics. Available online: https://www.wcrf.org/cancer-trends/colorectal-cancer-statistics/ (accessed on 11 December 2022).
2. Krafft, C.; Schie, I.W.; Meyer, T.; Schmitt, M.; Popp, J. Developments in spontaneous and coherent Raman scattering microscopic imaging for biomedical applications. *Chem. Soc. Rev.* **2016**, *45*, 1819–1849.
3. Santos, I.P.; Caspers, P.J.; Schut, T.B.; van Doorn, R.; Koljenović, S.; Puppels, G.J. Implementation of a novel low-noise InGaAs detector enabling rapid near-infrared multichannel Raman spectroscopy of pigmented biological samples. *J. Raman Spectrosc.* **2015**, *46*, 652–660. [CrossRef]
4. Çulha, M. Raman spectroscopy for cancer diagnosis: How far have we come? *Bioanalysis* **2015**, *7*, 2813–2824. [CrossRef] [PubMed]
5. Wang, W.; McGregor, H.; Short, M.; Zeng, H. Clinical utility of Raman spectroscopy: Current applications and ongoing developments. *Adv. Health Care Technol.* **2016**, *13*, 13–29. [CrossRef]
6. Camp, C.H., Jr.; Cicerone, M.T. Chemically sensitive bioimaging with coherent Raman scattering. *Nat. Photonics* **2015**, *9*, 295–305.
7. Krishna, H.; Majumder, S.K.; Chaturvedi, P.; Sidramesh, M.; Gupta, P.K. In vivo Raman spectroscopy for detection of oral neoplasia: A pilot clinical study. *J. Biophotonics* **2014**, *7*, 690–702. [CrossRef]
8. Wang, W.; Zhao, J.; Short, M.; Zeng, H. Real-time in vivo cancer diagnosis using raman spectroscopy. *J. Biophotonics* **2015**, *8*, 527–545. [CrossRef]

9. Santos, I.P.; Barroso, E.M.; Schut, T.C.B.; Caspers, P.J.; van Lanschot, C.G.F.; Choi, D.-H.; van der Kamp, M.F.; Smits, R.W.H.; van Doorn, R.; Verdijk, R.M.; et al. Raman spectroscopy for cancer detection and cancer surgery guidance: Translation to the clinics. *Analyst* **2017**, *142*, 3025–3047.
10. Short, M.A.; Tai, I.T.; Owen, D.; Zeng, H. Using high frequency Raman spectra for colonic neoplasia detection. *Opt. Express* **2013**, *21*, 5025–5034. [CrossRef]
11. Bergholt, M.S.; Lin, K.; Wang, J.; Zheng, W.; Xu, H.; Huang, Q.; Ren, J.; Ho, K.Y.; Teh, M.; Srivastava, S.; et al. Simultaneous fingerprint and high-wavenumber fiber-optic Raman spectroscopy enhances real-time in vivo diagnosis of adenomatous polyps during colonoscopy. *J. Biophotonics* **2016**, *9*, 333–342. [CrossRef]
12. 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]
13. He, H.; Yan, S.; Lyu, D.; Xu, M.; Ye, R.; Zheng, P.; Lu, X.; Wang, L.; Ren, B. Deep Learning for Biospectroscopy and Biospectral Imaging: State-of-the-Art and Perspectives. *Anal. Chem.* **2021**, *93*, 3653–3665. [CrossRef]
14. Ho, C.; Zhao, Z.; Chen, X.F.; Sauer, J.; Saraf, S.A.; Jialdasani, R.; Taghipour, K.; Sathe, A.; Khor, L.Y.; Lim, K.H.; et al. A promising deep learning-assistive algorithm for histopathological screening of colorectal cancer. *Sci. Rep.* **2022**, *12*, 2222. [CrossRef] [PubMed]
15. Lee, W.; Lenferink, A.T.M.; Otto, C.; Offerhaus, H.L. Classifying Raman spectra of extracellular vesicles based on convolutional neural networks for prostate cancer detection. *J. Raman Spectrosc.* **2019**, *51*, 293–300. [CrossRef]
16. Cao, Z.; Pan, X.; Yu, H.; Hua, S.; Wang, D.; Chen, D.Z.; Zhou, M.; Wu, J. A Deep Learning Approach for Detecting Colorectal Cancer via Raman Spectra. *BME Front.* **2022**, *2022*, 9872028. [CrossRef]
17. Ngiam, K.Y.; Khor, I.W. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol.* **2019**, *20*, 262–273. [CrossRef]
18. Jain, S.H.; Rosenblatt, M.; Duke, J. Is big data the new frontier for academic-industry collaboration? *JAMA* **2014**, *311*, 2171–2172. [CrossRef]
19. Weiss, K.; Khoshgoftaar, T.M.; Wang, D. A survey of transfer learning. *J. Big Data* **2016**, *3*, 9. [CrossRef]
20. Hu, J.; Zou, Y.; Sun, B.; Yu, X.; Shang, Z.; Huang, J.; Jin, S.; Liang, P. Raman spectrum classification based on transfer learning by a convolutional neural network: Application to pesticide detection. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2022**, *265*, 120366. [CrossRef]
21. Zhang, R.; Xie, H.; Cai, S.; Hu, Y.; Liu, G.-K.; Hong, W.; Tian, Z.-Q. Transfer-learning-based Raman spectra identification. *J. Raman Spectrosc.* **2020**, *51*, 176–186. [CrossRef]
22. Ho, C.-S.; Jean, N.; Hogan, C.A.; Blackmon, L.; Jeffrey, S.S.; Holodniy, M.; Banaei, N.; Saleh, A.A.E.; Ermon, S.; Dionne, J. Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning. *Nat. Commun.* **2019**, *10*, 4927. [CrossRef]
23. Fleming, M.; Ravula, S.; Tatishchev, S.F.; Wang, H.L. Colorectal carcinoma: Pathologic aspects. *J. Gastrointest. Oncol.* **2012**, *3*, 153–173. [PubMed]
24. Lykidis, D.; Van Noorden, S.; Armstrong, A.; Spencer-Dene, B.; Li, J.; Zhuang, Z.; Stamp, G.W. Novel zinc-based fixative for high quality DNA, RNA and protein analysis. *Nucleic Acids Res.* **2007**, *35*, 85. [CrossRef]
25. Chi, M.; Han, X.; Xu, Y.; Wang, Y.; Shu, F.; Zhou, W.; Wu, Y. An Improved Background-Correction Algorithm for Raman Spectroscopy Based on the Wavelet Transform. *Appl. Spectrosc.* **2019**, *73*, 78–87. [CrossRef]
26. Morháč, M. An algorithm for determination of peak regions and baseline elimination in spectroscopic data. *Nucl. Instrum. Methods Phys. Res. A* **2009**, *600*, 478–487. [CrossRef]
27. Airola, A.; Pahikkala, T.; Waegeman, W.; De Baets, B.; Salakoski, T. An experimental comparison of cross-validation techniques for estimating the area under the ROC curve. *Comput. Stat. Data Anal.* **2011**, *55*, 1828–1844. [CrossRef]
28. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.* **1989**, *1*, 541–551. [CrossRef]
29. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional neural networks: An overview and application in radiology. *Insights Imaging* **2018**, *9*, 611–629. [CrossRef]
30. Borchers, A.; Pieler, T. Programming pluripotent precursor cells derived from Xenopus embryos to generate specific tissues and organs. *Genes* **2010**, *1*, 413–426. [CrossRef]
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
32. Zhang, L.; Li, C.; Peng, D.; Yi, X.; He, S.; Liu, F.; Zheng, X.; Huang, W.E.; Zhao, L.; Huang, X. Raman spectroscopy and machine learning for the classification of breast cancers. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2022**, *264*, 120300. [CrossRef]
33. Kouri, M.A.; Spyratou, E.; Karnachoriti, M.; Kalatzis, D.; Danias, N.; Arkadopoulos, N.; Seimenis, I.; Raptis, Y.S.; Kontos, A.G.; Efstathopoulos, E.P. Raman Spectroscopy: A Personalized Decision-Making Tool on Clinicians' Hands for In Situ Cancer Diagnosis and Surgery Guidance. *Cancers* **2022**, *14*, 1144. [CrossRef] [PubMed]

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.