



MEDICINE 2040

Rewriting the future of healthcare

APPLIED DEEP LEARNING FOR THE ANALYSIS OF HIGH-THROUGHPUT DNA SEQUENCING DATA OF CANCER PATIENTS

Artem Danilevsky

Health Sciences, Tel Aviv University, Israel

High-throughput DNA sequencing, also known as next generation sequencing (or deep sequencing), has led to a quantum leap in our ability to read and understanding genetic information. Deep learning, is a particular field in machine learning allowing data-driven approaches to resolve extremely complex tasks. Deep learning models has proved to outperform conventional algorithms, and humans, in many tasks. Following training a model on large known data-sets, deep learning models can be applied to high-throughput DNA sequencing information in order to understand diverse patterns in noisy and composite data-sets. However, deep learning has hardly been used to interpret genetic information. Here we present our platform that allows leveraging deep learning methodologies onto DNA sequencing results. We demonstrate the capability of our platform to outperform current (published) approaches in classifying: (i) early detection of colorectal cancer from human feces; (ii) three types of human cancers from circulating tumor cells extracted from blood samples. Each data-set was trained solely on the DNA sequencing results without any preprocessing of the data. In both experiments we achieved better accuracy and Area Under the Curve (AUC) values than the original studies. Our platform will enable researchers to straightforwardly integrate deep learning into their high-throughput DNA sequencing data, and hence to interpret their genetic data hypothesis-free.