A Hybrid Feature Selection and Deep Learning Algorithm for Cancer Disease Classification

Authors : Niousha Bagheri Khulenjani, Mohammad Saniee Abadeh

Abstract : Learning from very big datasets is a significant problem for most present data mining and machine learning algorithms. MicroRNA (miRNA) is one of the important big genomic and non-coding datasets presenting the genome sequences. In this paper, a hybrid method for the classification of the miRNA data is proposed. Due to the variety of cancers and high number of genes, analyzing the miRNA dataset has been a challenging problem for researchers. The number of features corresponding to the number of samples is high and the data suffer from being imbalanced. The feature selection method has been used to select features having more ability to distinguish classes and eliminating obscures features. Afterward, a Convolutional Neural Network (CNN) classifier for classification of cancer types is utilized, which employs a Genetic Algorithm to highlight optimized hyper-parameters of CNN. In order to make the process of classification by CNN faster, Graphics Processing Unit (GPU) is recommended for calculating the mathematic equation in a parallel way. The proposed method is tested on a real-world dataset with 8,129 patients, 29 different types of tumors, and 1,046 miRNA biomarkers, taken from The Cancer Genome Atlas (TCGA) database.

Keywords : cancer classification, feature selection, deep learning, genetic algorithm

Conference Title : ICADLT 2020 : International Conference on Advances in Deep Learning Technologies

Conference Location : Dublin, Ireland

Conference Dates : February 13-14, 2020