

A Deep Learning Approach to Increase the Accuracy of Predicting Protein-Compound Interactions

**P. M. P. C. Gunathilake^{1*}, W. M. P. G. D. M. Roshan¹,
P. D. C. R. Jayarathna²**

¹Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Peradeniya 20400, Sri Lanka

²Doer Innovative Technologies Pvt. Ltd., Moon Valley Estate, Kundasale, 20168, Sri Lanka

**prabhath@pdn.ac.lk*

In drug repositioning and discovery, a system-level identification of drug-target direct interactions is important. However, even today a wide range of drug compounds and protein targets makes it challenging and expensive to identify the interactions between them. For protein-compound interaction prediction, traditional similarity-based computational models were used and they rarely utilize the hidden features from currently available large-scale unlabelled compound and protein data. They often limit their usage on relatively small-scale datasets. This work proposes a new schema for predicting the protein compound interaction and binding score using deep learning approach. Most accurate features from a large set of proteins and compounds were selected and converted into a form of numerical values as the input data. The known interactions between selected compound and protein pairs were taken in binary form as the target dataset. The numerical input dataset and the binary target dataset were used to train a deep neural network model under a supervised learning process. The neural network was optimized and trained with different activation functions and optimization algorithms. The trained neural network model was then used to predict the protein-compound interactions. Also, the second approach was made to predict the protein-compound binding score. The same set of protein-compound pairs were selected as the input of the neural network model. A multiple levels of numerical values representing the binding score between each protein-compound pair in the input dataset was taken as the target instead of the binary target in the interaction prediction method. Then the neural network model was optimized and trained to predict the binding score for a protein-compound pair. The cross validated accuracy for protein-compound binding score prediction neural network was 99.6% and for the protein-compound interaction prediction neural network it was 77.32%.

Key words: Deep Learning, Artificial Neural Network, Machine Learning, Protein Compound Interaction, Drug Discovery