The 4th Iranian Conference on
Systems Biology

Which one will be more applicable in the future of ligand based in computational drug design studies: neural network or deep learning?

--------one line distance ----------

Fahimeh Ghasemi1*, Alireza Mehridehnavi1 , Horacio Perez Sunchez2

1School of Advanced Technologies in Medicine, Isfahan University of Medical Sciences, Hezar-Jerib Ave., Isfahan, IR Iran, 81746 73461.

2 Computer Science Department, Universidad Católica San Antonio de Murcia (UCAM), E30107 Murcia, Spain

**Corresponding Author : f_ghasemi@amt.mui.ac.ir*

Abstract

Drug-discovery protocols in pharmaceutical industry have mainly relied for years on the high throughput screening methods for quickly assay of the biological or biochemical activity of a large number of drug-like compounds. The total drug development process costs, more than hundreds of millions of dollars, should be added to this list. Of the high throughput virtual screening approaches, quantitative structure-activity relationships (QSAR).

…….. ……..

Keywords: QSAR; Feature Selection; Deep learning; Neural Networks

Introduction

QSAR is fundamentally a protocol that applies knowledge of statistics and mathematics to perform prediction or classification of biological data related to the designed molecules. Artificial neural network (ANN) is one of the most popular non-linear modeling methods utilized in QSAR studies. The number of drug-like compounds which are attractive for pharmaceutical industry is increasing every day. The same is true for the number of the molecular descriptors describing the physicochemical features of these compounds. In this article, it is attempted to discuss about the advantages and disadvantages of the proposed neural network algorithms and especially innovative deep learning techniques utilized in ligand-based virtual screening

Materials and Methods

Despite all neural network advantages in drug discovery compared to other machine learning algorithms, these algorithms have two serious concerns: The existence of thousands of descriptors as well as the correlation between them that leads the network to redundancy problems and the inevitably the situation of getting stuck in local minima. Therefore, data mining algorithms have been suggested to reduce the number of descriptors.

Results

Redundancy problem has been a significant concern in QSAR, especially in neural network design, which negatively influences network efficiency. In QSAR studies, redundancy refers to the similarity and correlation between descriptors. Choosing a limited number of descriptors as input neurons was a useful suggestion. The second challenging concern in neural network is applying the appropriate technique to avoid over-fitting. Moreover, in high throughput virtual screening, thousands of molecules and descriptors had inevitably led to select a network with more than one hidden layer and many nodes in each layer. In 2006, a novel fast algorithm was introduced by Hinton et al based on the restricted Boltzmann machine (RBM) which was the infrastructure of deep learning (DL) methods in the processing fields (e.g. computer vision, speech processing and image processing) and led to neural network recovery.

Discussion, Conclusion and Suggestions

In this article, it is attempted to present the most comprehensive literature review of neural network in drug discovery including each proposed method's merits and demerits. It can be concluded that in biological activity prediction using neural network, there are two major issues, which are the consideration of limited number of compounds, and high throughput virtual screening based on thousands of molecules. Regarding the first issue, a limited number of compounds and a large number of descriptors lead to both redundancy and over-fitting as major problems for the model. On the second issue, a different situation is encountered: a large number of compounds with thousands of descriptors end up in opening a new path to use deep learning architecture. On the other hand, a large number of input data requires a network with a large number of computational elements. That's why in QSAR focus is on finding the best model based on deep architecture to avoid the situation of getting stuck in local minima and being prone to over-fitting. In the recent years, several algorithms are introduced based on deep learning which demonstrated much better performance than the previous learning algorithms in neural network. Most of these proposed algorithms focused on network parameters' fine tuning before the learning procedure starts.

References

Ghasemi, F, AR Mehridehnavi, and H Pérez-Sánchez. " Neural network and deep-learning algorithms used in QSAR studies: merits and drawbacks." Drug discovery today (2018).