Machine Learning Applied to the Field of Genomics

Simion-Daniel TĂTAR* and Gheorghe SEBESTYEN

Technical University of Cluj-Napoca, G. Barițiu Str., no. 26-28, 400027 Cluj-Napoca, Romania
E-mails: simiontatar@mail.utcluj.ro; Gheorghe.Sebestyen@cs.utcluj.ro

* Author to whom correspondence should be addressed

Abstract

As the quantity and quality of genomic data increases and the computational tools and infrastructure become readily available, the new algorithms provide us with new insights in the field of genomics. Traditional machine learning algorithms proved useful and new variants occurred. There is plethora of software solutions in this space based on these algorithms. A variety of genomic problems were tackled by these and are continuously studied. Emerging are a new class of algorithms based on artificial neural networks that constitute the branch of deep learning. Many solutions are proposed and modifications with useful optimizations. We conducted a review regarding machine learning applications in the field of genomics with an emphasis on deep learning algorithms and software solutions. The advantages and performances over the different data sets and the available solutions in form of software were enumerated. Where possible, a comparison either with traditional machine learning or with other similar deep learning approaches applied to the genomics field was conducted. Also, the architectures of deep learning models were of interest and presented.

Keywords: Machine Learning (ML); Deep Learning (DL); Genomics; Free and open-source software

Appl Med Inform 43(Suppl. S1) September/2021

S17