Abstract

This doctoral dissertation was aimed at the development and optimization of a methodology for infrared (IR) spectroscopic imaging, with the goal of achieving maximum data quality and creating models for histopathological classification of pancreatic cancer. Moreover, the research was focused on the development of techniques for the macromolecular orientation determination, which is related to the hypothesis stating, that they can provide additional information about the disease and improve classification models. Both of these issues are based on the use of infrared spectroscopy, which provides a wealth of information about the biochemical composition of the analyzed sample. What is crucial, the implementation of the aforementioned aspects is supported by thoroughly optimized data preprocessing procedures, with a focus on denoising and baseline correction methods.

In relation to the objectives outlined above, the thesis entitled "Infrared Imaging in histopathological classification and macromolecular orientation determination, supported by optimized data preprocessing methods" can be divided into four thematic areas, composed of publication sets. The first area is related to the optimization of spectral and spatial denoising techniques, to improve the quality of the signal and the effectiveness of histopathological classification. The following area focuses on histopathological classification and its effectiveness in diagnosing pancreatic tumors using IR imaging. The third research area is the determination of the in-plane (2D) and in-space (3D) macromolecular orientation using polarized infrared spectroscopy, with the focus pointed at the orientation of collagen fibers. The final topic is the implementation of the EMSC (Extended Multiplicative Signal Correction) baseline correction method for samples with cylindrical domains (e.g. fibers), which exhibit strong Mie-type scattering.

The theoretical part of the thesis covers the physical basis of the aspects related to the research, described in the publications included in the work. The chapter begins with a description of IR spectroscopy, focusing on the fundamentals of vibrational spectroscopy and quantitative analysis. The transmission Fourier Transform Infrared Spectroscopy (FT-IR) setup was characterized, including the optical system and image formation aspects. Two additional far-field IR spectroscopic methods are also introduced, namely Quantum Cascade Laser (QCL) microscopy and Optical-PhotoThermal Infrared (O-PTIR). The following part focuses on specific data preprocessing methods, including spectral and spatial denoising techniques. The group of denoising techniques includes simple filters (mean, median, etc.), low-pass frequency filter, Savitzky-Golay filter, Wavelets, Principal Component Analysis (PCA), and the Minimum Noise Fraction (MNF) method. The aspects of baseline correction, resulting from scattering effects dominant in the transmission measurement mode, are then discussed. The main attention is focused on EMSC-type algorithms, utilizing the full Mie-type scattering theory and the van de Hulst approximation. The EMSC algorithm for scattering correction in samples with cylindrical domains is described in detail. The third subchapter of the theoretical introduction describes the use of polarized IR spectroscopy for the determination of the transition dipole moments orientation. The 2D method provides limited information about the projection of the dipole moment onto the plane, whereas the 3D method utilizes two dipole moments simultaneously and provides spatial information about their orientation. By analyzing

the directions of vibrations, the orientation of transition dipole moments is used to determine the orientation of macromolecular chains. The chapter concludes with a description of the Random Forest classification method, which was used in pancreatic cancer histopathology research.

The key part of this dissertation is the chapter featuring publication reprints. It groups the papers, that form the fundaments of this thesis, into the four thematic threads mentioned above. Each of the topics has a brief description of the context and motivation for the conducted research, as well as the crucial results and conclusions. The first and largest set of publications is focused on the influence of denoising on spectroscopic data and the histopathological classification results. A range of spectral and spatial denoising methods were thoroughly optimized across a wide range of parameters. In the case of signal quality improvement, simulated data sets with varying levels of noise and spatial resolution were used. This allowed for estimation the effectiveness of the methods, in terms of improving the signal-to-noise ratio (SNR) and minimizing the signal distortion introduced by denoising. The obtained results revealed the power of MNF and PCA methods, highlighting the potential for reduction of measurement time. Furthermore, scripts for generating simulated data, along with the data itself, were published as open access. In addition, denoising was optimized to improve the Random Forests classification accuracy. It was shown that applying PCA or MNF to FT-IR data increased pixel-level classification sensitivity by up to 43-44%. For classification based on data from QCL microscopy, the application of MNF demonstrated the potential for the reduction of spectral information to as few as 20 variables (bands). Such denoising optimizations had a crucial impact on the classification described further.

Pancreatic histopathology was the next research area, with the aim to create models allowing cancer detection based on FT-IR spectroscopy and machine learning methods (Random Forest). To achieve this, data were collected for over 600 pancreatic biopsies, from approximately 250 patients, using both a standard (15x) and a high-resolution objective (36x), with the latter providing more detailed spatial information. Models were trained to distinguish between 6 classes, enabling tumor and inflammation detection with an accuracy of 95%. Additionally, the ability to detect the surgical margin, highlighting a clear boundary between healthy and tumorous tissue, was presented.

The theories regarding the influence of fibrous tissue (forming tumor microenvironment) on disease progression and the effectiveness of treatment, were the motivation for the development of methods for molecular orientation determination, based on polarized FT-IR spectroscopy. Results of 2D method application to collagen-rich tissues were presented, clearly indicating the directions of molecular orientation. Additionally, the Hermans function exposed areas of high and low levels of orientation, respectively for fibrous tissue and epithelial cells. This was the first application of this method to such a complex system as a tissue sample. In the following step, the first implementation of the theoretical 3D orientation determination method to experimental data was presented. For this purpose, a well-described in literature system of polycaprolactone polymer in the form of a spherulite was used as a model sample. Furthermore, the method was applied to O-PTIR, data allowing for sub-micron resolution measurements. The determined orientation of macromolecules was consistent with theoretical predictions, forming concentric circles around the nucleation center, and

highlighting the potential of this method in non-tomographic and non-destructive 3D orientation determination.

Fibrous tissue analyzed in orientation studies revealed a high tendency to scatter light, significantly disrupting the analysis. Therefore, the fourth and final topic addressed in the dissertation is the scattering correction. The implementation of the EMSC algorithm for Mie-type scattering correction in samples with cylindrical domains was presented. Additionally, the algorithm was extended to include linearly polarized light properties, optimizing it for orientation determination applications. To speed up the calculations, graphical processing units (GPUs) were utilized, with prepared MATLAB library later published as an open access. A polymer fiber with strong scattering properties was used for algorithm's validation, and their effectiveness exceeded other methods.