Persistent Homology and Analysis of Histological Data

Kadantsev, Georgii (School: School 564)
Sinitsyn, Aleksandr (School: School 564)

Colorectal cancer is the second highest cause of cancer occurrence and death in men and women in the United States combined. A specialist usually confirms their diagnosis by a careful microscopical examination of a tissue sample. As a thorough analysis like this can be quite difficult when time is of the essence, close attention has been given to developments towards a computer-based diagnosis system in recent years. In a computer-aided examination of a tissue sample its digital copy is usually divided into many small images called patches. Each patch is then analysed individually. In this project we study these histology images using methods from topological data analysis (TDA). TDA is a modern approach to data analysis which aims to extract certain topological features from data. The primary characteristic of a patch for us is persistent entropy of an image, which is extracted from its 0-th persistent homology. It can be viewed as a certain numerical measure for the chaos present in an image described in a language of TDA. The goal of this project is to show that the concept of persistent entropy can be helpful in diagnosis of colorectal cancer. We have developed a fast and original algorithm for computing 0-th persistent homology and persistent entropy. We have analysed a big dataset of healthy tissue images and tumor images and observed a significant difference between the entropy of two patch classes. These findings can become pivotal in a new computer-based system for colorectal cancer diagnosis.

Awards Won: