Improvement of histopathological classification based on chemical FT-IR imaging with data augmentation

FT-IR (Fourier-Transform Infrared Spectroscopy) imaging is a method that gives biochemical and morphological information about the sample[1]. This method in connection with statistical tools (chemometrics and machine learning algorithms), is successfully used in different biopsies tissue types recognition. Nowadays, one of the most frequent types of cancer and at the same most frequent cause of cancer related death in the case of women is breast cancer. One of the crucial stages in breast cancer diagnosis is histopathological analysis of biopsy taken from the patient. Breast tissue is highly differentiated and number of markers used for histopathological analysis is limited which making diagnosis harder. FT-IR imaging is free from these limitations because recognition of tissue type is based on biopsy biochemical composition. Models of classification breast cancerous and healthy tissue types based on their FT-IR images measured in transmission mode were recently successfully developed[2,3]. However, in the order to introduce above mentioned model for use in the clinic, a few obstacles must be overcome. Multiple sources of variability should be included for proper robust model creation, for example, high number of patients, sample preparation approaches, diversity of equipment and laboratory conditions. The solution to this labour-intensive problem are methods of data augmentation that expand data calibration model with new samples mathematically created, based on original data. The second issue which has an impact on introduction of this method to the clinic is connected with expensive sample substrates – optical windows that must be used in transmission mode when infrared radiation goes through the sample. There is a possibility to measure in transflection mode in which low price substrates can be used. However, in this mode radiation is reflected from the sample carrier and goes through the sample twice which produces additional signal distortion effect.

Consequently, **the goal of this research** is the creation of robust models of breast cancer detection and tissue types differentiation with use of FT-IR imaging, in transmission and transflection mode, and data augmentation methods.

The project will provide information about the influence of data augmentation methods on model robustness and breast tissue classification results. Also, the question about the possibility of breast tissue types differentiation with FT-IR images of biopsies measured in transflection mode will be answered.

1. Wrobel, T. P. & Bhargava, R. Infrared Spectroscopic Imaging Advances as an Analytical Technology for Biomedical Sciences. *Anal. Chem.* **90**, 1444–1463 (2018).

2. Mittal, S., Wrobel, T. P., Leslie, S., Kadjacsy-Balla, A. & Bhargava, R. A four class model for digital breast histopathology using High- Definition Fourier transform infrared (FT-IR) spectroscopic imaging. in *Progress in Biomedical Optics and Imaging - Proceedings of SPIE* (eds. Gurcan, M. N. & Madabhushi, A.) **9791**, 1–8 (International Society for Optics and Photonics, 2016).

3. Mittal, S. *et al.* Simultaneous cancer and tumor microenvironment subtyping using confocal infrared microscopy for all-digital molecular histopathology. *Proc. Natl. Acad. Sci.* **115**, E5651–E5660 (2018).