Classification of Multiple Cancer Types with Deep Convolutional Neural Network

Authors : Nan Deng, Zhenqiu Liu

Abstract : Thousands of patients with metastatic tumors were diagnosed with cancers of unknown primary sites each year. The inability to identify the primary cancer site may lead to inappropriate treatment and unexpected prognosis. Nowadays, a large amount of genomics and transcriptomics cancer data has been generated by next-generation sequencing (NGS) technologies, and The Cancer Genome Atlas (TCGA) database has accrued thousands of human cancer tumors and healthy controls, which provides an abundance of resource to differentiate cancer types. Meanwhile, deep convolutional neural networks (CNNs) have shown high accuracy on classification among a large number of image object categories. Here, we utilize 25 cancer primary tumors and 3 normal tissues from TCGA and convert their RNA-Seq gene expression profiling to color images; train, validate and test a CNN classifier directly from these images. The performance result shows that our CNN classifier can archive >80% test accuracy on most of the tumors and normal tissues. Since the gene expression pattern of distant metastases is similar to their primary tumors, the CNN classifier may provide a potential computational strategy on identifying the unknown primary origin of metastatic cancer in order to plan appropriate treatment for patients.

Keywords : bioinformatics, cancer, convolutional neural network, deep leaning, gene expression pattern

Conference Title : ICMLB 2017 : International Conference on Machine Learning and Bioinformatics

Conference Location : Vancouver, Canada

Conference Dates : August 07-08, 2017