Deep Neural Networks for Protein-Ligand Interaction Prediction - Survey

Vano Mazashvili (Student)

DIAG - Department of Computer, Control and Management Engineering, Sapienza University of Rome, Rome, Italy

ARTICLE INFO

Keywords:
Deep Learning
Ligand-Protein interaction
Molecular Bioinformatics
Protein Structure
Molecular Docking
Drug Design
Artificial Intelligence

ABSTRACT

Proteins are the building blocks of all biological organisms. For drug discovery, after the target protein identification, possible interactable ligand candidates are chosen, which can alter the function of the given molecule. Interpreting Protein-Ligand interactions is a critical component in understanding most biological processes. The protein-ligand docking modeling is generally used in drug discovery to figure out the binding affinity and molecular interplay of the ligand on protein molecules. Finding the optimal binding parameters involves molecular modeling and is a dynamic strenuous and time-consuming process if done in vitro. Deep Learning, a subset of Machine Learning has been used to ease and streamline the given task. Effectively identifying the optimal binding molecules and then predicting the protein-ligand binding affinity with the help of the Deep Neural Networks is a fundamental step towards safe and fast in silico drug discovery. This survey focuses on the categorization of existing methodologies in the field of deep learning used for protein-ligand interaction prediction and states the motivations and challenges of the given solutions.

1. Introduction

Understanding protein-ligand interaction is a fundamental task in molecular and structural biology and drug discovery. Finding the target proteins involved in medical conditions first and then developing the appropriate ligands capable of changing the protein's structure and function in a controlled and desired fashion is a strenuous task, considerably stretching the drug development timeline. This can be associated with high Research and Development costs, translating into high prices.

Several in silico methods are used to aid in the process of protein-ligand pair selection, which essentially involves the prediction of drug-target interaction. Although less reliable, these approaches have been used successfully in practice. Although the older methods suffered from accuracy and data availability issues, deep learning has found great improvements in given metrics.(1)

In this survey, we investigate the deep learning models used in ligand-protein binding site prediction, exploring the problems and challenges associated with the given solution methods. We categorize given solutions, focus on cutting-edge methodologies, and reveal the future trajectory of this domain.

2. Problem Description

Despite the great advances in drug research aided by the exploding field of machine learning, the drug development process remains a resource-intensive endeavor. Searching for the ideal protein-ligand pair is time-consuming and cannot be solved with brute force. There is a need for a specific algorithm or heuristic that can narrow down the vast field of candidate molecules available. The problem is multifaceted; To create a comprehensive solution for the drug synthesis pipeline, the appropriate model for drug target analysis should be created. Although we come across more complications: To create accurate, interpretable, and generalizable protein-ligand interactions, intricate biochemical and molecular properties should be considered while analyzing the PLIs. Considering limited data, modeling the explicit local interactions efficiently and accurately is a challenge, meaning that, learning the patterns from the low representation data yields poor predictions as the same molecules can have various alignments, angles, and docking pockets. Besides that, generalization prediction accuracy and robustness are questionable as there are shifts in domain distributions between the datasets and real-life scenarios.

Competent and highly automated Deep Learning solutions will considerably shorten the drug development timeline. In addressing the stated issues, we hope to refine state-of-the-art solutions by developing new models capable of learning and interpreting complex biomolecular properties.

3. Existing Methods Categorized

Here we categorized the existing problems and different methods with their advantages and disadvantages, how they handle generalization, and we mention their prediction accuracy. We also dive into the novel approaches and analyze their effectiveness and shortcomings.

3.1. Multimodal Approaches

The Bilinear Attention Network is utilized for the prediction of DTI for the DrugBAN framework. (1) The model uses two data streams, one for the target protein structures and the other for the ligand. The graph convolutional network and the convolutional neural network encode the molecular graph of the drug and the one-dimensional protein sequence respectively. This is done to increase the information dimension since, in higher dimensions, less molecular information is lost. This is achieved by converting the simplified one-dimensional molecular input line entry system into a two-dimensional molecular graph.(1)

✉ mazashvili.1993251@studenti.uniroma1.it (V. Mazashvili)

Eventually, the drug and protein representations are fed to the bilinear attention network module. The given block captures DTI by acquiring pairwise attention weights and a joint drug-target representation. The output is a probability of the residue-ligand affinity. (1)

A more geometric approach is taken for EquiPocket (12). Instead of topological information, the model takes advantage of the spatial structure of the molecules. Unlike DrugBAN, EquiPocket treats proteins as graphs, where, instead of being a one-dimensional residue-based representation, we have a three-dimensional interpretation. Not only that, it leverages a multimodal architecture by introducing surface points, which are not necessarily atoms, but a representation of the 3-dimensional surface. (1; 12). Representing the target molecules as graphs enables us to leverage the power of GNN, where we can encode the irregular protein structures and yield a translation/rotation (E(3)-equivariant) invariant solution, improving generalization. (12)

It is also helpful to consider protein's global structure in multimodality, as working with *just* local interactions is a resource-intensive job that isn't maximally generalizable. Global structure modeling gives us more information on how the protein fold is achieved and thus how the pocket is formulated (12), which helps to predict the PL site.

Combining these two information streams grants us multi-level structure features, reflecting high generalizability and good prediction accuracy. (12)

In another multimodal SE(3)-equivariant solution presented in a paper by D. Lee et al. (6) we have a Binding Site Prediction (BSP) task disassembled into two components. Binding Site Detection (BSD) and Binding Residue Identification (BRI). The BSD and BRI modules are constructed accordingly. The authors used transfer learning to initiate the parameters for the BSD component to address data scarcity. Compared to (12), (6) uses a more robust SE(3)-equivariant attention mechanism with SE(3)-equivariant grid featurization. The authors focus on different practices to boost the performance of the given model. The model essentially does two things:

Evaluating the Druggability of the Binding Sites - After generating the candidate binding sites in the chosen protein molecule from the outside source, the BSD module outputs the druggability value for each. To achieve this, the residues around the center mass of the protein structure are featurized with the 3-dimensional grids.

(6) That means that each feature now encodes the neighborhood of its residues. Although Y. Zhang et al. (12) criticize CNNs usage with 3-Dimensional grids, due to their sensitivity to rotations, the given model still achieves not only E(3), but more robust SE(3)-equivalency. In addition, contrary to taking the whole structure information as a whole, (12) the given method seems more efficient regarding computation power because it is concentrated on point embeddings, rather than the whole information. However, one shortcoming of this method could be the unawareness of data distribution shifts.

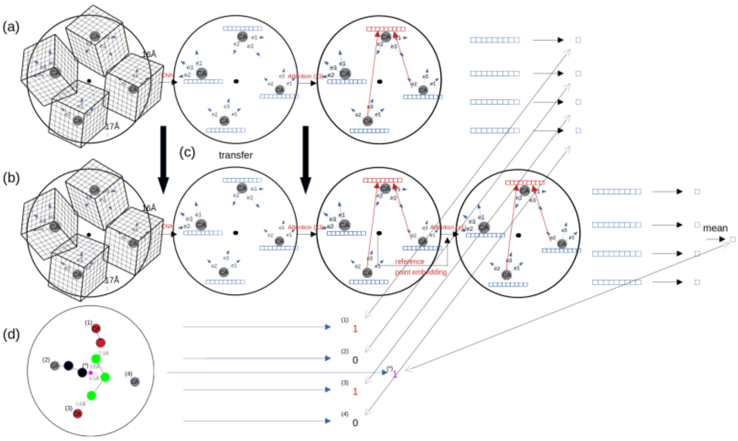

Figure 1: (a) BRI module (b) BSD module (c) Transfer learning (d) Ground truth generation. (6) Notice, that for the BRI module, the last point-wise feed-forward layer without a mean-reduction operation from the BSD is non-existent(6)

The embedded features from the GNN are first extracted and then aggregated, resulting in a single scalar value correlating to the specific candidate binding site.

Key Residue Identification - For this subproblem, the BRI model comes into play. It shares the same architecture as the BSD without the last point-wise feed-forward layer and a mean-reduction operation (6). This means that the output shows the specific residues in binding sites.

Transfer learning is used to extract the parameters with the weights from the BRI module and use them to initialize the weights for the BSD module. (6) The authors justify this decision with an intuition that the pattern of the resulting specific residue combinations can determine the binding affinity on its own.

In addition, the abundant new labels help BSD perform better.

A similar E(3)-equivariant multimodal architecture is implemented in FABind, using ligand, receptor, and pair representations. (11) The binding pocket is represented by the center of the residues or amino acids containing the mixture. The model uses the FABind layer, which, on its own consists of the Equivariant Graph Convolutional (EGCL) or an independent message passing layer, cross-attention update, and interfacial message passing layers to not only solve pocket prediction tasks but also docking, executed in a streamlined fashion. (11) Notably, each layer besides the cross-attention update layer is E(3)-equivariant since it does not encode the structure. The given FABind layers are deployed for both pocket prediction and docking tasks.

The FABind layer is an integral component of the model architecture. This deep diffusion model is capable of learning complex noncovalent intermolecular interactions, such as hydrogen bonding and hydrophobic interactions. Respectively, the aforementioned sub-layers, take a pair of

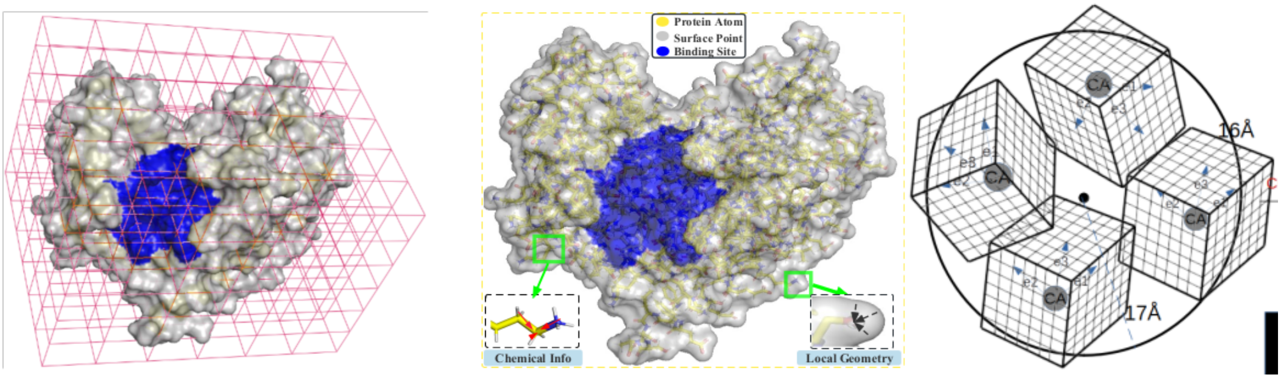

Figure 2: CNN-based methods by voxelization vs EquiPocket protein graph (12) vs featurized surroundings of the candidate binding site centers (6). Note the difference between the older voxelization method (no E(3)-equivariancy), E(3)-invariant 3D surface representation and SE(3)-equivariant residue voxelization

protein and ligand representations as input (and a pair representation from the second, cross-attention update layer) and (11)

a. Update the node embeddings and coordinates in each component

b. Capture the ligand-residue correlations and modify the embeddings

c. Attentively update coordinates and representations and focus on the contact surface

The FABind layer is designed to be equivariant to the input representations, meaning that it can process proteins of different sizes and shapes.

Unlike the other papers, the FABind model treats pocket prediction as a binary classification problem. (11) Here it is assumed that the amino acid or a residue can or cannot be in the pocket. (11) Therefore, sensibly, the pocket classifier is trained with the binary cross-entropy loss function. (11) The pocket center constraint is constructed to include/exclude classified residues from the pocket center. As the Gumbel-Softmax function gives us a differentiable approximation of the discrete selection process, it is used to assign probabilistic weighed decisions to each residue classified, respectively, as the given computation involves discrete decisions. This aids in model precision as the choice has an additional dimension of probability. (11)

FABind model deploys a very efficient regression-based method to compute the protein-ligand docking. The iterative refinement is introduced to the FABind layers, allowing for refinement of the initially predicted ligand pose by feeding it back to the FABind layer for multiple iterations. (11)

Adequate results can be achieved just using the multistream convolutional neural network, where the streams represent drug, protein, and combined characterization respectively. (3) It becomes exponentially harder to represent the molecules and their interplay with high accuracy as the number of parameters increases. The drug-target binding affinity is affected by the toxicity, solubility, IC50 (half-maximal inhibitory concentration), drug-receptor affinity, and efficacy of the competitive, noncompetitive, and uncompetitive enzyme inhibition, etc. (3) KIBA dataset made it possible to take these merits and combine them into a KIBA score(3), where we obtain the protein inhibition efficacy.

Contrary to the narrative, (3), chooses compound SMILES strings of the KIBA dataset as input, ditches outlier molecules in terms of size, and enforces an upper and lower limit for molecule size. The resulting database entries were concatenated and, if necessary, padded, and fed to the ResDTA model. It is worth noting that the SMILES, combined and Sequence representations generated by the 3 CNN modules, are provided to the network with the residual skip connections, significantly improving the accuracy. However, being able to perform well in 5-fold cross-validation, with the overall concordance index score of 0.887 and MSE of 0.0002, (3)the critique of the model would be a theoretical lack of generalizability, as molecular outliers were ditched in training and tests were not performed on the datasets with the shifted domain distributions.

3.1.1. *Fusion Methods*

This section delves into fusion/assembly techniques that enhance the robustness/accuracy of the individual models.

Classical Hybrid Models

As we observed, classical neural network architectures mostly leverage either 3D molecule structures or biomechanical data such as covalent bonds from the graph representation. Although promising results, combining the two possibly complementary feature representations in one model should yield an improvement.(5) By fusing 3D-CNN and SG-CNN, (5) aims to improve performance and accuracy by capturing 3D atomic features and implicit atomic interactions and covalent and noncovalent interactions from the graph representation by introducing an arbitrary number of edge types and a node information propagation technique.(5) Although the [fusion] model is simple (averaging the final predictions of the CNN models) (5) it still manages to capture the complementary information and outperform, or be comparable to the regular methods in Pearson r, Spearman r, MAE, and RMSE metrics.

Quantum Hybrid Machine Learning

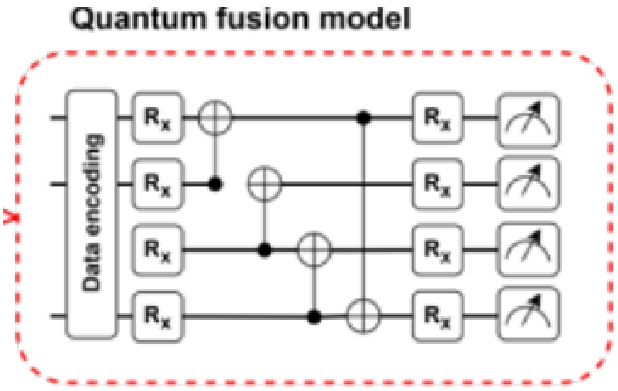

Figure 3: The quantum fusion model is responsible for the combination of the outputs from 3D-CNN and SG-CNN classical networks. The output vectors of size 16 and 6 respectively, are encoded onto quantum states and then fed into feed-forward QNN, 3D generalized rotation gates with all-to-all CNOT connections (2)

One of the more novel approaches in the machine-learning field is the utilization of quantum computers. In particular, when dealing with exponentially increasing data dimensionality, Quantum Machine Learning (QML) usually outperforms classical machine learning models.(2) Accordingly, in (2) a novel approach is proposed, incorporating a Hybrid quantum-classical model. It consists of three main components, 3D-CNN, spatial graph CNN (SG-CNN), and quantum fusion model. Essentially, this model is an improved version of the (5), the notable difference being the

QNN incorporated in fusion. Just before the output layers, the information is taken out from 3D-CNN and SG-CNN modules' fully connected layers and passed to the quantum fusion model, to collect the information from the aforementioned networks. The architecture of the fusion model is as follows: (2)

a. Quantum Encoding - Encodes data into quantum states

b. Parametrized Quantum Circuit (PQC) - leveraging a substantial level of entanglement and expressibility, PQC uses classically optimized parametrized quantum gates.

In other words, the data are translated into quantum states, where it is learned upon by the parametrized, thus trainable, quantum circuit, potentially capturing complex relationships between the inputs. As a result, the problem of convergence stability has been addressed with good prediction accuracy and improved generalization capacity. (2) The results show boosted RMSE, MAE and minimized R^2, Pearson, and Spearman coefficients compared to (5), overall 6% increase in accuracy with better, faster, and smoother convergence. (5) (2)

3.2. Unimodal Approaches

Unimodal approaches enable us to use more streamlined networks and avoid unnecessary data augmentation like concatenations and padding. The singular representation should be less biased towards a particular modality while being computationally efficient. Although it is harder to capture comprehensive information from a single stream rather than multiple, unimodal approaches are still valid choices for the given task.

PIGNet2 is an example of the multimodal network approach, where the input data is a protein-ligand complex, fed to the GCNN layers. The resulting graph representation is then passed to the fully connected layers, making the final prediction about the binding affinity. (8)

The paper employs novel data augmentation strategies, one of which is positive data augmentation (PDA), generating crystalized conformations of near-native protein-ligand complexes. The basic mechanism behind this strategy is small structural perturbations in the protein-ligand complexes, which are evaluated by the physics-informed scoring system and appropriately added to the dataset to aid the network's prediction performance. (8) It is noteworthy, that D. Lee et al. (6) uses a similar approach, but it is based on residual orientation perturbations.

3.3. Techniques to Improve Cross-Domain Prediction

Cross-domain adaptation makes the model more robust and accurate in real-life scenarios. As it is easy for a machine learning model to "operate" in the domain (1), there is a need to improve the generalization to achieve better performance in various scenarios. The drug development field is especially prone to generalization shortcomings because there is a scarcity of diverse data sources affected by various experimental conditions. This creates a domain distribution shift between the source and target domains. (1)

One way to combat this issue is a CDAN domain adaptation module. Similarly to Conditional Generative Adversarial Networks, CDAN exploits discriminative information, conditioning the domain classifier on the predicted class probabilities to align the distributions between source and target domains. (1)

One way to improve the cross-domain prediction is to use a specific data augmentation strategy. For instance, PIGNet2 introduces negative data augmentation (NDA) approaches like cross-docking and random docking. In this case, instead of PDA, our objective is to introduce novel structures into the training data, promoting generalization (8).

4. Discussion and Open Challenges

In this survey, we explore the use of deep learning in the prediction of drug targets. The papers presented in this survey show the complexity and importance of the drug development problem. A Plethora of solutions and learning techniques including geometric deep learning, quantum-classical hybrids, multimodal and unimodal networks, and holistic frameworks indicate the flexibility of deep learning and underline the multifaceted nature of the problem. However, more challenges remain than we tried to solve, such as data generalization, efficiency, performance, and explainability.

I want to state that there is a deficit of good-quality data, and it is one of the biggest challenges we face right now. On top of that, there is a significant shift in domain distributions in the existing data. More time and energy should be devoted to obtaining and labeling high-quality data.

Another challenge is a lack of explainability. As the task is related to human health, it requires high interpretability. The black-box nature of deep learning is an issue. To deal with this, we will close the bridge between the in silico synthesis and the real-life clinical validation phases.

Finally, some recent articles have developed hybrid quantum-classical techniques that have promise. It might be revolutionary to develop this field of study and discover more efficient ways to harness quantum computing power for medication target prediction.

Overall, It is certain that we will develop more advanced and integrated holistic deep learning solutions in the future that are interpretable, fast, and accurate, helping us through the hardest times, and providing us with an easy and flexible framework for drug synthesis.

References

[1] Bai, P., Miljković, F., John, B., and Lu, H. Interpretable bilinear attention network with domain adaptation improves drug-target prediction, 2023.

[2] Banerjee, S., Yuxun, S. H., Konakanchi, S., Ogunfowora, L., Roy, S., Selvaras, S., Domingo, L., Chehimi, M., Djukic, M., and Johnson, C. A hybrid quantum-classical fusion neural network to improve protein-ligand binding affinity predictions for drug discovery, 2023.

[3] Ghosh, P., and Haque, M. A. Resdta: Predicting drug-target binding affinity using residual skip connections, 2023.

[4] He, Y., Peng, S., Chen, M., Yang, Z., and Chen, Y. A transformer-based prediction method for depth of anesthesia during target-controlled infusion of propofol and remifentanil, 2023.

[5] Jones, D., Kim, H., Zhang, X., Zemla, A., Stevenson, G., Bennett, W. F. D., Kirshner, D., Wong, S. E., Lightstone, F. C., and Allen, J. E. Improved protein–ligand binding affinity prediction with structure-based deep fusion inference. *Journal of Chemical Information and Modeling 61*, 4 (2021), 1583–1592. PMID: 33754707.

[6] Lee, D., Byun, J., and Shin, B. Boosting convolutional neural networks' protein binding site prediction capacity using se(3)-invariant transformers, transfer learning and homology-based augmentation, 2023.

[7] Lee, H.-J., Emani, P. S., and Gerstein, M. B. Improved prediction of ligand-protein binding affinities by meta-modeling, 2023.

[8] Moon, S., Hwang, S.-Y., Lim, J., and Kim, W. Y. Pignet2: A versatile deep learning-based protein-ligand interaction prediction model for binding affinity scoring and virtual screening, 2023.

[9] Ngo, N. K., and Hy, T. S. Target-aware variational auto-encoders for ligand generation with multimodal protein representation learning, 2023.

[10] Papillon, M., Sanborn, S., Hajij, M., and Miolane, N. Architectures of topological deep learning: A survey on topological neural networks, 2023.

[11] Pei, Q., Gao, K., Wu, L., Zhu, J., Xia, Y., Xie, S., Qin, T., He, K., Liu, T.-Y., and Yan, R. Fabind: Fast and accurate protein-ligand binding, 2023.

[12] Zhang, Y., Huang, W., Wei, Z., Yuan, Y., and Ding, Z. Equipocket: an e(3)-equivariant geometric graph neural network for ligand binding site prediction, 2023.

[13] Ünlü, A., Çevrim, E., Sarigün, A., Çelikbilek, H., Güvenilir, H. A., Koyaş, A., Kahraman, D. C., Olğaç, A., Rifaioğlu, A., and Doğan, T. Target specific de novo design of drug candidate molecules with graph transformer-based generative adversarial networks, 2023.