JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

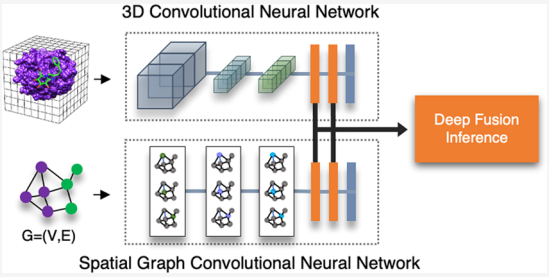
Improved Protein−Ligand Binding Affinity Prediction with Structure-Based Deep Fusion Inference

Derek Jones,*,⊥ Hyojin Kim,*,⊥ Xiaohua Zhang, Adam Zemla, Garrett Stevenson, W. F. Drew Bennett, Daniel Kirshner, Sergio E. Wong, Felice C. Lightstone, and Jonathan E. Allen

Cite This: J. Chem. Inf. Model. 2021, 61, 1583−1592 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Predicting accurate protein−ligand binding affinities is an important task in drug discovery but remains a challenge even with computationally expensive biophysics-based energy scoring methods and state-of-the-art deep learning approaches. Despite the recent advances in the application of deep convolutional and graph neural network-based approaches, it remains unclear what the relative advantages of each approach are and how they compare with physics-based methodologies that have found more mainstream success in virtual screening pipelines. We present fusion models that combine features and inference from complementary representations to improve binding affinity prediction. This, to our knowledge, is the first comprehensive study that uses a common series of evaluations to directly compare the performance of three-dimensional (3D)-convolutional neural networks (3D-CNNs), spatial graph neural networks (SG-CNNs), and their fusion. We use temporal and structure-based splits to assess performance on novel protein targets. To test the practical applicability of our models, we examine their performance in cases that assume that the crystal structure is not available. In these cases, binding free energies are predicted using docking pose coordinates as the inputs to each model. In addition, we compare these deep learning approaches to predictions based on docking scores and molecular mechanic/generalized Born surface area (MM/GBSA) calculations. Our results show that the fusion models make more accurate predictions than their constituent neural network models as well as docking scoring and MM/GBSA rescoring, with the benefit of greater computational efficiency than the MM/GBSA method. Finally, we provide the code to reproduce our results and the parameter files of the trained models used in this work. The software is available as open source at https://github.com/llnl/fast. Model parameter files are available at ftp://gdo-bioinformatics.ucllnl.org/fast/pdbbind2016_model_checkpoints/.



INTRODUCTION

Predicting accurate binding affinity between a small molecule and a target protein presents a fundamental challenge in drug development. Recently, deep learning models have been proposed as an alternative to traditional physics-based free energy scoring functions. The benefit of the deep learning approach is in learning binding interaction rules directly from an atomic representation without relying on hand-curated features that may not capture the mechanism of binding.[1,2] Two types of three-dimensional (3D)-structure-based deep learning approaches are most commonly used, 3D-convolutional neural networks (3D-CNN) and spatial graph convolutional neural network (SG-CNN).[3,4] The 3D-CNN, which has been previously applied to the problem of binding affinity prediction,[5] uses a 3D voxel representation of atoms. The 3D-CNN representation implicitly accounts for pairwise relationships among atoms through their relative positioning in the 3D voxel grid but does not predetermine which atomic interactions to represent other than defining a minimum atomic resolution. This comes at the cost of having to learn a larger number of parameters to represent the 3D voxel grid. In contrast, the SG-CNN uses explicit distance thresholds to determine which pairs of atoms to consider in pairwise interactions (e.g., covalent, noncovalent), with the potential benefit of requiring fewer parameters in the model. Both approaches show promise, with complementary strengths and weaknesses. However, both methods have yet to be compared directly with each other or to traditional physics-based scoring functions. The benefit of fusion models lies in combining their possibly complementary feature representations. Several fusion models have been proposed for the task of video activity recognition by bridging the dimension and feature difference among multiple input types (e.g., visual and temporal data).[6,7] Previous work in the computer vision domain[8] addressed several distinct strategies for fusion models and their

Received: November 10, 2020
Published: March 23, 2021