Deep learning for the prediction of the chemotherapy response of metastatic colorectal cancer: comparing and combining H&E staining histopathology and infrared spectral histopathology

Benjamin Brunel, Pierre Prada, Florian Slimano, Camille Boulagnon-Rombi, Olivier Bouché, Olivier Piot

Deep learning for the prediction of the chemotherapy response of metastatic colorectal cancer: comparing and combining H&E staining histopathology and infrared spectral histopathology

Benjamin Brunel[ab‡], Pierre Prada[a‡], Florian Slimano[a], Camille Boulagnon-Rombi[c], Olivier Bouché[ad] and Olivier Piot*[a]

Colorectal cancer is a global public health problem with one of the highest death rates. It is the second most deadly type of cancer and the third most frequently diagnosed in the world. The present study focused on metastatic colorectal cancer (mCRC) patients who had been treated with chemotherapy-based regimen for which it remain uncertainty about the efficacy for all eligible patients. This is a major problem, as it is not yet possible to test different therapies in view of the consequences on the health of the patients and the risk of progression. Here, we propose a method to predict the efficacy of an anticancer treatment in an individualized way, using a deep learning model constructed on the retrospective analysis fo the primary tumor of several patients. Histological sections from tumors were imaged by standard hematoxylin and eosin (HE) staining and infrared spectroscopy (IR). Images obtained were then processed by a convolutional neural network (CNN) to extract features and correlate them with the subsequent progression-free survival (PFS) of each patient. Separately, HE and IR imaging resulted in a PFS prediction with an error of 6.6 and 6.3 months respectively (28% and 26% of the average PFS). Combining both modalities allowed to decrease the error to 5.0 months (21%). The inflammatory state of the stroma seemed to be one of the main features detected by the CNN. Our pilot study suggests that multimodal imaging analyzed with deep learning methods allow to give an indication of the effective-ness of a treatment when choosing.

1. Introduction

Colorectal cancer is a global public health issue, with one of the highest number of deaths. It is the second most deadly and third most diagnosed type of cancer in the world.[1] In our study, we focused on metastatic colorectal cancers (mCRC) treated with chemotherapy and targeted therapies combination. Targeted therapies, such as bevacizumab[2], directed against the Vascular Endothelial Growth Factor (VEGF) or cetuximab/panitumumab[3], targeting Epidermal Growth Factor receptor (EGFR), have shown clinical activity when combined to chemotherapy in patients with cancer.[4] However, this improvement is conditioned by the development of resistance of cancer cells to targeted therapies.[5] The appearance of numerous pro-angiogenic drivers during anti-VEGF treatment[6], and different mutations in KRAS gene during anti-EGFR treatment[7], may lead to a resistance of cancer cells to targeted therapies. Therefore, this uncertainty about the effectiveness of such treatments is a major problem, especially if we consider the time required to test different therapies and the consequences that can have on the health of patients. For that reason, predicting the effectiveness of a treatment response in an individualized manner would be a crucial advantage.

To gain insight on the condition and characteristics of a tumor, histology remains a reference method. By staining histological sections with standard hematoxylin and eosin staining (HE) the different cellular structures of a tissue (tumor, necrosis, fibrous stroma…) can be identified. Therefore, the colored section guides pathologists and oncologists to make a diagnosis and to propose the best treatments, based on the morphological types of cells (such as tumor cells, immune cells...). Indeed, the response to treatment varies greatly depending on the differentiation of cancer cells, the surrounding stroma, the abundance of immune cells, and the mutational profile in cancer cells which can lead to its resistance.[8]

Mid-Infrared (IR) imaging has proven to be a valuable analytical method for characterizing tumors and their environment. IR spectroscopy probes the vibrations of

[a.] *Université de Reims Champagne-Ardenne, EA7506-BioSpectroscopie Translationnelle (BioSpecT), Reims, France.*
[b.] *Université de Franche-Comté, CNRS, institut FEMTO-ST, F-25000 Besançon, France.*
[c.] *Laboratoire d'anatomie et cytologie pathologique, CHU Reims, 51100 Reims, France.*
[d.] *Service d'Oncologie Digestive, CHU Reims, 51100 Reims, France.*
‡ *These authors contributed equally.*

chemical bonds and thus provides a biomolecular fingerprint of cells and tissues. Each pixel within the image comprises an entire mid-IR spectrum that can give information on the biochemical content of the sample at a scale of few micrometers. For example, it is possible to reveal a molecular heterogeneity of tumor, and to characterize their microenvironment, without labeling nor particular preparation of the sample.[9,10] Histology studies based on the numerical analysis of spectral images acquired on human colic tissues are gaining prominence.[11,12] Similarly to the segmentation done in standard histology, IR spectroscopy images can be numerically segmented by using multivariate analysis, to visualize the different histological structures of the sample. This method, called spectral histopathology, makes it possible to obtain a digitally colored image, that is informative of the physiopathology status of the tissue. Recently, studies have automated spectral histopathology[13] and optimized it[14], on colon samples of patients. Spectral histopathology is set as a promising diagnostic and prognostic tool on which clinicians can rely on to choose the most appropriate treatment.[15,16]

Deep learning algorithms, and more precisely convolutional neural networks (CNN), are powerful tools to detect spatial characteristics in an image. In the biomedical field, such characteristics can be correlated to a disease.[17] Models are trained on labeled data and can then be applied to unknown data. For example, CT images analyzed by deep learning have being put forward to guide the choice of treatment in CRC.[18] In the field of histopathology, deep learning algorithms have been proposed to perform tumor detection or grading tasks.[19] In the field of vibrational biospectroscopy, their use begins to emerge.[20,21]

In this study, we propose to evaluate a method to predict the best treatment for mCRC patients from images of histological sections of the primary tumor, obtained before the treatment was set up. Our approach consists in applying deep learning algorithm on HE staining and infrared pseudocolor-coded images, first separately and then in a combined manner, to test the possibility to predict the treatment efficacy evaluated by considering the progression free survival (PFS). To the best of our knowledge, this pilot study is the first to compare and combine standard and spectral histopathology, using deep learning analysis, to provide indicators of treatment response in mCRC.

2. Materials and methods

2.1 Study design and population

The study was conducted on surgical sections of primary tumors of CRC. Our samples were provided by the tumor biobank of the French University Hospital of Reims. The criteria for samples inclusion were: i) to correspond to stage IV metastatic CRC, ii) localized in the left side of the colon (left transverse colon, descending sigmoid colon and rectum if not irradiated), iii) and treated in the first line with FOLFIRI + bevacizumab palliative chemotherapy regimen (irinotecan + folinic acid + bolus and infusional 5-fluorouracil combined with an anti-angiogenic (bevacizumab)). The effectiveness of treatments depends on the side (left or right) of the primary tumor.[22] We chose to carry out our study on left-sided CRCs because the range of PFS is wider for left-sided CRCs, consequently the prediction is of greater importance. Also, these lesions are more frequent.

The primary endpoint was the progression free survival (PFS) of the patients from whom the primary tumor was taken. PFS is the period of time between the first day of treatment and the day when cancer progression is observed. In practice, this period without progression corresponds to the length of time during which first-line treatment is maintained, the treatment being considered effective since it prevents the progression of the tumor (assuming that only the treatment has an influence on the progression of the cancer). A progression was considered if an increase of at least 20% in the size of the target lesion was observed on a CT scan (Response Evaluation Criteria in Solid Tumors (RECIST)), or if there was a change in the treatment or if the patient died. A progression was also taken into consideration if it was observed after a therapeutic discontinuation because the patient's condition was too deteriorating, or because the patient had refused the treatment. However, a progression was not counted, if it was observed during a therapeutic break.

2.2 Ethical approval

The study was conducted in accordance with the Helsinki Declaration. Informed written consent was obtained from each patient to inclusion of the tumor in biobank. Patients' records were anonymized prior to analysis. The data base was constituted in accordance with the reference methodology MR004 (MR00414012022) of the French National Commission on Informatics and Liberty (CNIL). As per French regulations concerning the retrospective study, no informed consent or additional ethical committee review was required.

2.3 Sample preparation

Frozen histological sections were prepared from the surgical resection of primary tumors of patients treated in Reims University Hospital. In our study, samples from 6 patients could be usable, after a histological examination of adjacent tissue sections, stained with HE. The main characteristics of the 6 patients for whom samples had been analyzed is given in Table 1.

Two adjacent tissue sections were performed; the thickness of sections was 6 µm for the first one and then 3µm for the second one. The first section was placed on a calcium fluoride window (CaF2) for FTIR imaging and the second one was placed on a glass slide for histology (HE staining).

Table 1 Clinical characterization of patients included in this study (F: female, M: male, m: mutated, w: wild).

Sex	Age (years)	KRAS	PFS (months)
M	44	w	14.9
F	72	m	15.3
F	58	m	15.4
F	66	m	20.7
M	69	w	35.1
M	83	m	41.3

2.4 Segmentation of HE images

HE images were segmented into histological classes before deep learning analysis. Segmentation was first done by histopathologists on a small portion of the surface of each image, using QuPath software.[23] Six distinctive classes were identified: cancer cells, immune cells, necrosis, cryptic cells, glandular lumen and stroma. The rest of the surface was segmented using machine learning trained on the set of the labelled portions of all images, as explained below.

First, the colors of the different HE images were normalized by estimating the stains vectors.[24] Doing so, colors were expressed as linear combinations of the two stains vectors (HE). Images were down sampled by a factor of 5 as the pixel size (0.17 µm) was very small as compared to the size of the histological structures and to the precision of their manual segmentation. Then, the classifier was trained at the level of the individual labelled pixels (around 24 million). Classification was not only based on the 2 colors of the pixels, but also on features calculated from the surrounding pixels. Selected features were: local (disk of 7 µm of radius) average, standard deviation and skew of colors, larger (disk of 43 µm of radius) average, standard deviation and skew of colors, and average (disk of 43 µm of radius) of the square of a ridge detection. A linear discriminant analysis (LDA) reduction was then performed to go from 16 to 6 features.

To predict the class of pixels, the 6-dimensional space of pixels was discretized into 35 bins for each dimension (ranging from the minimal to the maximal value of the feature), leading to 356 bins. Using the pixels of the training set, the mode of each bin was calculated. Applied on the testing set, the class of each pixel was selected as the class attributed to the corresponding bin. Pixels that correspond to unclassified bins form an "unclassified" class. This model has the advantage of not using any assumptions about the shape of the class distributions, and in particular about a linear shape. The number of pixels in each class was far from balanced. This could result in small classes being ignored by the model. To avoid this, pixels were weighted with the proportion of their class to the power of -2/3. This balancing step was tuned to give an approximate equality between the sensitivity and the precision of the prediction of each class.

The final step of segmentation consisted in smoothing the segmented images obtained with the classifier. Small patches (below 300 pixels) of each class were deleted, and each pixel was classified in the majority class of local pixels (disk of 9 µm of radius). These last three steps were repeated 3 times. To assess the accuracy of the segmentation process, the labelled pixels were separated into training and test data. We used cross-validation, which is more reliable than an evaluation based on a single separation.[25] Three folds seemed sufficient, given the large number of pixels (24 million). Finally, the model was applied to the unlabelled portion of HE images (figure 1a, c and e).

2.5 Instrumentation and FTIR data collections

FTIR imaging acquisitions of colonic samples were obtained using the spectrum spotlight TM 400 microscope coupled to spectrum One FTIR spectrometer (Perkin Elmer, Villebon sur Yvette, France).

The device was equipped with a cadmium and mercury telluride (MCT) detector cooled with liquid nitrogen. The frozen sections were imaged in IR with wavenumbers between 800 cm^{-1} and 4000 cm^{-1}, with a spectral resolution of 4 cm^{-1} and a pixel size of 25 x 25 μm^2.

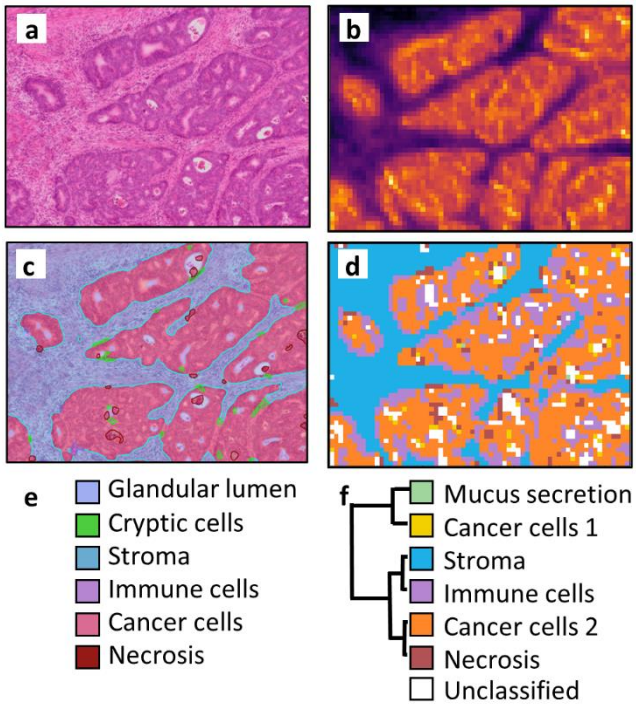

Fig. 1 Segmentation of HE stained (a,c,e) and IR spectroscopy (b,d,f) images. The same area for both modalities was selected as an example. HE images (a) are segmented semi-automatically (c) into six classes (e). IR images (at 1080 cm-1) (b) are segmented using k-means (d) into six classes related to one another as shown in the dendrogram (f). White pixels in (d) are unclassified pixels.

Before image acquisition, a background spectrum was recorded on a clean surface of the CaF2 window and automatically removed from the tissue spectrum for each pixel using the Spectrum Image Software (Perkin Elmer, Waltham, MA, USA). During acquisition, the sample was placed in a chamber purged with dry air. In addition, atmospheric correction of water vapor and CO2 was applied using Spectrum Image Software. Images preprocessing was carried out using internal algorithms written in Matlab (the Mathworks, Natick, MA, USA). An extended multiplicative signal corrections (EMSC) model was used to eliminate outlier spectra, correct scattering effects, and normalize spectra, with the target being the average of all spectra.[26]

2.6 Segmentation of IR spectroscopy images

K-means clustering [27] was applied to FTIR images to group the spectra according to their spectral similarity in 6 clusters. In order to visualize the k-means clustering image, different colors were assigned to each class and we obtained a reconstruction of the spectral image (figure 1d). The 6 spectral classes corresponded to histological structures slightly different from the 6 histological classes fixed from the HE stained images; these spectral classes are cancer cells type 1, cancer cells type 2, mucus secretion, necrosis, immune cells and stroma (figure 1f). Cancer cells type 1 might correspond to a mucinous carcinoma as they had a spectral signature close to mucus secretion and were mainly found in the histological cut of a single patient.

Figure 2 shows the IR absorbance spectra of the centroids of the different classes, with peaks at 968 cm-1(N+-CH3), 1083 cm-1 (symmetric PO2-), 1235 cm-1 (asymmetric PO2-), 1390 cm-1 (CH3), 1455 cm-1 (CH2), 1540 cm-1 (amide II), 1650 cm-1 (amide I). Differences in intensity could be observed mainly at 1083 cm-1, related to the proportion of DNA, and at 1650 cm-1 related to the proportion of protein and their conformation.

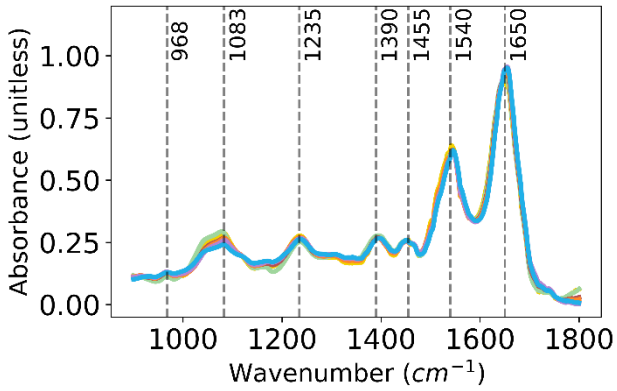

Fig. 2 FT-IR spectra of the centroids of the different classes determined by k-means. The colors of the lines correspond to those in figure 1f.

2.7 CNN for HE stained images analysis

To highlight the existence of a statistical correlation between HE images and the treatment efficacy, evaluated by the PFS, a CNN was used. Several images were associated with each patient, but to evaluate the model, a separation between training and test sets had to be done at the patient level. Otherwise, the presence of different images of the same patient in both sets would overestimate the model's ability to generalize to an unknown patient. As the number of patients was small, eliminating several patients for training would have had a major impact on model training, so we opted to leave-one-patient out validation. Consequently, all but one of the patients were used for training, and the last was used for validation. By changing the left-out patient, the model was evaluated on all patients, making the evaluation more robust. The evaluation criterion was the root mean square error (RMSE) of the PFS prediction. Image processing parameters values presented below were found by empiric optimization. They will be discussed in the "Results and Discussion" section. As the full spatial resolution was not necessary to measure general histological features and affects greatly the calculation time for training the CNN, we tried to down sample images, eventually by a factor of 27. Then, images were cut into squares, a more appropriate shape for CNN.

The choice of the size of the squares was a compromise between a larger number of small squares to train the network, and larger squares to be more representative of the tissue architecture. The selected size was 45 x 45 pixels, corresponding to 1160 x 1160 µm^2. Between 2 and 32 squares were extracted from each cut. To avoid too much imbalance between the patients in the training set, the number of squares was limited to a maximum of 4 by cut, with a random selection. Segmented squares were then reshaped into stacks of binary images, one for each class. Rather than keeping all the classes in the next processing step, we tried to select only some of them. Better RMSE was obtained when selecting only the two classes corresponding to immune cells and stroma.

Finally, a CNN was trained from the binary squares, with a leave-one-patient out validation scheme. Training data was augmented by adding 7 copies of each image, rotated (90°, 180°, 270°) or flipped and rotated (0°, 90°, 180°, 270°). The CNN was build using Keras API, whose architecture is presented in figure 3a. The input layer consisted in two (stroma and immune cells) 45x45 images. Then followed 3 repetitions of i) a 2D convolutional layer with 16 neurons, a kernel of 2x2 and a softplus activation function, ii) a batch normalization[28] and iii) a max pooling layer with a kernel of 2x2. The resulting outputs were flattened, a dropout of 20% was applied during training and a final batch normalization was applied. Finally, a linear regression was performed on the outputs to predict the PFS.

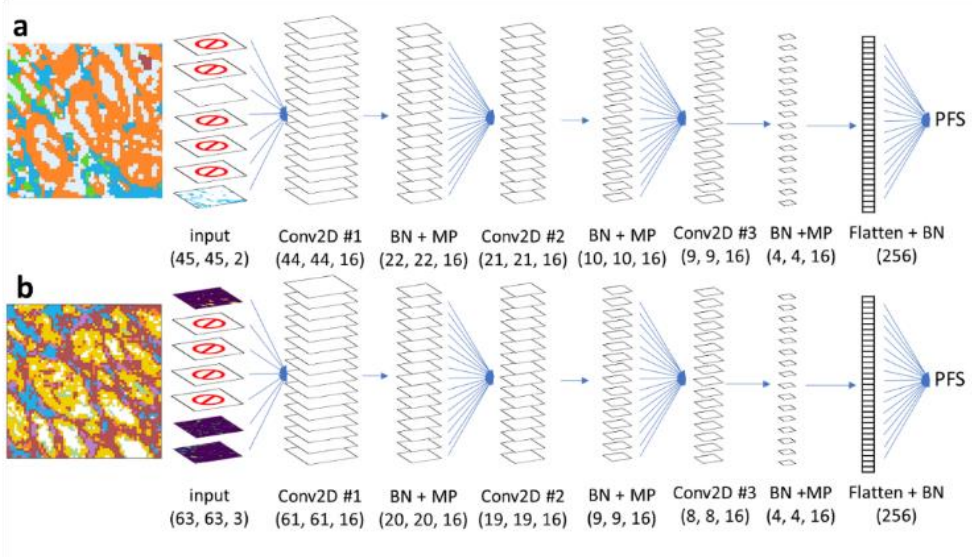

Fig.3 CNN architecture used for HE images analysis (a) and for IR images analysis (b). Abbreviations: 2D convolutional layer (Conv2D), batch normalization (BN), max pool (MP).

Then followed 3 repetitions of i) a 2D convolutional layer with 16 neurons, a kernel of 2x2 and a softplus activation function, ii) a batch normalization[28] and iii) a max pooling layer with a kernel of 2x2. The resulting outputs were flattened, a dropout of 20% was applied during training and a final batch normalization was applied. Finally, a linear regression was performed on the outputs to predict the PFS. Briefly, convolutional layers allow to extract morphological features of the input images by applying convolution filters. Max pooling layers allow both the concentration of information and a change of scale for the search for features. Batch normalization normalizes the output values of the layer, to speed up the learning process. Dropout allows 20% of the neurons to be deactivated (weight =0) randomly at each batch during the training, to make the model less sensitive to noise. For more details on CNN, we refer the reader to work of LeCun et al.[29] In the training process, the entire training set was presented to the network 900 times (number of epochs).

2.8 CNN for IR spectroscopy images analysis

Similar processing was done for the analysis of IR spectroscopy images segmented by k-means clustering. Using a leave-one-patient out validation, with the criterion of the RMSE on PFS, a slightly different CNN was built. The segmented images were cut into squares of 63x63 pixels. Selected areas did not include borders of the cuts (margin of approximately 150 µm), as they were more likely to be damaged during sectioning. For each cut, two squares were extracted. Segmented squared were then reshaped into stacks of binary images, one for each class. K-means clustering allows to identify structures, but the information about variation inside classes is lost. To consider this spectral variability, the k-means binary images were modulated by a score calculated as follows. A principal component analysis (PCA) was calculated class by class, then the score of each pixel according to the first component of its class was extracted. The PCA scores were centered and reduced, then limited to the interval]-0.5, +0.5[to avoid extreme scores. The binary images were modulated by this score by multiplying them by (1-2*scores). Thus, the pixels out of class (0) remain at 0 and the pixels of the class took a value strictly higher than 0 and lower than 2. The intensity of each pixel positioned it within its class. This process was done for all classes except for the class of unclassified pixels. Figure 4 shows an example of the result obtained. The addition of the intra-class spectral variability improved the prediction given by the CNN.

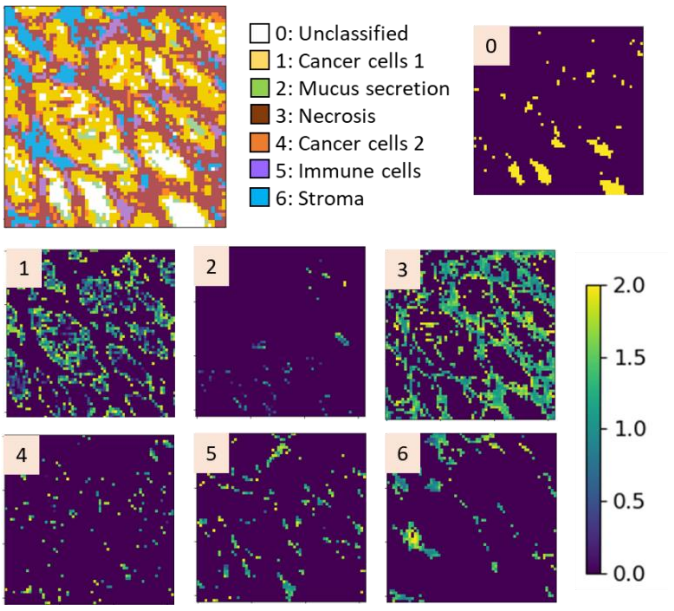

Fig. 4 Example of spectral images obtained from k-means segmentation, then modulation by PCA scores. The central image represents the k-means segmentation, with the corresponding classes label below. Then, for each class (number in the top left corner), a PCA was calculated, and the intensity of pixels was modulated according their PCA score.

Like for HE, classes to be included in the processing step were selected. Better RMSE was obtained when selecting only the class corresponding to unclassified pixels, immune cells and stroma. Finally, a CNN was trained from the squares, with a leave-one-patient out validation scheme, using data augmentation as described above. Architecture of the CNN is presented in figure 3b. The input layer consisted in three (unclassified, immune cells, and stroma) 63x63 images.

Then followed 3 repetitions of i) a 2D convolutional layer with 16 neurons, a kernel of 3x3 then 2x2, and 2x2 and a "softplus" activation function, ii) a batch normalization[28] and iii) a max pooling layer with a kernel of 3x3 then 2x2, and 2x2. The resulting outputs were flattened, a dropout of 20% was applied during training and a final batch normalization was applied. Finally, a linear regression was performed on the outputs to predict the PFS. The CNN was trained for about 530 epochs.

2.9 Fusion of HE and IR CNN

Once the two CNN were trained separately, we fuse them to test whether the combined information of HE staining and IR spectroscopy would improve the prediction of the PFS prediction. The two CNNs were combined at the last layer level (concatenation of flatten layers) allowing the extraction of features to be specific to each modality. The resulting CNN could have been trained from scratch, in the same way as the two separate networks. However, it would have twice as many parameters, without the possibility to increase accordingly the number of images to train on. As an alternative, the already trained HE and IR CNN were used, and only the last layer (linear regression) was trained. This resulted in a faster training (100 epochs) as compared to two previous training from scratch (900 and 530 epochs). The number of squares per patient was set to two for each modality, meaning part of the HE squares were not considered here. Note that HE and IR squares did not necessarily correspond to the same areas of the histological cuts.

3. Results and discussion

3.1 Segmentation of HE images

Visual segmentation in histology requires an anatomical pathologist and is time consuming. To automate this step, we applied a segmentation model, trained to recognize the different tissue structures, previously identified by a pathologist. Figure 5 presents the resulting confusion matrix, and table 2 presents the corresponding overall and class by class accuracies. Segmentation parameters were adjusted so that the average class by class accuracy was maximized, rather than the overall accuracy. This choice aimed at having the best prediction for each class, and not just for the most represented classes.

Many recent studies have shown promising results using automated segmentation in the case of CRC.[30–32] Our goal here was not to develop a competitive classifier for histology recognition, but to have a quick approach for labelling entire surfaces of the HE images, as manual segmentation can be time consuming. For a review of current methods of segmentation of HE stained histological, we refer the reader to Wang et al.[33] and Bankhead et al.[34] The most difficult distinction seemed to be between cancer and cryptic cells. As showed in figure 5 and table 2, we obtained sufficiently satisfactory results to be used in the further analysis.

3.2 Deep learning for PFS prediction from HE and IR images

The CNNs described in figure 3 were applied on HE (figure 3a) and IR images (figure 3b). Using a leave-one-patient out validation, prediction for each patient when leaved out was obtained. Histological sections of each patient were divided into several sub-images, leading to the corresponding number of PFS predictions (see table S1 and S2 in Supporting Information for results detailed for each patient and each section). Figure 6 shows the median (square) and standard deviation (bar) of these values, for HE (a) and IR (b) modalities.

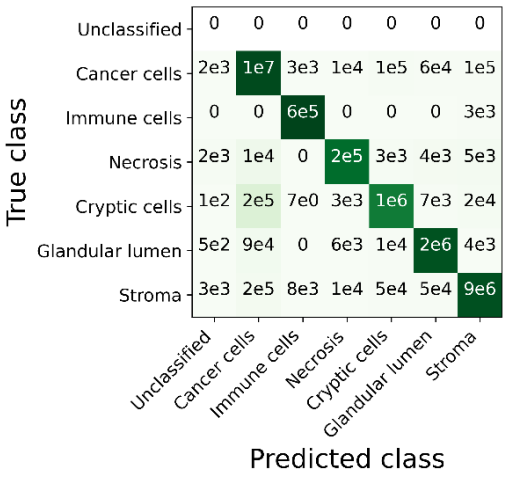

Fig. 5 Confusion matrix of the segmentation method tested on labelled pixels.

Table 2 Overall and class by class accuracies of the segmentation.

Class	Accuracy
Cancerous cells	97.6 %
Immune cells	99.5 %
Necrosis	87.1 %
Non-cancerous cryptic cells	81.2 %
Glandular lumen	95.2 %
Stroma	95.6 %
Mean	92.7 %
Overall	96.0 %

The RMSE obtained over all patients was of 6.6 and 6.3 months (28% and 26% of the average PFS) for standard HE staining and IR imaging respectively. Short PFS are overestimated, while long PFS are underestimated. This bias is known as regression dilution and is due to real PFS measurement uncertainties.[35]

We then applied a CNN taking as input both HE stained and IR spectroscopy images simultaneously (see table S3 in Supporting Information for results detailed for each patient and each cut). Figure 6c shows the median (square) and standard deviation (bar) of these values. The RMSE obtained over all patients was of 5.0 months (21% of the average PFS).

Several tests were carried out on the parameters of the processing sequence. These parameters affect the prediction output, but the values retained also provide information on the characteristics on which the prediction is based.

HE images were down sampled with a factor of 27. This indicates that small details of the cuts are not relevant for the prediction, and thus only the global features of the different classes were taken into account. However, it could come from the segmentation step, which was performed at a limited resolution (26 µm). In that case, small details, such the shape of cancerous cells, could provide some information regarding the state of the cancer. The size of the square sub-images was chosen as the largest possible for both for HE and IR modalities. This suggests that having large images, representative of the whole cut, is more important that having many small images for CNN training.

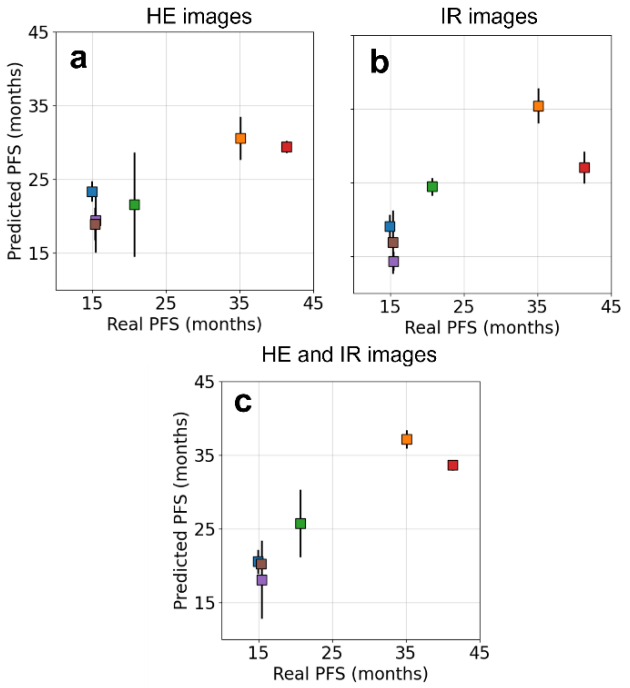

Fig. 6 PFS predicted by the CNN on HE images (a), IR images (b), and both (c) as a function of the real PFS. Error bars correspond to the standard deviation of the predictions made on the different sub-images of a single patient.

When selecting the most relevant classes, immune cells and stroma were selected for both HE and IR modalities. It suggests that CNN correlates the PFS with the presence of immune cells in the stroma. And indeed, a visual examination of the sections allows to observe a correlation between the number of immune cells and the PFS. This is consistent with the literature saying that inflammation of the stroma is linked with better prognosis.[36] Also, larger percentage of stroma in tumor have been showed to be associated with worse prognosis, with higher tumor growth and invasion.[37] Previous work by Kather et al. also underline the importance of the stroma in the prognostic score.[38] Jones et al have also shown, thanks to machine learning algorithms, the importance of the composition of the tumor stroma and inflammation in the evaluation of the risk of recurrence from HE sections.[39] Other studies show the important impact on the prognosis of certain markers related to inflammation[40,41], these markers can also be linked to the TNM score for more precision.[42]

We noted an interesting difference in the segmentation of the 'immune cells' class, between IR and HE images. For the HE images, this class is mainly composed of local clusters of immune cells. But for the k-means images, the class also takes into account areas of inflammatory stroma. This difference could explain why the HE model seems to have an increased variability in its predictions (figure 6a and table S1 and S2). Indeed, the prediction is based on the detection of a small number of clusters, more susceptible to variations from one sub-image to another.

To determine the architecture of CNNs, we started with the standard combination of a convolution layer, followed by batch normalization and a max pooling layer.[20] This combination was repeated a few times to allow working at different scales on the images, using max pooling. As the number of samples was limited, the total number of trainable parameters was limited to a few thousands, by taking a simple 2x2 kernel for the convolutions and adjusting the number of layers to end up with a few hundred features before the final regression. Exact values of parameters were then tuned empirically.

CNN used for the two modalities were close, although different architectures were tested separately. Likewise, the selected classes were close, leading mainly to the same information: the inflammation of the stroma. The fact that the same feature could be found from two different modalities, segmentation and optimization of the processing sequence tends to show that the method is robust. The prediction obtained with the fusion CNN was better, but not by much. This result is coherent with the hypothesis that the two separate methods identify similar features. The increase in precision could also partly come from the two modalities imaging different areas of the histological cuts. As so, the combined information is more representative than for single

modalities. Indeed, the representativeness of the sub-pictures had a strong impact on the RMSE.

While the number of patients is limited, partly due to restricted inclusion conditions to limit the impact of confounding factors, RMSE obtained were relatively low. This number should be compared to the precision of the real PFS measurement. Progression is usually checked upon every two months during a check-up scanner. This means that the confidence interval of real PFS is at least of 2 months. With more patients included, it would be possible to increase the depth of the network, allowing more subtle features to be extracted, and obtain lower errors.

4. Conclusions

As a pilot study, the present results suggest the possibility of anticancer treatment efficacy prediction with an acceptable accuracy. Standard histopathology (HE) and spectral histopathology can extract information correlated with treatment response, including stromal inflammation, from histological sections. The accuracy of the deep learning model is better when based on both modalities jointly. The modulation of k-means clustering by PCA scores allows to take into account spectral variability within the same stromal structure, which improves the accuracy. Other approaches to take advantage of the richness of the spectral information will be tested in further investigations. To obtain more robust results, it would be important to find new samples and confirm these promising results on a larger number of patients. Overall, the present study illustrates the interest of new numerical methods, especially from recent developments in neural networks, in the field of spectroscopy and histology. It also contributes to the improvement of prognostic methods to better personalized therapy for cancer patients.

Author Contributions

Benjamin Brunel: Conceptualization, Methodology, Software, Writing - Original Draft, Writing - Review & Editing, Funding acquisition. Pierre Prada: Software, Investigation, Writing - Original Draft, Writing - Review & Editing. Florian Slimano: Writing - Review & Editing, Supervision. Camille Boulagnon-Rombi: Resources, Writing - Review & Editing. Olivier Bouché: Resources, Writing - Review & Editing, Supervision. Olivier Piot: Conceptualization, Writing - Review & Editing, Supervision, Funding acquisition.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work has been supported by the 'Agence Regionale de Santé (ARS)' for funding the doctoral position of Pierre Prada and by the 'Fondation ARC pour la recherche sur le cancer'. It has been made within the framework of the Graduate School NANO-PHOT (École Universitaire de Recherche, PIA3, contract ANR-18-EURE-0013). The authors thank the Platform of Cellular and Tissular Imaging (PICT) at University of Reims Champagne-Ardenne. We also thank Nicole Bouland for technical support. All the calculations were performed on the French HPC Center ROMEO HPC (https://romeo.univ-reims.fr).

References

1 H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal and F. Bray, *CA. Cancer J. Clin.*, 2021, **71**, 209–249.
2 H. Hurwitz, L. Fehrenbacher, W. Novotny, T. Cartwright, J. Hainsworth, W. Heim, J. Berlin, A. Baron, S. Griffing, E. Holmgren, N. Ferrara, G. Fyfe, B. Rogers, R. Ross and F. Kabbinavar, *N. Engl. J. Med.*, 2004, **350**, 2335–2342.
3 E. Van Cutsem, C.-H. Köhne, E. Hitre, J. Zaluski, C.-R. Chang Chien, A. Makhson, G. D'Haens, T. Pintér, R. Lim, G. Bodoky, J. K. Roh, G. Folprecht, P. Ruff, C. Stroh, S. Tejpar, M. Schlichting, J. Nippgen and P. Rougier, *N. Engl. J. Med.*, 2009, **360**, 1408–1417.
4 J. M. Phelip, D. Tougeron, D. Léonard, L. Benhaim, G. Desolneux, A. Dupré, P. Michel, C. Penna, C. Tournigand, C. Louvet, N. Christou, P. Chevallier, A. Dohan, B. Rousseaux and O. Bouché, *Dig. Liver Dis. Off. J. Ital. Soc. Gastroenterol. Ital. Assoc. Study Liver*, 2019, **51**, 1357–1363.
5 A. Baraniskin, B. Buchberger, C. Pox, U. Graeven, J. W. Holch, W. Schmiegel and V. Heinemann, *Eur. J. Cancer Oxf. Engl. 1990*, 2019, **106**, 37–44.
6 C. Montemagno and G. Pagès, *Front. Cell Dev. Biol.*
7 M. J. Sorich, M. D. Wiese, A. Rowland, G. Kichenadasse, R. A. McKinnon and C. S. Karapetis, *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.*, 2015, **26**, 13–21.
8 P. Wu, W. Gao, M. Su, E. C. Nice, W. Zhang, J. Lin and N. Xie, *Front. Cell Dev. Biol.*
9 J. Nallala, M.-D. Diebold, C. Gobinet, O. Bouché, G. D. Sockalingum, O. Piot and M. Manfait, *The Analyst*, 2014, **139**, 4005–4015.
10 V. Gaydou, M. Polette, C. Gobinet, C. Kileztky, J.-F. Angiboust, P. Birembaut, V. Vuiblet and O. Piot, *Chem. Sci.*, 2019, **10**, 4246–4258.
11 R. Wolthuis, A. Travo, C. Nicolet, A. Neuville, M.-P. Gaub, D. Guenot, E. Ly, M. Manfait, P. Jeannesson and O. Piot, *Anal. Chem.*, 2008, **80**, 8461–8469.
12 A. Travo, O. Piot, R. Wolthuis, C. Gobinet, M. Manfait, J. Bara, M.-E. Forgue-Lafitte and P. Jeannesson, *Histopathology*, 2010, **56**, 921–931.
13 T. N. Q. Nguyen, P. Jeannesson, A. Groh, O. Piot, D. Guenot and C. Gobinet, *J. Biophotonics*, 2016, **9**, 521–532.
14 I. Farah, T. N. Q. Nguyen, A. Groh, D. Guenot, P. Jeannesson and C. Gobinet, *The Analyst*, 2016, **141**, 3296–3304.
15 V. Balan, C.-T. Mihai, F.-D. Cojocaru, C.-M. Uritu, G. Dodi, D. Botezat and I. Gardikiotis, *Materials*, 2019, **12**, 2884.
16 S. Pahlow, K. Weber, J. Popp, B. R. Wood, K. Kochan, A. Rüther, D. Perez-Guaita, P. Heraud, N. Stone, A. Dudgeon, B. Gardner, R. Reddy, D. Mayerich and R. Bhargava, *Appl. Spectrosc.*, 2018, **72**, 52–84.
17 D. Shen, G. Wu and H.-I. Suk, *Annu. Rev. Biomed. Eng.*, 2017, **19**, 221–248.

18 C.-H. Li, D. Cai, M.-E. Zhong, M.-Y. Lv, Z.-P. Huang, Q. Zhu, C. Hu, H. Qi, X. Wu and F. Gao, *Front. Genet.*, 2022, **13**, 880093.
19 J. Van der Laak, G. Litjens and F. Ciompi, *Nat. Med.*, 2021, **27**, 775–784.
20 S. Berisha, M. Lotfollahi, J. Jahanipour, I. Gurcan, M. Walsh, R. Bhargava, H. Van Nguyen and D. Mayerich, *The Analyst*, 2019, **144**, 1642–1653.
21 B. Debus, H. Parastar, P. Harrington and D. Kirsanov, *TrAC Trends Anal. Chem.*, 2021, **145**, 116459.
22 R. Moretto, A. Elliott, D. Rossini, R. Intini, V. Conca, F. Pietrantonio, A. Sartore-Bianchi, C. Antoniotti, C. Rasola, M. Scartozzi, M. Salati, N. Pella, M. A. Calegari, M. Carullo, F. Corti, G. Mauri, M. Fassan, G. Masi, P. Brodskiy, H.-J. Lenz, A. Shields, S. Lonardi, M. Korn and C. Cremolini, *Br. J. Cancer*, 2022, **127**, 957–967.
23 P. Bankhead, M. B. Loughrey, J. A. Fernández, Y. Dombrowski, D. G. McArt, P. D. Dunne, S. McQuaid, R. T. Gray, L. J. Murray, H. G. Coleman, J. A. James, M. Salto-Tellez and P. W. Hamilton, *Sci. Rep.*, 2017, **7**, 16878.
24 M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt and N. E. Thomas, in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2009, pp. 1107–1110.
25 R. R. Picard and R. D. Cook, *J. Am. Stat. Assoc.*, 1984, **79**, 575–583.
26 N. K. Afseth and A. Kohler, *Chemom. Intell. Lab. Syst.*, 2012, **117**, 92–99.
27 T. N. Q. Nguyen, P. Jeannesson, A. Groh, D. Guenot and C. Gobinet, *The Analyst*, 2015, **140**, 2439–2448.
28 S. Ioffe and C. Szegedy, in *Proceedings of the 32nd International Conference on Machine Learning*, PMLR, 2015, pp. 448–456.
29 Y. LeCun, Y. Bengio and G. Hinton, *Nature*, 2015, **521**, 436–444.
30 M. Bilal, S. E. A. Raza, A. Azam, S. Graham, M. Ilyas, I. A. Cree, D. Snead, F. Minhas and N. M. Rajpoot, *Lancet Digit. Health*, 2021, **3**, e763–e772.
31 O. L. Saldanha, P. Quirke, N. P. West, J. A. James, M. B. Loughrey, H. I. Grabsch, M. Salto-Tellez, E. Alwers, D. Cifci, N. Ghaffari Laleh, T. Seibel, R. Gray, G. G. A. Hutchins, H. Brenner, M. van Treeck, T. Yuan, T. J. Brinker, J. Chang-Claude, F. Khader, A. Schuppert, T. Luedde, C. Trautwein, H. S. Muti, S. Foersch, M. Hoffmeister, D. Truhn and J. N. Kather, *Nat. Med.*, 2022, **28**, 1232–1239.
32 P. L. Schrammen, N. Ghaffari Laleh, A. Echle, D. Truhn, V. Schulz, T. J. Brinker, H. Brenner, J. Chang-Claude, E. Alwers, A. Brobeil, M. Kloor, L. R. Heij, D. Jäger, C. Trautwein, H. I. Grabsch, P. Quirke, N. P. West, M. Hoffmeister and J. N. Kather, *J. Pathol.*, 2022, **256**, 50–60.
33 S. Wang, D. M. Yang, R. Rong, X. Zhan and G. Xiao, *Am. J. Pathol.*, 2019, **189**, 1686–1698.
34 P. Bankhead, *J. Pathol.*, 2022, **257**, 391–402.
35 J. A. Hutcheon, A. Chiolero and J. A. Hanley, *BMJ*, 2010, **340**, c2289.
36 K. Jakubowska, W. Kisielewski, L. Kańczuga-Koda, M. Koda and W. Famulski, *Oncol. Lett.*, 2017, **14**, 3869–3877.
37 J. H. Park, C. H. Richards, D. C. McMillan, P. G. Horgan and C. S. D. Roxburgh, *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.*, 2014, **25**, 644–651.
38 J. N. Kather, J. Krisam, P. Charoentong, T. Luedde, E. Herpel, C.-A. Weis, T. Gaiser, A. Marx, N. A. Valous, D. Ferber, L. Jansen, C. C. Reyes-Aldasoro, I. Zörnig, D. Jäger, H. Brenner, J. Chang-Claude, M. Hoffmeister and N. Halama, *PLOS Med.*, 2019, **16**, e1002730.
39 H. J. S. Jones, C. Cunningham, H. A. Askautrud, H. E. Danielsen, D. J. Kerr, E. Domingo, T. Maughan, S. J. Leedham and V. H. Koelzer, *Histopathology*, 2021, **79**, 947–956.
40 T. Yamamoto, K. Kawada and K. Obama, *Int. J. Mol. Sci.*, 2021, **22**, 8002.
41 H. Wang, F. Xu, M. Zhang, J. Liu, F. Wang and Q. Zhao, *DNA Cell Biol.*, 2020, **39**, 1181–1193.
42 X. Bai and L. Feng, *Nutr. Cancer*, 2020, **72**, 1170–1177.