Comprehensive Histopathology and Fast Cancer Imaging in Pancreatic Biopsies: Infrared Imaging with Machine Learning Approach

Danuta Liberda[1#], Paulina Koziol[1,2#], Tomasz P. Wrobel[1*]

[1] Solaris National Synchrotron Radiation Centre, Jagiellonian University, Czerwone Maki 98, 30-392
Krakow, Poland.
[2] Institute of Physics, Jagiellonian University, Lojasiewicza 11, 30-348 Krakow, Poland

*email: tomek.wrobel@uj.edu.pl
# equal contribution

Abstract

Infrared (IR) based histopathology offers a new paradigm in looking at tissues and can provide a complementary information source for more classical histopathology. In this article we report such results for pancreatic cancer histopathology based on IR imaging and machine learning, providing a classification model based on data from over 600 biopsies (coming from 250 patients) imaged with IR diffraction-limited spatial resolution. This forms one of the largest IR datasets analyzed up to now, with almost 700 million spectra of different tissue types. This study forms a steppingstone for further avenues and we also report initial results in two of them. The first is a translation to a fast Quantum Cascade Laser microscope for intraoperative margin detection with a timescale in minutes. The second is subtle pathology grading, placing a biochemically grounded baseline for neoplasia (PanIN1-2) in the benign/cancerous space.

Introduction

Pancreatic ductal adenocarcinoma (PDAC), the most common pancreatic cancer type, accounts for less than 3% of all cancers, but it remains the third leading cause of cancer-related deaths, both for men and women in the United States with 53,000 estimated new cases and an 80% death rate in one year [1]. Furthermore, PDAC is projected to become the second cause of cancer-related death by 2030 [2]. The dreadful prognosis of patients with this disease, including a less than 8% 5-year survival after diagnosis, is due to the lack of early symptoms and/or specific biomarkers for early diagnosis and the paucity of available chemotherapy. In PDAC, three types of biomarkers are desirable: those that help in the detection of the disease onset (diagnosis biomarkers); those that predict responses to treatments (predictive biomarkers) and those that forecast the likely course of the disease, including survival and recurrence patterns (prognosis biomarkers). One of the important issues of effective diagnosis and patient treatment is lesions progression monitoring. In the case of pancreas - pancreatic intraepithelial neoplasia (PanIN) that can progress to PDAC [1]. PanIN can be divided in to three grades (PanIN1-3) that have different ability to be developed in to pancreatic cancer [2].

Infrared (IR) based histopathology offers a new paradigm in looking at tissues and can provide a complementary information source for more classical histopathology [3,4]. Especially, correct understanding of the involved inflammation, fibrosis and neoplasia is required for a method to be successful. IR histopathology is being developed for an increasing

number of cancer types, e.g. breast [5], colon [6], lung [7], brain [8] prostate [9], esophagus [10] - highlighting the increasing potential of the method. The technological developments, especially on the speed and spatial resolution front, make this approach feasible in the clinic [11]. The key to the optimization of IR measurements and reducing analysis time is the spectral information reduction: whether it is selection of specific spectroscopic bands [12] or significant reduction of spectral resolution [13]. Both conditions can be met using Quantum Cascade Lasers (QCL) based microscopes. It has been reported that a QCL data collection needed for classification of a centimeter-scale size biopsies could be done in the order of minutes [13,14] with a commercially available Spero-QT (Daylight Solutions, CA, USA) microscope. However, to optimize a method for a fast screening or imaging modality, a proper foundation needs to be laid beforehand even after initial proof of concept studies [15,16].

In this article we report such results for pancreatic cancer histopathology based on FT-IR imaging and machine learning, providing a classification model based on data from over 600 biopsies (coming from 250 patients) imaged with IR diffraction-limited spatial resolution. This forms one of the largest IR datasets analyzed up to now, with almost 700 million spectra of different tissue types. This study forms a steppingstone for further avenues and we also report initial results in two of them. The first is a translation to a fast Quantum Cascade Laser microscope for intraoperative margin detection with a timescale in minutes. The second is subtle pathology grading, placing a biochemically grounded baseline for neoplasia (PanIN1-2) in the benign/cancerous space.

Results

Patient pathology recognition methodology

This research was performed using Tissue Micro Arrays (TMAs), which are typically applied in studies on tissue composition and differentiation. This is caused by the high variability of TMAs due to many biopsies that are taken from patients differentiated by age, sex, pathology diagnosis, grade. After the typical procedure of needle biopsies collection, to prevent their damage they are embedded with paraffin and cut into thin 5 µm slices (Figure 1 A). In this research, we used two neighboring slices. Biopsy pathology recognition needs to be done based on histopathological annotations performed by an experienced histopathologist (Figure 1B). Therefore, the first TMA slice is H&E stained and imaged with a visible light microscope. The second slice is imaged with an IR spectrometer. The result of this measurement is data in the form of a cube containing IR images in the X and Y direction, and spectra in the Z direction (Figure 1C). Based on histopathological annotation mentioned before corresponding regions are marked on IR image – spectra that come from this area (after preprocessing) are used for final histopathological model creation. This model in the next step is applied for biopsy IR image digital histopathological description. Patient pathology is finally recognized using Receiver Operating Characteristic (ROC) curves, based on the number of pixels (from a chosen class/classes) present in predicted images (Cancer class in the Patient pathology classification panel in Figure 1D). Model prediction power is evaluated using Area Under the Curve (AUC).

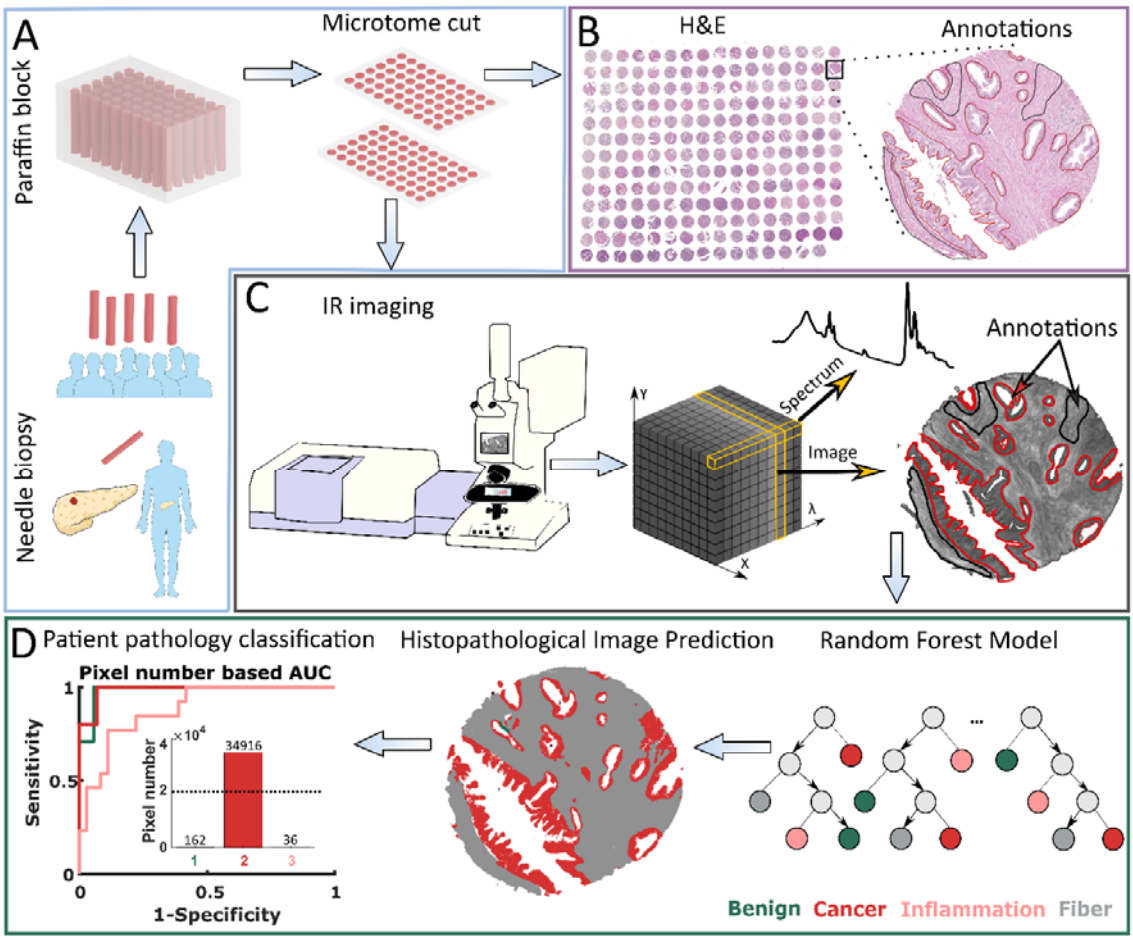

Figure 1. Patient pathology recognition pipeline: A) Tissue Micro Array preparation, B) Histopathological annotation on H&E stained biopsies images, C) Biopsies' Infrared Imaging and obtained data structure, D) Model creation and patient pathology recognition.

Comprehensive histopathological model

The FT-IR histopathological Random Forest model was based on 6 TMAs (more information in Supplementary Materials Table 1). Our model can differentiate six classes of pancreatic tissue: Benign, Cancer, Necrosis, Inflammation, Fiber, and Blood. Differences between analyzed classes are to some extent reflected in their mean spectra (Figure 2B). At the first glance, there are visible differences in shape and relative intensity of Amide A and B (3550 – 3000 cm^{-1}) protein bands and DNA/RNA bands (1300 – 1000 cm^{-1}). Patients in TMAs were divided into model, validation, and test sets (Supplementary Materials Table 4) for model optimization and prediction power evaluation. In Figure 2A predicted TMAs (for Standard Definition) images with zooms on single cores and regions of interest are presented. Benign biopsies with green color can be easily differentiated from remaining cases in all TMAs. Class benign can be described as a combination of lobules and islands annotations, coming from normal pancreas tissue (for example single core from PA1002a TMA presented in Figure 2A), and normal adjacent to cancer tissue. The microenvironment of cancer is stroma-rich, which is visible among others in the single core from PA2072a TMA presented in Figure 2A. In both presented cancerous cores in the upper part of Figure 2A there can be also found necrotic regions associated with cancer changed biopsy tissue. Frequently, we observed lymphocyte

clusters like in the single biopsy in PA961e TMA presented in Figure 2A, infiltrating into the cancerous tissue surrounded by stroma. Most chronic pancreatitis cases are visible in BBS14011 TMA. In the single biopsy predicted image, one may observe both: lymphocyte infiltration into the tissue and inflamed changed lobules. In the single core in BIC14011a TMA in Figure 2A, there is visible very well classified aggregation of erythrocytes in the affected stroma. The classification results on the pixel level (tissue level) were quantified using Area Under the Curve (AUC) metric in Figure 2C. All classes reached AUC values higher than 0.95, which is a very good result taking into account that AUC values equal to one corresponds to the situation where each pixel was correctly classified. We also calculated AUC metrics on the core and patient level for test set (49 patients) giving results 0.99 (Supplementary materials Figure S4). A similar model based on the same histopathological annotations and TMAs was built for High Definition data. In comparison to Standard Definition, this model unraveled small cells (i.e. lymphocytes) located between tissue. In general, predicted classes are not consistent like in Standard Definition, there are cases with pixels that are highly mixed (Supplementary Materials S3A). This is reflected in AUC values – there are three classes with results below 0.95 (Supplementary Materials S3C). However, this prediction still gives very good classification results on the patient level (Supplementary Materials S4B). Finally, we used intraoperative biopsy (taken in clinical conditions) as the external sample for Standard Definition model (Figure 2D). The most valuable result achieved for this sample is a very well classified transition from benign to cancerous state with inflammation infiltration. Zoom on interesting intraoperative biopsy regions can be found in Supplementary Materials S2. In predicted image in Figure 2D intraoperative margin can be easily marked.

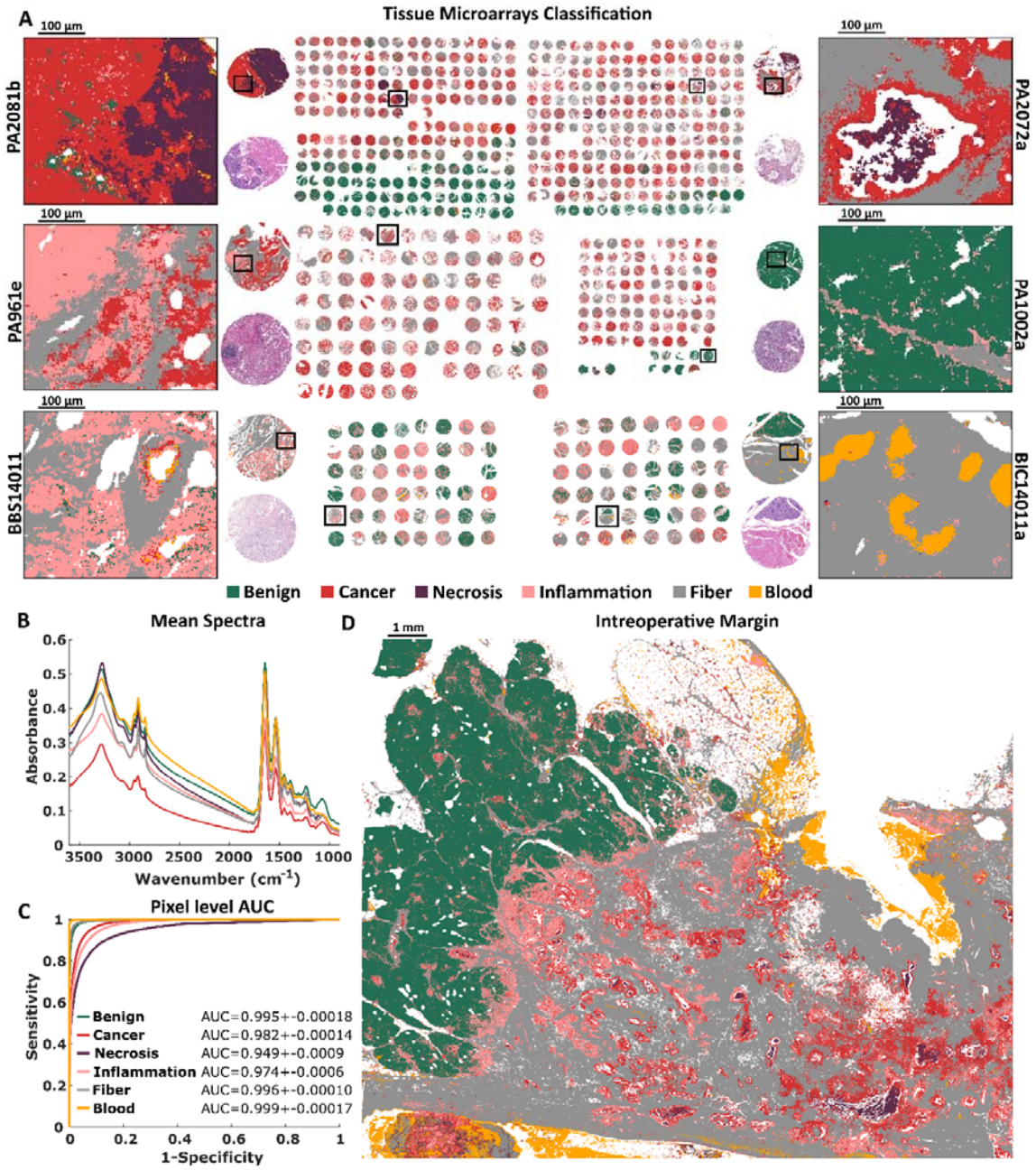

Figure 2. Histopathological recognition of six pancreatic tissue classes with Random Forest classification for: A) Tissue Micro Arrays with zooms on single cores with chosen regions, and D) Intraoperative margin. B) Tissue Micro Arrays mean spectra, and C) Pixel level AUC values of recognized tissue classes.

Finally, we decided to evaluate duct character in BIC14011a TMA containing cases of normal duct, Pancreatic Intraepithelial Neoplasia (PanIN) with grades between 1 and 2, and duct adenocarcinoma. On the grounds that Partial Least Squares Regression (PLSR) gives a possibility to predict value of dependent variable Y (duct character from -1 Benign up to 1 Cancer), we decided to apply it to build a two-class FT-IR regression model predicting PanIN ducts character. Prediction results closer to 1 indicate more advance disease progression. In

Figure 3A we presented the distribution of Y predicted for spectra coming from ducts annotated by histopathologists as Benign (green), PanIN (blue), and Cancer (red). Y predicted distribution of PanIN class is closer to Benign distribution showing their prevailing similarity, but there are also PanIN cases covering Y predicted space between 0 and 1. This trend is presented in Figure 4B where predicted duct character was plotted on biopsies IR images starting from the most cancerous biopsy F2 ending with the most normal-like character duct A4 (sorted based on distribution median value). Biopsies from F2 up to F3 have highly cancerous character, from C4 up to C2 they can be defined as a cancerous character, from D5 up to E4 duct character is between benign and cancer, and in the case of biopsies from B8 to A4 duct character goes toward benign.

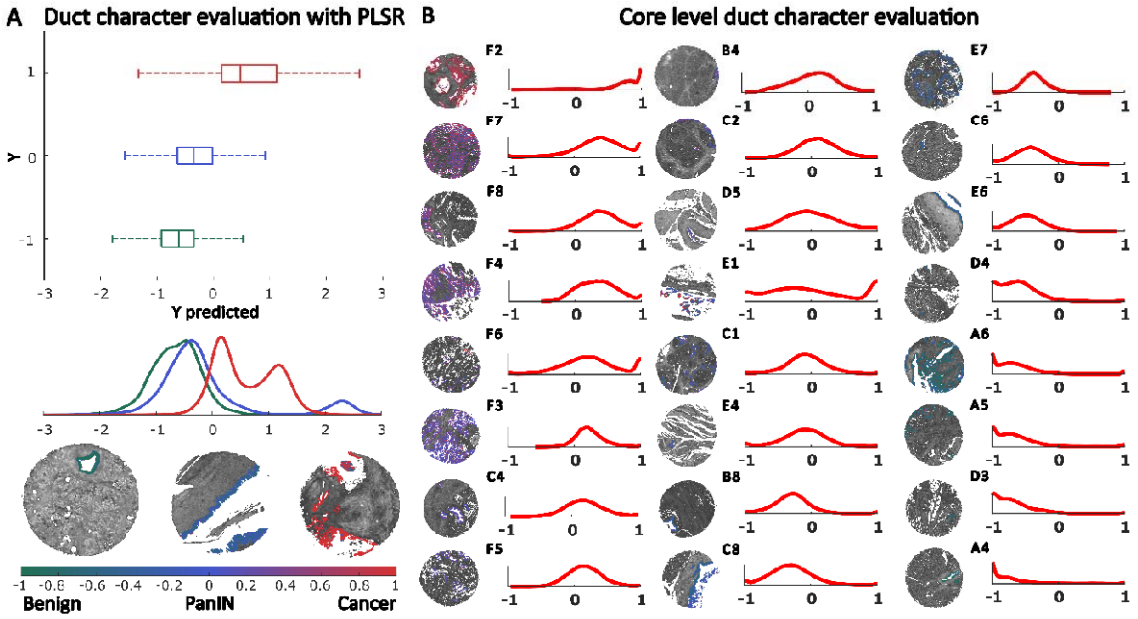

Figure 3. Duct histopathological character evaluation with two class (Benign, Cancer) Partial Least Squares Regression FT-IR model: A) Observed dependent variable (Y) versus its predicted values (Y predicted) for three classes: Benign (color close to green, -1), PanIN (color between turquoise and violet), and Cancer (color close to red, 1), with density plots. B) Core level duct character evaluation with density plots based on duct pixels values (from 1 to -1), sorted in columns from the most cancerous to the most benign biopsies.

Fast pathology recognition

To meet the clinic requirements a method needs to be rapid. Therefore, we created a Random Forest model based on data collected with a very fast system equipped with QCL Laser. Usage of those lasers allows for chosen frequencies measurement (Discrete frequency imaging). As a first step, we created spectra based three-class model with two TMAs (BIC14011a and PA2081b), that is relatively simple in comparison to presented in previous section six-class model. However, it allows for an easier definition of pathologically changed tissue regions with very high AUC values above 0.99 (for 109 patients) on the patient level. Numbers of patients divided into validation sets can be found in Supplementary Materials Table 5. The patient classification was based on the ratios of classes Cancer and Benign pixels number thresholding. In Figure 4C four exemplary biopsies coming from BIC14011a and

PA2081b TMAs are shown. For the majority of biopsies in both TMAs (Supplementary Materials S6) prediction quality is similar to the FT-IR model (Figure 2). However, there are also cases like biopsy in the right bottom corner of Figure 4C panel where cancerous regions are little mixed with benign pixels. In the final second step, we test fast intraoperative biopsy prediction model, based on two TMAs mentioned above and only 8 frequencies providing chemical contrast (additional 5 for baseline and 1 for normalization). In other words, we optimized the model choosing the smallest possible frequencies number allowing for relevant model performance. As an external test, we used a biopsy sample taken in a clinical environment. Mean spectra of analyzed tissue classes with frequencies chosen for fast model predicting intraoperative biopsy are presented in Figure 4A. In the Figure 4D, we presented an IR image at 1650 cm^{-1} intensity (maximum of protein chain builder Amide I band) of this intraoperative biopsy. As mentioned earlier the most important for clinical application is the time factor. The QCL model built in this research can preprocess and predict intraoperative biopsy in time below two minutes. Taking into account that measurement time of 14 frequencies can be less than five minutes it gives the total time needed for digital staining below seven minutes. The prediction results allow for visible at first glance differentiation of benign and cancerous changed tissue regions (Figure 4E). Cancer cells are located between dense stroma. In the benign biopsy region, there are visible fibers surrounding ducts.

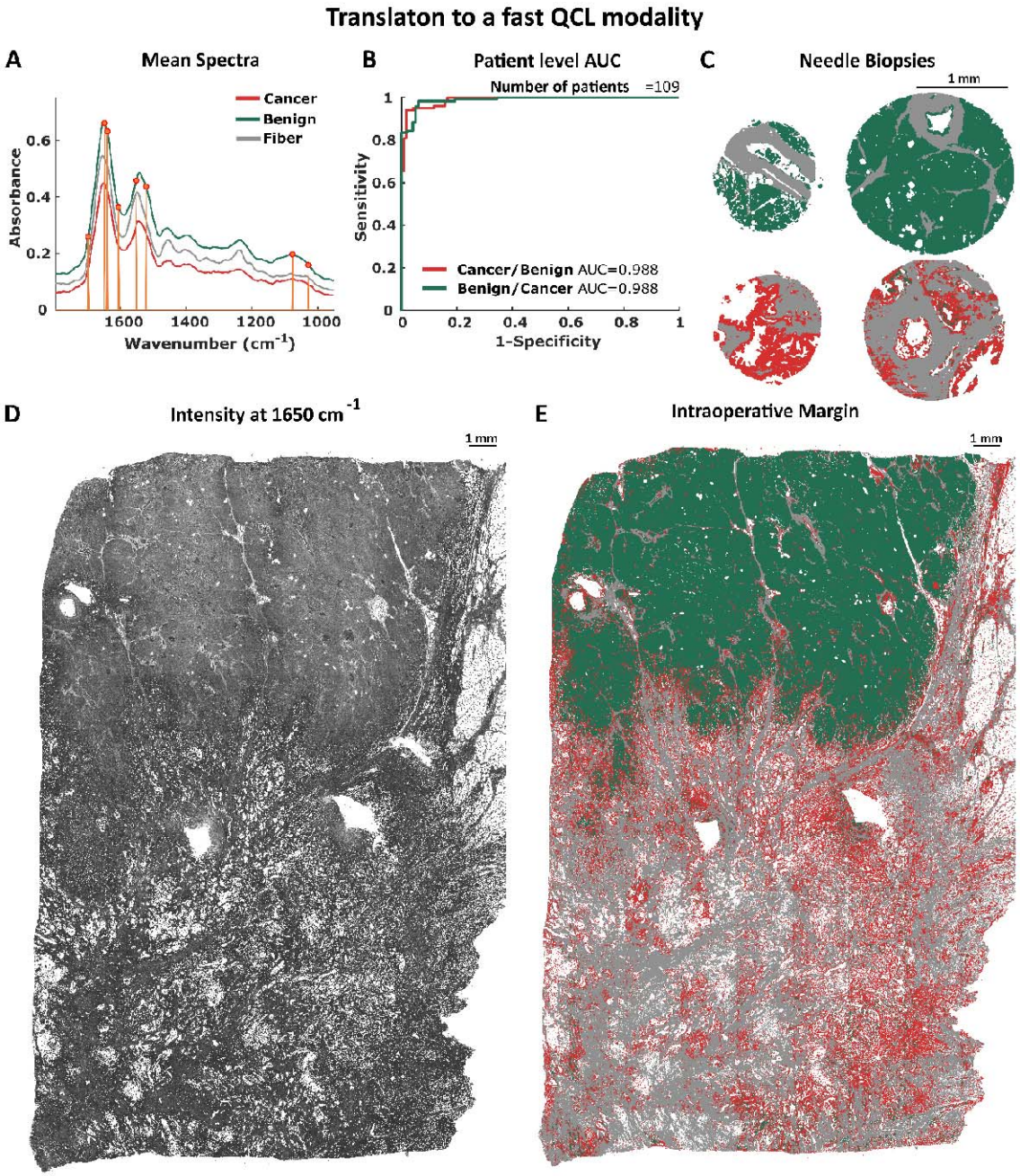

Figure 4. Model translation from FT-IR to a fast QCL modality with the aim of fast pathology recognition and intraoperative margin determination. A) Mean spectra for three tissue classes: Benign, Cancer, Fibers; and spectral frequencies (orange dots) used in model creation for fast intraoperative margin prediction. B) Patient level AUC values calculated based on the number of pixels in histopathological significant classes: Benign and Cancer. C) Predicted needle biopsies. D) Intraoperative biopsy image at 1650 cm^{-1} frequency. E) Intraoperative biopsy prediction.

Discussion

The infrared imaging combined with the machine learning algorithm was applied to develop a model predicting comprehensive pancreatic histopathology and fast cancer detection. In the first step FT-IR Random Forest model was built based on a broad and reach in biochemical information spectra. This model has enabled recognition of six tissue classes: Benign, Cancer, Necrosis, Inflammation, Fiber, and Blood. Automated pathology recognition that can be performed based on this model allows differentiation of cancerous, benign, and chronic pancreatitis cases with sensitivity higher that 95%. Moreover, in the event of a nonobvious case, an image of a predicted biopsy can be analyzed further to diagnose a patient finding specific regions of affected tissue, i.e. immune cells and erythrocyte infiltration, fibrosis, changes in a number of characteristic tissue structures. This is an important aspect from a diagnostic point of view to categorize disease progression, especially with PanIN grading that can lead to aggressive duct adenocarcinoma. Therefore, we successfully developed the PLS model to provide a tool for duct character evaluation. Taking into consideration the requirements of the clinics and analytical laboratories, we decided to translate the FT-IR model to the QCL modality to speed up the measurement time factor. The opportunity to measure only chosen frequencies influence also time consumption by data preprocessing and prediction steps, because of a smaller volume of data that needs to be processed. The created model can be used for fast patients screening but also intraoperative biopsy prediction with the aim of surgical margin annotation. Prediction results on the pixel level are slightly worse than in the case of the FT-IR model, but it does not significantly influence classification performance on the patient level. It is a compromise between the speed of QCL modalities and highly biochemically informative FT-IR modalities. In this research, we optimized measurement time and developed a model that can perform intraoperative biopsy digital staining (including measurement, data preprocessing, and prediction) in a total time less than seven minutes, which is competitive compared to the typical clinical diagnostic approach.

Methods

All calculations described in this study were done in MATLAB software.

Biological material

Human pancreatic tissue was used in this study, with sample set consisting of six Tissue Microarrays (TMAs): PA2081b, PA2072a, PA1002a, PA961e, BBS14011, BIC14011a purchased from Biomax Inc. and two surgical resections. TMAs are assembled from tissue cores coming from needle biopsies with diameters between 1-1.5 mm (depending on the TMA). Resection material was specifically selected to present boarded between cancer and control tissue (surgical margin), therefore sample sizes are roughly 16x35 mm and 16x13 mm. Samples were placed in a paraffin block and cut using microtome for 6 μm thick slices, with one slice placed on IR transparent BaF_2 salt plate for transmission measurement, and consecutive slice mounted on glass for Hematoxylin and Eosin (H&E) staining allowing histopathological annotations. Initially paraffin covered IR samples were deparaffinized with 24 hours hexan bath to avoid paraffin absorption around 1462 cm^{-1}.

High patient statistic was achieved by the use of TMAs with 663 tissue cores in total. However, in most cases, more than one tissue core came from a single patient. Additionally, overlapping cases existed between some TMAs, thus, final patient cohort comprised of 250

cases (with two additional patients represented by surgical resection). Ethical approval was granted by Ethics committee at Jagiellonian University in Krakow (no. 1072.6120.304.2020).

Fourier Transform Infrared Imaging

Bruker Vertex 70v spectrometer coupled with Hyperion 3000 microscope and 64x64 FPA MCT detector was used for FT-IR measurements. Spectra were measured within 3850-900 cm^{-1} range, with 8 cm^{-1} spectral resolution and zero filling factor of 1, giving 765 spectral points. Signal was co-averaged 4 times for a sample region and 64 times for background collection. Two types of objectives were used during measurements – 15x (SD) and 36x (HD), giving projected pixel sizes of 2.7 μm and 1.1 μm, respectively. All TMAs were measures with both resolutions, but surgical resections samples were only measured using 15x objective. Total number of sample spectra collected with FT-IR imaging in this study reached 672 407 632 (including both resolutions).

The first step of preprocessing was data denoising using Minimum Noise Fraction (MNF) method with 20 bands used for reconstruction [17,18]. Due to partially spatial character of MNF's noise estimation and RAM memory constrains, TMAs processing was done in a core-by-core manner. In case of resection material, denoising was done for region of 2x2 measurement tiles. Such approach allows to prevent information leakage between patients, which could cause over optimistic results in further classification. For the classification purpose, FT-IR spectra were transferred into so-called metrics. Metrics provide spectral information extracted from selected IR bands defined by experienced spectroscopist (list of spectral regions presented in Table S2 in Supplementary Materials). Metrics calculation for each spectral region includes local linear baseline removal followed by calculation of band integration, maximum value and center of gravity. Finally, those three characteristics are normalized to the corresponding value of Amide I. Total number of metrics extracted from single spectrum is 123.

Quantum Cascade Laser Imaging

Measurements with QCL Microscope were performed by DRS Daylight Solutions, CA, USA, using Spero QT spectrometer. Imaging modality was provided by 480x480 Microbolometer FPA, with 4x objective, leading to 4.25 μm projected pixel size. Images were collected in transmission mode with 2 cm^{-1} spectral resolution and 1794-952 cm^{-1} range, resulting in 422 spectral variables. Two full TMAs and one resection material were measured with this modality, giving 31 615 960 sample spectra.

Two types of preprocessing approaches were applied to QCL datasets to extract biochemical information for the classification purpose. In the first case, for the means of model translation, data were denoised with MNF using 20 bands for reconstruction [19] and set of 81 metrics was created using spectral ranges defined in Table S3 in Supplementary Materials, similarly to FT-IR. For the second approach, data preprocessing steps were specifically arranged to allow spectral information reduction favoring fast surgical margin detection. Therefore, spectral denoising step was in this case omitted and rubber baseline correction was applied with segments between following wavenumbers: 1794, 1752, 1592, 1478, 1134, and 952 cm^{-1}. All variables were normalized to 1650 cm^{-1} intensity (Amide I), giving 422 possible features to feed the classifier. Finally, predicted image of the intraoperative biopsy were smoothed using a 2D majority filter with a 5 by 5 pixels window size.

Random Forest Classification

Classification model creation requires multiple steps, starting from tissue types annotation and Region of Interests (ROIs) definition for chosen classes, which are required to extract data later used to feed the classifier. Annotations of Infrared images for all TMAs were done utilizing ground truth annotations provided by an experienced Histopathologist, based on visual inspection of H&E images. Such annotations were later transformed into ROI masks, allowing correct pixels (metrics/spectra) selection. Furthermore, to assure classification using only tissue originating pixels, a tissue mask was created after thresholding absorbance for 1650 cm^{-1} (most intensive band). Such a mask was used to remove background spectra from ROI masks and to select samples for final prediction images.

To prevent from Random Forest classifier being biased, all classes had the same number of samples (pixels) during a particular model training process. This number was determined by the number of samples available for the smallest class. Moreover, to ensure training data variability, equal number of samples from each TMA was taken, unless the number of available samples was insufficient. In such case, lacking samples were evenly topped up from other TMAs. In all cases, samples were always randomly selected.

FT-IR based six-class models (SD and HD) were built with 50 Trees, which was sufficient to prevent overtraining, but at the same time did not unnecessarily complicate the model. The same number of trees was kept for model's translation to QCL modality, despite the fact that number of classes was reduced to three. However, fast margin classifier training procedure was more extended that previous ones and some steps need clarification. Initial model was built with 50 trees and used chemical information in the form of full spectra (not metrics, as described in the previous section). One of the biggest advantages of RF is its ability to provide information about Importance of each spectral variable. Variables with highest importance and lowest coherence artifacts (decision based on visual inspection) were used to train model for fast margin detection, using only 8 variables providing chemical contrast (additional 5 for baseline and 1 for normalization). This model was equally sensitive with smaller number of trees, thus it was reduced to 20 trees to shorten the prediction time.

A crucial factor for stable model creation is its proper validation, assuring its stability and lack of overtraining. For FT-IR based models, patients were divided according to their annotations (benign, inflammation, cancer) in a random manner into model, validation and test sets. Subset of patients from each TMA was assigned to a test set, keeping in mind overlapping cases between TMAs. Number of patients in validation and test sets are presented in Supplementary Materials Table 4. A four-fold cross validation (using model data) was used during the model optimization, which is a long and iterative process. To assess models' performance during the cross validation, confusion matrix along with Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC) were calculated. Finally, when created model reached satisfactory parameters, a test set was used for its final evaluation with ROC and AUC defining final diagnostic ability. Described validation approach was applied to both, SD and HD datasets. Nonetheless, adopted validation method is strongly dependent from available patient cohort. Thus, in case of QCL based models with only two TMAs available, five-fold cross validation (Supplementary Material Table 5) was applied with confusion matrix, ROC and AUC determining models performance.

Partial Least Squares Regression

Partial Least Squares Regression (PLSR) was used for the evaluation of duct character in BIC14011a TMA imaged with FT-IR. It consists of 48 cores and 24 cases. Data preprocessing (denoising, baseline correction, normalization) was the same as in the case of FT-IR Random Forest model creation. Data standardization preprocessing step (logarithmic transformation) was added to this pipeline, because of the dependence of variance from metrics values. Leave-one-patient-out cross-validation was applied for the best model performance. Two-class (benign and cancerous duct) models were built using 4 Latent Variables (LVs). The LVs number was chosen through evaluation of mean Root Mean Square Error of Cross-Validation (RMSECV) calculated for model sets (Supplementary Materials S5B). Classes were coded as follows: benign -1, cancer 1. To evaluate the character of PanIN – duct neoplasia – we created a two-class model.

Acknowledgements

This research was supported by „Pancreatic cancer comprehensive histopathology based on IR chemical imaging" project, which was carried out within the Homing programme (grant no. Homing/2016-2/20) of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund. This research was performed using equipment purchased in the frame of the project co-funded by the Malopolska Regional Operational Program Measure 5.1 Krakow Metropolitan Area as an important hub of the European Research Area for 2007-2013, project no. MRPO.05.01.00-12-013/15.The authors declare no conflicts of interest.

References

[1] R.H. Hruban, N. Fukushima, Pancreatic adenocarcinoma: Update on the surgical pathology of carcinomas of ductal origin and panins, Mod. Pathol. 20 (2007) 61–70. https://doi.org/10.1038/modpathol.3800685.

[2] A. Andea, F. Sarkar, V.N. Adsay, Clinicopathological Correlates of Pancreatic Intraepithelial Neoplasia: A Comparative Analysis of 82 Cases With and 152 Cases Without Pancreatic Ductal Adenocarcinoma, Mod. Pathol. 16 (2003) 996–1006. https://doi.org/10.1097/01.MP.0000087422.24733.62.

[3] T.P. Wrobel, R. Bhargava, Infrared Spectroscopic Imaging Advances as an Analytical Technology for Biomedical Sciences, Anal. Chem. 90 (2018) 1444–1463. https://doi.org/10.1021/acs.analchem.7b05330.

[4] M.J. Baker, H.J. Byrne, J. Chalmers, P. Gardner, R. Goodacre, A. Henderson, S.G. Kazarian, F.L. Martin, J. Moger, N. Stone, J. Sulé-Suso, Clinical applications of infrared and Raman spectroscopy: State of play and future challenges, Analyst. 143 (2018) 1735–1757. https://doi.org/10.1039/c7an01871a.

[5] S. Mittal, T.P. Wrobel, M. Walsh, A. Kajdacsy-Balla, R. Bhargava, Breast cancer histopathology using infrared spectroscopic imaging: The impact of instrumental configurations, Clin. Spectrosc. 3 (2021) 100006. https://doi.org/10.1016/j.clispe.2021.100006.

[6] S. Tiwari, A. Kajdacsy-Balla, J. Whiteley, G. Cheng, S.M. Hewitt, R. Bhargava, INFORM: INFrared-based ORganizational Measurements of tumor and its microenvironment to predict patient survival, Sci. Adv. 7 (2021). https://doi.org/10.1126/sciadv.abb8292.

[7] F. Großerueschkamp, A. Kallenbach-Thieltges, T. Behrens, T. Brüning, M. Altmayer, G. Stamatis, D. Theegarten, K. Gerwert, Marker-free automated histopathological annotation of lung tumour subtypes by FTIR imaging, Analyst. 140 (2015) 2114–2120. https://doi.org/10.1039/c4an01978d.

[8] J.R. Hands, P. Abel, K. Ashton, T. Dawson, C. Davis, R.W. Lea, A.J.S. McIntosh, M.J. Baker, Investigating the rapid diagnosis of gliomas from serum samples using infrared spectroscopy and cytokine and angiogenesis factors, Anal. Bioanal. Chem. 405 (2013) 7347–7355. https://doi.org/10.1007/s00216-013-7163-z.

[9] J.T. Kwak, A. Kajdacsy-Balla, V. Macias, M. Walsh, S. Sinha, R. Bhargava, Improving prediction of prostate cancer recurrence using chemical imaging, Sci. Rep. 5 (2015) 1–10. https://doi.org/10.1038/srep08758.

[10] D. Liberda, M. Hermes, P. Koziol, N. Stone, T.P. Wrobel, Translation of an esophagus histopathological FT-IR imaging model to a fast quantum cascade laser modality, J. Biophotonics. 13 (2020) 1–9. https://doi.org/10.1002/jbio.202000122.

[11] M. Schnell, S. Mittal, K. Falahkheirkhah, A. Mittal, K. Yeh, S. Kenkel, A. Kajdacsy-Balla, P.S. Carney, R. Bhargava, All-digital histopathology by infrared-optical hybrid microscopy, Proc. Natl. Acad. Sci. U. S. A. 117 (2020) 3388–3396. https://doi.org/10.1073/pnas.1912400117.

[12] S. Mittal, K. Yeh, L. Suzanne Leslie, S. Kenkel, A. Kajdacsy-Balla, R. Bhargava, Simultaneous cancer and tumor microenvironment subtyping using confocal infrared microscopy for all-digital molecular histopathology, Proc. Natl. Acad. Sci. U. S. A. 115 (2018) E5651–E5660. https://doi.org/10.1073/pnas.1719551115.

[13] C. Kuepper, A. Kallenbach-Thieltges, H. Juette, A. Tannapfel, F. Großerueschkamp, K. Gerwert, Quantum Cascade Laser-Based Infrared Microscopy for Label-Free and Automated Cancer Classification in Tissue Sections, Sci. Rep. 8 (2018) 1–10. https://doi.org/10.1038/s41598-018-26098-w.

[14] N. Goertzen, R. Pappesch, J. Fassunke, T. Brüning, Y.D. Ko, J. Schmidt, F. Großerueschkamp, R. Buettner, K. Gerwert, Quantum Cascade Laser-Based Infrared Imaging as a Label-Free and Automated Approach to Determine Mutations in Lung Adenocarcinoma, Am. J. Pathol. 191 (2021) 1269–1280. https://doi.org/10.1016/j.ajpath.2021.04.013.

[15] K. Szymoński, E. Lipiec, K. Sofińska, K. Skirlińska-Nosek, K. Milian-Ciesielska, J. Szpor, M. Czaja, S. Seweryn, N. Wilkosz, G. Birarda, F. Piccirilli, L. Vaccari, M. Szymoński, Spectroscopic screening of pancreatic cancer, Clin. Spectrosc. 3 (2021) 100016. https://doi.org/10.1016/j.clispe.2021.100016.

[16] V. Notarstefano, S. Sabbatini, C. Conti, M. Pisani, P. Astolfi, C. Pro, C. Rubini, L. Vaccari, E. Giorgini, Investigation of human pancreatic cancer tissues by Fourier Transform Infrared Hyperspectral Imaging, J. Biophotonics. 13 (2020) 1–10. https://doi.org/10.1002/jbio.201960071.

[17] M.K. Raczkowska, P. Koziol, S. Urbaniak, C. Paluszkiewicz, W.M. Kwiatek, T.P. Wrobel, Influence of denoising on classification results in the context of hyperspectral data□: High Definition FT-IR imaging, Anal. Chim. Acta. 1085 (2019) 39–47.

[18] P. Koziol, M.K. Raczkowska, J. Skibinska, S. Urbaniak-Wasik, C. Paluszkiewicz, W. Kwiatek, T.P. Wrobel, Comparison of spectral and spatial denoising techniques in the

context of High Definition FT-IR imaging hyperspectral data, Sci. Rep. 8 (2018) 1–11. https://doi.org/10.1038/s41598-018-32713-7.

[19] P. Koziol, M.K. Raczkowska, J. Skibinska, N.J. McCollum, S. Urbaniak-Wasik, C. Paluszkiewicz, W.M. Kwiatek, T.P. Wrobel, Denoising influence on discrete frequency classification results for quantum cascade laser based infrared microscopy, Anal. Chim. Acta. (2018). https://doi.org/10.1016/j.aca.2018.11.032.