Abstract of Doctoral Dissertation

Degree requested Doctor of Life Science Applicant's name Khalifa Mohammad Helal

Title of Doctoral Dissertation

Raman Microscopic Histology Using Machine Learning Techniques for Non-Alcoholic Fatty Liver Disease
(非アルコール性脂肪肝疾患における機械学習技術を用いた Raman 組織学)

Histopathology is a standard means to diagnose the disease states of cells or tissues from their morphological features, but it requires the expertise of histopathologists and is therefore susceptible to human bias, and it can be affected by insufficient biochemical information on a microscopic level. Raman micro-spectroscopy can provide additional biochemical information that is not available to morphological examination, and has large potential to assist histological inspection and to benefit diagnosis of disease as objectively as possible in label-free manner. Detailed analysis of Raman microscopic data is essential to detect the spectral changes originated from underlying biochemical changes in cells or tissues due to disease.

This thesis is concerned with the development of diagnostic tools by integrating Raman microscopic imaging with methods of machine learning and information theory to analyze Raman hyper-spectral images of rat liver tissues comprising a dietary model of non-alcoholic fatty liver disease (NAFLD), with each liver tissue having been histopathologically diagnosed as normal, non-alcoholic fatty liver (NAFL), or non-alcoholic steatohepatitis (NASH). We employ two strategies to analyze Raman images to investigate the different states of NAFLD.

In the first study, dimension reduction (manifold learning) and ensemble-learning-based random forest classification are performed on the Raman spectra obtained from the regular spatial grid averaging of Raman images for predicting different states (dietary and histological), and enhancing the diagnostic capabilities to distinguish the states of tissues at early stages. We identify a set of important Raman bands in differentiating the Raman spectra arising from different states of tissues. Furthermore, NAFL state is distinguished into two pahses, namely, 'slowly progressive NAFL' (NAFL-α) and 'rapidly progressive NAFL' (NAFL-β) in terms of Raman imaging, and main Raman shifts to separate these two NAFL models are identified.

As a second diagnostic approach, using the dietary model of NAFLD in rats, we apply machine learning and information theory to evaluate cellular-level information in liver tissue samples. The method first increases the signal-to-noise ratio while maintaining spatial and spectral structures of Raman images as much as possible through extension of the simple linear iterative clustering superpixel algorithm developed in the area of image analysis. Second, using the unsupervised machine learning with rate distortion theory and the Poisson error arising from photon counting, it identifies a set of characteristic spectra having distinct Raman information across the tissues, and thus discovers diverse chemical environments in the liver tissues, allowing for the quantification of representative biochemical components such as vitamin A, lipids, and cholesterols which can be very important insights into the disease states of cells or tissues. Third, armed with microscopic information about the biochemical composition of the liver tissues, we are able to group tissues having similar composition using agglomerative hierarchical clustering, providing a descriptor enabling us to infer tissue states, contributing valuable information to histological inspection. Excessive lipid deposition with the appearance of cholesterol signatures indicates the severity of the disease state of the NAFLD tissues.

We expect that Raman microscopy coupled with the proposed techniques will offer new clinical tools that will aid pathologists in more precise NAFLD diagnosis by providing detailed molecular information about the liver tissue, and provide objectivity and precision through recognition of the biochemical properties of the samples.