2024, 21(11): 3605-3616
https://doi.org/10.21123/bsj.2024.9212
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Deep Learning based Models for Drug-Target Interactions

Ali K. Abdul Raheem*[1,2], ***Ban N. Dhannoon***[3]

[1] Software department, College of Information Technology, University of Babylon, Hillah, Babil, Iraq.
[2] Department of Qur'anic Sciences, College of Islamic Sciences, University of Warith Al-Anbiyaa, Kerbala, Iraq.
[3]Department of Computer Science, College of Science, Al-Nahrain University, Baghdad, Iraq.
*Corresponding Author.

Received 22/06/2023, Revised 12/09/2023, Accepted 14/09/2023, Published Online First 20/04/2024, Published 01/11/2024

Abstract

The typical drug development approach is slow, costly, and fraught with failure - scientists examine millions of compounds, but only a few make it to preclinical or clinical testing. Machine learning (ML), a subset of AI, is a fast-expanding subject many pharmaceutical businesses increasingly utilize. Incorporating machine learning technologies into the drug development process can aid in automating repetitive data processing and analysis processes. ML techniques may be used at several stages of drug development, including drug property prediction, drug-target interaction (DTI) prediction, and De Novo drug design. DTIs are a critical component of the drug development process. When a drug (a chemical molecule) attaches to a target (proteins or nucleic acids), it is said to bind; it alters its biological behavior/function, returning it to normal. DTI prediction is an essential part of the Drug Discovery process since it may speed up and decrease costs, but it is challenging and costly because experimental assays take a long time and are expensive. In recent years, deep learning-based approaches have demonstrated encouraging results in predicting DTI. This paper developed two deep-learning architectures to predict drug-target interactions. The first model uses message-passing neural networks (MPNN) for drug encoding and bidirectional gated recurrent units (Bi-GRU) for protein-encoding. The second model uses Bi-GRU for drug encoding and protein encoding. The two models were trained and evaluated on several benchmark datasets. Our results demonstrate that our models outperform state-of-the-art DTI prediction methods and are a promising approach for predicting DTI with high accuracy.

Keywords: Bi-GRU, Deep Learning, Drug-target interactions, Drug Discovery, MPNN, Prediction computational models.

Introduction

Drug discovery and development have been sped up because of the advances in computational science. Artificial intelligence (AI) is widely used in both industry and academia. Machine learning (ML), an essential component of AI, has been used in a variety of contexts, including data production and analytics[1]. Drug discovery is one area that stands to gain significantly from this machine learning achievement. ML may be used to accelerate and minimize the labor-intensive and costly process of discovering novel medications [2,3]. Drug development, bioinformatics, and cheminformatics have benefited from introducing these computer-assisted computational techniques. Bringing a new medicine to market is complicated and time-consuming, costing pharmaceutical companies an average of $2.6 billion and 10 years of research and development as shown in Fig.1[4] .

2024, 21(11): 3605-3616
https://doi.org/10.21123/bsj.2024.9212
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

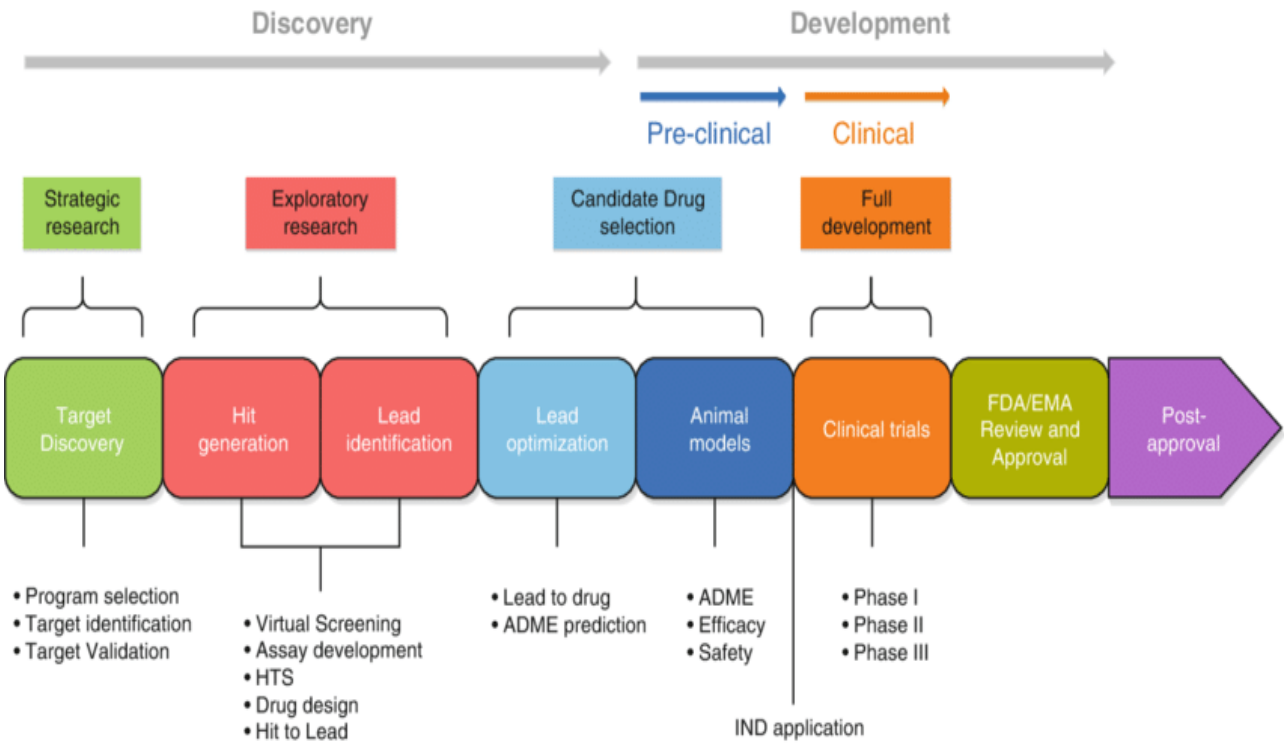

Figure 1. Drug Discovery Process [4].

DTIs are an important stage in the drug discovery process. Because it is difficult to identify potential active chemicals in silico, leveraging large chemical libraries can improve drug development efficiency by removing the need to test all chemical compounds against a given target protein [5]. However, identifying drug-target interactions in Wet-lab (in vitro) research is highly expensive, time-consuming, and complex [6]. Computational (in silico) techniques can successfully enhance traditional in vitro activity detection procedures, allowing the identification of interacting drug-target combinations and speeding up drug discovery. DTIs have been studied for decades by researchers using clinical observations and biological studies. Using these experimental procedures, however, is still consumes time and costly. They must also contend with a high rate of attrition [7]. As a result, in recent years, in drug development and molecular pharmacology, computational approaches to predict drug-target interactions have emerged as a hot research topic.

When a medicine (a chemical molecule) Connects a protein or nucleic acid to a target, it modifies its biological behavior/function, returning it to normal [8]. DTI prediction is an important aspect of the DD process since it can both expedite and save costs [9]. However, it is challenging and expensive since experimental assays take a long time and are pricey [10]. As a result, researchers have increased their efforts to identify the association between medications and targets to speed up drug development and shorten the time to market [11]. Based on existing DT trials, computer-generated DTI predictions can be used to efficiently evaluate the interaction strength of new drug-target (DT) combinations. Thus, when dealing with a vast amount of complex information, The DD process is hastened by systematically suggesting a new set of candidate molecules (e.g., interactions (hydrophobic, ionic, hydrogen bonding, and/or van der Waals forces) between molecules) [12,13].

The process of drug discovery is costly and time-consuming, which can lead to increased healthcare costs for patients. The identification of DTI is a substantial part of the drug discovery process. Therefore, there is a need to reduce the cost of predicting DTI in order to accelerate drug discovery and make it more affordable. Additionally, personalized medicine can be developed with a precisely learned molecule representation in a DTI model, benefiting numerous patient cohorts. Furthermore, predicting binding affinity values of drug-target pairs remains a challenge, even with the increasing availability of affinity data in DT knowledge bases. These issues highlight the need for advanced learning techniques such as machine and deep learning architectures to improve the prediction of DTI and binding affinities.

2024, 21(11): 3605-3616
https://doi.org/10.21123/bsj.2024.9212
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

This paper aims to develop a cost-effective and accurate model for predicting drug-target interactions that can contribute to the development of personalized medicine. also to improve the efficiency and accuracy of predicting binding affinity values. The paper was able to contribute to these problems and achieve the goals by proposing two different models for predicting drug-target interactions.

The structure and content of the paper are organized into several main sections, each serving a specific purpose. the introduction section provides an overview of the problem of drug-target interactions and their significance in drug discovery. The objectives and contributions of the paper are clearly stated in the related work section, the paper reviews existing literature and research on drug-target interaction prediction using deep learning methods. In the materials and methods section, the paper describe the datasets used in the study, their sources, and the preprocessing steps applied to ensure data reliability. Also presents the deep learning models designed specifically for drug-target interaction prediction. The paper describe the architectural design of each model and the rationale behind its development. then the results and discussion section presents the experimental outcomes obtained from evaluating the proposed systems. Performance comparisons against baseline methods and existing state-of-the-art approaches are discussed. Finally the conclusion section summarizes the key findings of the study and emphasizes the novelty and contributions of the proposed deep learning models.

Related Work

Numerous recent publications have been published that use machine learning techniques to anticipate drug-target interactions. This study focuses on the most recent works on this topic that employ the methodologies and applications discussed in the previous section.

- Abdul Raheem and Dhannoon [14] In this study, they discussed the steps of drug discovery and methods of machine learning that can be applied in each step, as well as given an overview of each of the research works in this field. they also presented a proposal to use machine learning techniques to solve drug discovery problems.
- Ye Q, et al.[15] The researchers addressed the issue of various targets having vastly varied quantities of interactions. In greater detail, several positive samples are available for targets with a high number of interactions (TWLNI). However, for targets with a low interaction rate, a limited number of interactions might result in a relatively modest positive sample count. As a result, researchers created two distinct categorization algorithms for these two categories of targets. Based on the aforesaid concept, a novel approach for Multiple Classification Methods (MCSDTI) used to predict DTI is developed, which predicts interaction utilizing distinct classification strategies of TWSNI and TWLNI. Furthermore, they examined TWSNI and TWLNI individually, which allowed them to avoid the problem that when all targets are reviewed simultaneously, the conclusion might be predominantly driven by targets with a significant quantity of interactions. The scientists analyzed five datasets: Nuclear receptors (NR), G protein-coupled receptors (GPCR), enzymes (E), ion channels (IC), and a drug bank are all examples of proteins. These databases are created by smiles. The results show that their method's AUCs are 3.31%, 1.27%, 2.02%, 2.02%, and 1.04% higher than the second best methods for TWLNI on the NR, IC, GPCR, and E, and 1.00%, 3.20%, and 2.70% higher than the second best approaches for TWSNI on the NR, IC, and E.
- De Souza et al.[16] In This work, they proposes a method for predicting the binding affinity between drugs and targets using a similarity-based approach with a 2D convolutional neural network (CNN). The model is validated on public datasets and demonstrates its effectiveness in estimating drug-target binding affinity.
- Shim and Hong[17] The authors develop a deep learning model, HoTS, which predicts binding areas and drug-target interactions using protein sequence information. The model performs well in predicting binding areas and outperforms other models in drug-target interaction prediction.
- Lee and Nam [18] This paper presents MPS2IT-DTI, a deep learning model for drug-target interaction prediction. The model uses a convolutional neural network and achieves better performance than existing state-of-the-art methodologies. It represents molecular and

2024, 21(11): 3605-3616
https://doi.org/10.21123/bsj.2024.9212
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

protein sequences as images, avoiding the need for an embedding layer.

- Öztürk et al.[19]. The study proposes a deep learning-based approach for calculating binding affinity in protein-ligand interactions. The model uses 1D representations of targets and pharmaceuticals and outperforms other methods, including a state-of-the-art approach, by utilizing CNNs to create high-level representations.
- Mukherjee et al.[20] was introduced a Deep GLSTM as a graph convolutional network and LSTM-based strategy for predicting binding affinity between FDA-approved drugs and SARS-CoV-2 viral proteins. The model is trained on multiple datasets and generates a list of top medicines with high binding affinity, potentially useful for developing new medications.
- Shao et al.[21] was presented as DTIGCCN as a novel drug-target interaction prediction model. It combines spectral-based graph convolutional networks and convolutional neural networks to extract refined features and exploit drug-target correlations. Experimental results demonstrate the superiority of this model over traditional approaches.
- Tsubaki and Tomii [22] The study focuses on end-to-end representation learning for compounds and proteins. It proposes a novel approach that integrates a graph neural network for compounds and a convolutional neural network for proteins. The suggested strategy performs competitively or better than previous approaches, even on imbalanced datasets.
- Ranjan et al.[23] The researchers aim to create highly active chemicals that bind to the protein structure of SARS-CoV-2. They develop a framework using a Gated Graph Neural Network (GGNN), knowledge graphs, and early fusion. The framework successfully generates highly precise unique compounds, tested on viral proteins RdRp and 3CLpro.
- Wen et al.[24] introduced a Deep DTIs as a deep-learning-based algorithmic framework for identifying drug-drug interactions and target interactions in drug repositioning. The model utilizes unsupervised pretraining and achieves competitive performance compared to other state-of-the-art methods. Drug and target data are obtained from the Drug Bank database.
- Wen et al.[25] presented a graph-convolutional framework for predicting protein-ligand interactions. The authors utilize unsupervised graph auto-encoders and demonstrate that Graph-CNNs can capture protein-ligand binding interactions without relying on target-ligand co-complexes. The model outperforms or matches other ligand-scoring methods on benchmark datasets.
- Zhao et al. [26] proposed a DPP network to capture the relationships between drug-protein pairs (DPPs) in drug-target interaction modeling. The proposed framework, GCN-DTI, utilizes graph convolutional networks and deep neural networks to predict DTIs.

Materials and Methods

The general proposed system diagram presented in Fig. 2 provides a holistic view of the drug-target interaction prediction pipeline, showcasing the integration of deep learning models with input data and the overall workflow. This system architecture forms the basis for the subsequent detailed explanation of the specific deep learning models and their contributions to predicting drug-target interactions.

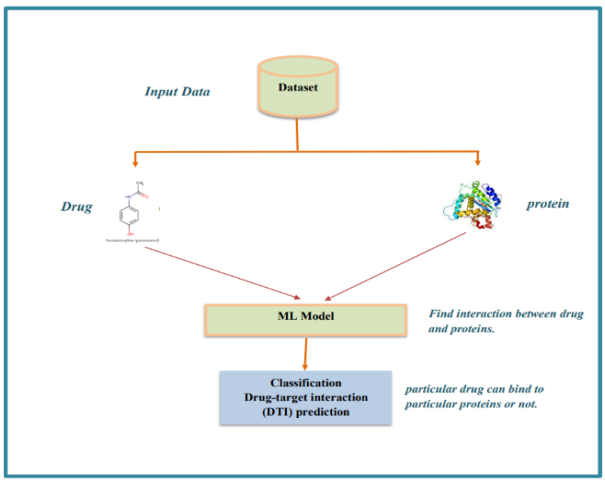

Figure 2. Methodology of the proposed system

2024, 21(11): 3605-3616
https://doi.org/10.21123/bsj.2024.9212
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

3.1 Davis dataset

The Davis dataset is a choice, for research on predicting interactions between drugs and targets. It includes an amount of data with over 72,000 measurements of binding affinity between 442 drug compounds and 73 target proteins. Researchers often use this dataset to test their models due to its size and the variety of compounds and targets it covers [27].

Moreover the Davis dataset is frequently employed in deep learning approaches for predicting drug target interactions. It has also been utilized to create models that forecast the binding affinity between drugs and targets a task that's more complex than classification. In essence the Davis dataset serves as a tool for scientists investigating drug target interactions and binding affinity prediction playing a role in advancing various successful deep learning models, in this field.

3.2 Dataset Preprocessing

3.2.1 SMILES and Protein Sequences Encoding

In SMILES, the labels or unique letters represent atoms, bonds, and other molecular features. The specific set of labels used in SMILES can vary depending on the context and the molecules being represented. While in protein, there are a total of 20 standard amino acids that are commonly found in proteins. Each amino acid is represented by a unique letter or symbol. However, if there are additional categories, such as ambiguous or non-standard amino acids, post-translational modifications, or special symbols used to denote specific features, the number of categories may exceed 25.

These 20 amino acids form the basis of protein sequences, and each amino acid contributes to the structure and function of proteins in different ways. It's important to note that if you are considering a specific set of protein sequences or a specific domain of proteins, additional non-standard amino acids or symbols may represent unique features within those sequences.

Steps to SMILES and protein sequences encoding as following:

- Tokenize each SMILES or amino acid as a symbol
- Builds the vocabulary and assigns a unique index to each symbol.
- Replaces each symbol with its corresponding index in the vocabulary.

create token dictionaries for amino acids and SMILES

```
amino_acid_dict = {
    'A': 0, 'C': 1, 'D': 2, 'E': 3, 'F': 4, 'G': 5, 'H': 6, 'I': 7,
    'K': 8, 'L': 9, 'M': 10, 'N': 11, 'P': 12, 'Q': 13, 'R':
    14, 'S': 15, 'T': 16, 'V': 17, 'W': 18, 'Y': 19}

smiles_dict = {'#': 0, '(': 1, ')': 2, '+': 3, '-': 4, '.': 5, '/':
    6, '1': 7, '2': 8, '3': 9, '4': 10, '5': 11, '6': 12, '7': 13,
    '8': 14, '9': 15, '=': 16, 'B': 17, 'C': 18, 'F': 19, 'H':
    20, 'I': 21, 'K': 22, 'N': 23, 'O': 24, 'P': 25, 'S': 26,
    '[': 27, '\\': 28, ']': 29, 'c': 30, 'n': 31, 'o': 32, 's': 33}
```

3.2.2 Convert SMILES to Graph

This manner converts SMILES (Simplified Molecular Input Line Entry System) representations of molecules into graph structures to perform graph-based deep learning. The general procedure used in the manner to convert SMILES to graphs:

1. Parsing SMILES: The example uses the RDKit library, a powerful cheminformatics toolkit, to parse the SMILES representations and obtain a molecule object. RDKit provides functions to analyze and manipulate molecular structures.
2. Atom Mapping: After obtaining the molecule object from SMILES, the example maps each atom to a unique index. It creates a dictionary where each atom is a key, and the corresponding value is an integer index.
3. Atom Attributes; The provided information outlines characteristics related to atoms, within the molecule. These attributes encompass details such as the number, hybridization status, hydrogen count and other relevant properties. Incorporating these attributes enables the inclusion of details, within the graphical depiction.
4. Bond Mapping: Each bond connecting atoms is assigned an index, for reference. Similar to atom assignment a dictionary is established where each bond serves as a key linked to an identifier.
5. Bond Features: Additionally the process involves extracting attributes for every bond within the molecule. These features may encompass details such as the type of bond

2024, 21(11): 3605-3616
https://doi.org/10.21123/bsj.2024.9212
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

(triple) ring status, stereochemistry and other relevant information. By encapsulating these characteristics the connectivity and properties of the chemical bonds are encoded.

6. Graph Construction: Utilizing the assigned atom and bond indices a visual representation of the molecule in graph form is generated. Nodes are created to represent atoms while edges denote the bonds between them incorporating features accordingly. The resulting graph visually depicts the arrangement with atoms represented as nodes and bonds as edge.

7. Graph Neural Network (GNN) Input: The constructed graph structure serves as input for the MPNN based model implementation. MPNNs harness the graph topology. Employ message propagation techniques to disseminate information throughout the nodes and edges of the graph structure. This mechanism enables the model to capture relationships and interactions, within the structure involving atoms and bonds.

In general, the procedure includes analyzing SMILES assigning indices, to atoms and bonds extracting characteristics for atoms and bonds creating a depiction and applying the graph for learning using MPNN. This enables learning based on graphs, for molecular information.

The Proposed Systems

Model 1: BiGRU-DTA

The BiGRU DTA model is a network design created to predict the relationship, between drug compounds and target proteins. It utilizes a gated recurrent unit (Bi GRU) to extract features from both protein and drug sequences, followed by a connected neural network for the final forecast as shown in Fig. 3.

The Bi GRU network, a type of network (RNN) effectively captures sequential patterns in input sequences by analyzing data in both forward and backward directions. It consists of two layers of GRU units with one processing the input sequence in a manner and the other in reverse. The results from these layers are. Passed on to the next layer.

In this model separate Bi GRU networks handle the drug and protein sequences independently extracting features from them in both directions. The outputs from these networks are. Fed into the connected neural network for accurate prediction of binding affinity.

This approach offers benefits compared to methods. Firstly the Bi GRU network is adept at capturing long distance relationships within input sequences, which's crucial for predicting binding affinities. Additionally it is adaptable to varying sequence lengths making it beneficial for datasets, with sequence sizes.

Finally the Bi GRU network has parameters compared to CNNs, which helps reduce the risk of overfitting especially when working with datasets. In general this alternative approach using Bi GRU networks shows promising results in forecasting the binding affinity, between drug compounds and target proteins emerging as a substitute for the CNN based method. The parameter settings of the model for predicting interactions between drugs and targets are described in Table 1. These parameters play a role, in determining the models efficiency and accuracy. The table below presents the ranges and specific values assigned to each parameter during the experimentation phase:

Table 1. Parameter Settings for Model-1

Parameters	Range
Number of filters	32*1; 32*2; 32*3
Epoch	100
Hidden neurons	1024; 1024; 512
Batch size	256
Dropout	0.1
Optimizer	Adam
Learning rate (lr)	0.001

2024, 21(11): 3605-3616
https://doi.org/10.21123/bsj.2024.9212
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

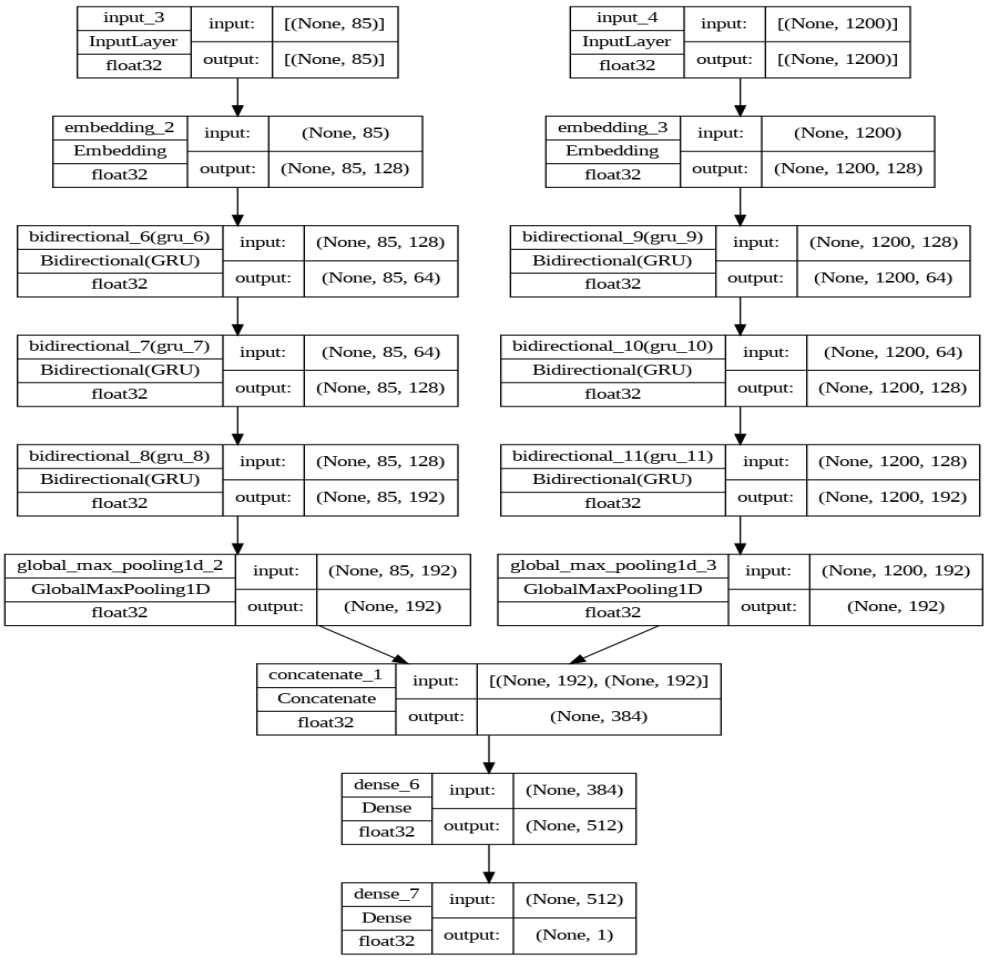

Figure 3. Model-1 with two Bi-GRU blocks to learn from compound SMILES and protein sequences.

Model2: MPNN-GRU-DTA

The model includes two network components; one, for dealing with drug data called a message passing neural network (MPNN) and another for handling protein data known as a bidirectional gated recurrent unit (Bi GRU) as shown in Figure 4. This approach effectively captures the characteristics and trends to each type of information.

The MPNN part is designed to process details about drugs in a graph format representing the structure of each drug. By using the MPNN design the model can. Encode structural aspects of drug molecules capturing key features that influence how they interact with target proteins. On the hand the Bi GRU component focuses on working with sequences of protein data. This type of network effectively identifies patterns within protein sequences by examining them in both forward and backward directions.

After analyzing drug and protein data the model combines the results from both the MPNN and Bi GRU networks. This merging allows the model to leverage a combination of information from both the properties of drugs and the sequential patterns found in proteins. The combined output then goes through layers, including connected layers to produce the final prediction for interactions, between drugs and target proteins.

These layers are designed to understand patterns and representations, from the combined features ultimately making predictions about how a particular drug interacts with its target protein. In general this model structure, which uses MPNN for drugs and Bi GRU for proteins while combining their results presents a method, for foreseeing Drug Target interactions. By capturing both the sequential elements of the information the model leverages the advantages of these two types of data to improve prediction accuracy in the challenging task of predicting DTI.

2024, 21(11): 3605-3616
https://doi.org/10.21123/bsj.2024.9212
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

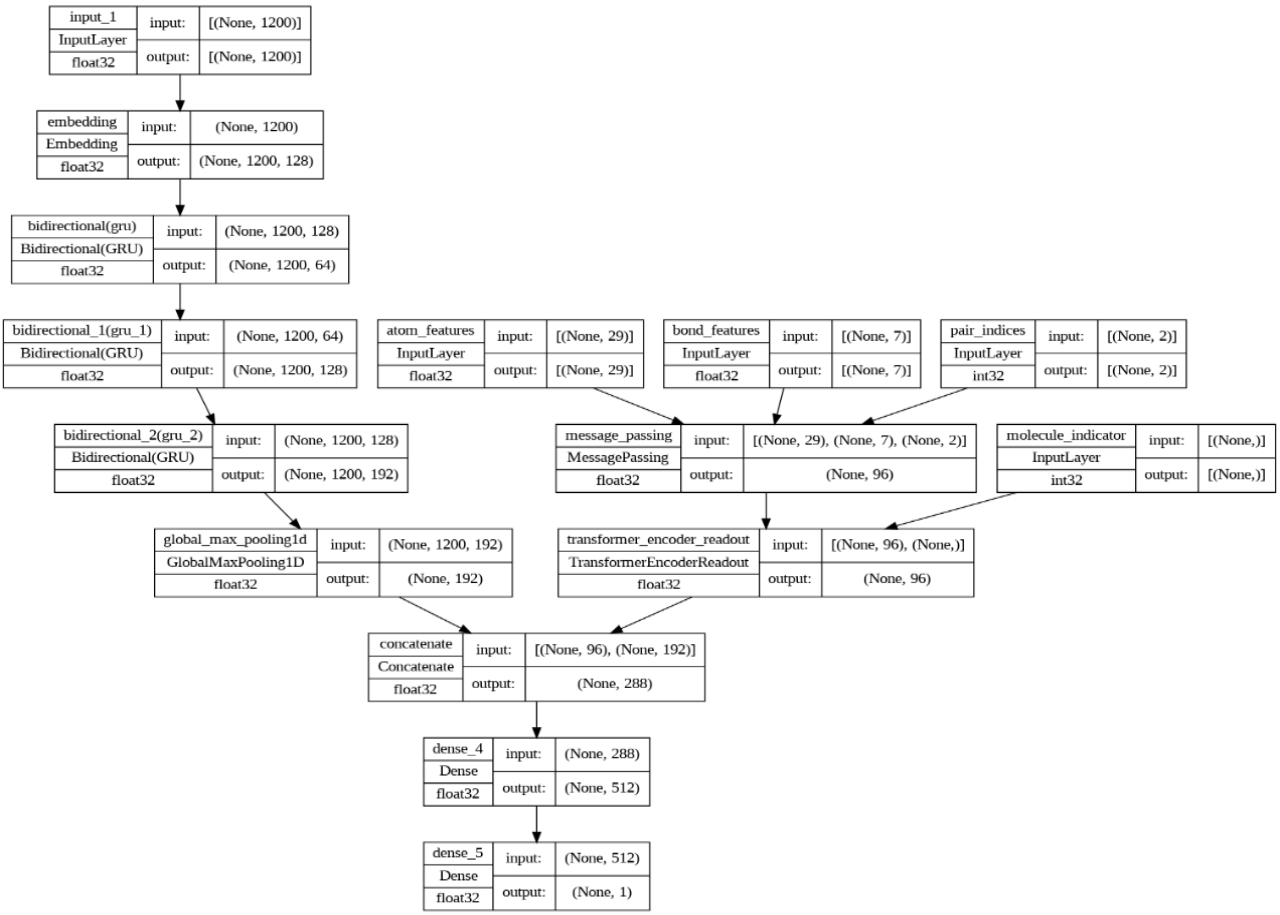

Figure 4. Model-2 with Bi-GRU to learn from protein sequences and MPNN for compound Graph.

Results and Discussion

This paper introduced two deep-learning models. The first model employs two BI-GRU blocks to acquire representations for drugs and targets by considering their sequences. The performance of these models was evaluated by different training sets for the Davis dataset, and the results are presented in Table 2 and Fig. 5, showing the average MSE scores over the independent test set for all six models trained with identical parameters.

The second model employed two distinct neural network components: MPNN for processing drugs and Bi-GRU for processing proteins. This approach effectively captures the unique characteristics and patterns associated with each data type. Tables 3 and Fig. 6 report the average MSE scores over the independent test set of the seven models trained with the same parameters using the seven different training sets for the Davis dataset.

Table 2. The average MSE scores of the test set for model-1

	Proteins	**Compounds**	**MSE**
KronRLS	S–W	Pubchem Sim	0.379
SimBoost	S–W	Pubchem Sim	0.282
DeepDTA	S–W	Pubchem Sim	0.608
DeepDTA	CNN	Pubchem Sim	0.419
DeepDTA	S–W	CNN	0.420
DeepDTA	CNN	CNN	0.261
The Proposed Method	**Bi-Gru**	**Bi-Gru**	**0.229**

2024, 21(11): 3605-3616
https://doi.org/10.21123/bsj.2024.9212
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

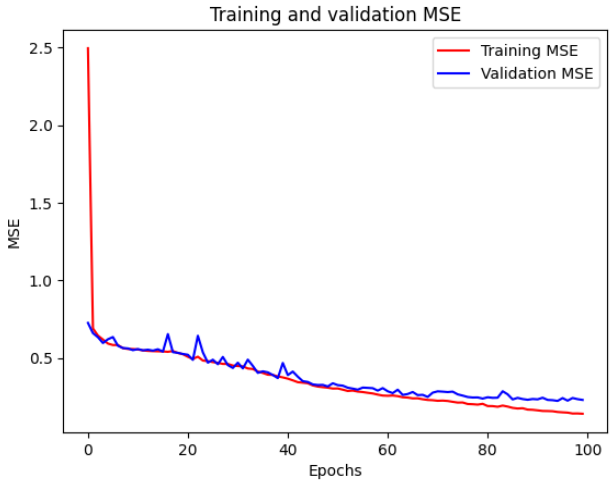

Figure 5. Training and validation MSE for model-1

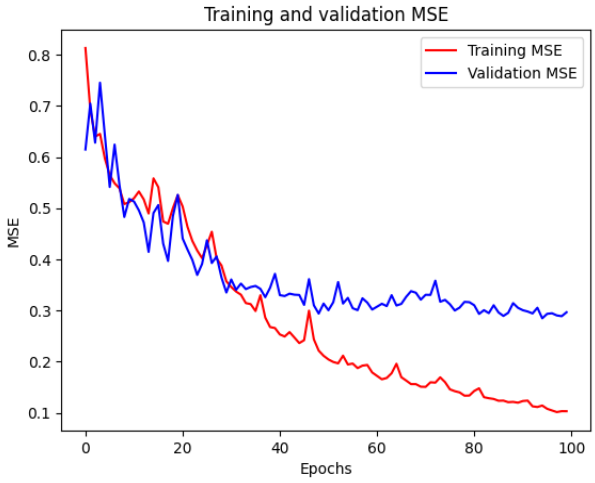

Figure 6. Training and validation MSE for model-2

Table 3. The average MSE scores of the test set for model-2

Model	MSE
GANsDTA[24]	0.276
DeepGS	0.252
GraphDTA (GCN)	0.254
GraphDTA (GATGCN)	0.245
GraphDTA (GAT)	0.232
GraphDTA (GIN)	0.229
DeepGLSTM	0.232
The Proposed Method	**0.202**

The selection of the dataset used in each study plays a crucial role in determining the model's performance [28]. The availability of additional diverse and larger datasets could further enhance the generalizability and robustness of our models also Future studies should assess the transferability and generalization capabilities of our models to new data points. Drug-target interactions are influenced by a myriad of biological factors, including protein structure, ligand binding sites, and cellular environments.

Finally Our study focused on the computational prediction of drug-target interactions; however, experimental validation is critical before applying any predicted interactions in a clinical setting. The absence of direct experimental validation is a limitation, and future research should consider experimental verification to confirm the predicted interactions' accuracy.

Conclusion

In conclusion, this study introduces two deep-learning models to enhance the prediction accuracy of drug-target interactions. The first model utilizes two BI-GRU blocks to effectively learn representations for drugs and targets based on their sequential characteristics. Through evaluation of the independent test set of the Davis dataset, the performance analysis, as shown in Table 2 and Fig. 5, demonstrates that our proposed model achieves a superior average mean squared error (MSE) score of 0.2295. This indicates its efficacy in predicting drug-target interactions compared to alternative models, such as KronRLS, SimBoost, and DeepDTA, which exhibit comparatively higher MSE scores.

Moreover, the second model in this study incorporates two distinct neural network components, specifically MPNN for drug processing and Bi-GRU for protein processing. Our model successfully captures important features by leveraging the unique characteristics and patterns associated with each data type, leading to improved predictive capabilities. The data shown in Table 3 and Figure 6 emphasize the MSE scores, on the test set of the Davis dataset for seven different models. Our new approach achieves a MSE score of 0.2846 showing its effectiveness in predicting drug target interactions when compared to advanced models like

GANsDTA, DeepGS and various GraphDTA designs.

Overall the results of this research highlight the importance of using learning models to enhance our knowledge and forecasting of drug target interactions. Our proposed models, featuring BI GRU and MPNN Bi GRU architectures demonstrate performance in capturing the connections, between drugs and targets. These models present opportunities to expedite drug discovery and development processes by providing predictions of drug target interactions. The results obtained contribute to the field by advancing the knowledge and potential applications of deep learning techniques in pharmaceutical research.

Author's Declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images that are not ours have been included with the necessary permission for re-publication, which is attached to the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee at University of Babylon.
- Ethics statement:
  No animal studies are present in the manuscript.
  No human studies are present in the manuscript.
  No potentially identified images or data are present in the manuscript.

Author's Contribution Statement

B. N. D. Suggesting a research project and preparing a research plan. A. K. A. Executing and developing the models, obtaining the results, and presenting them to the supervisor.

References

1. Patel L, Shukla T, Huang X, Ussery DW, Wang S. Machine Learning Methods in Drug Discovery. Molecules. 2020; 25(22): 5277. https://doi.org/10.3390/molecules25225277.
2. Dhannoon BN. Predication and Classification of Cancer Using Sequence Alignment and Back Propagation Algorithms inBrca1 and Brca2 Genes. Int J Pharm Res 2019; 11. https://doi.org/10.31838/ijpr/2019.11.01.062.
3. Jassim OA, Abed MJ, Saied ZHS. Deep Learning Techniques in the Cancer-Related Medical Domain: A Transfer Deep Learning Ensemble Model for Lung Cancer Prediction. Baghdad Sci J. 2023; http://dx.doi.org/10.21123/bsj.2023.8340
4. Duelen R, Corvelyn M, Tortorella I, Leonardi L, Chai YC, Sampaolesi M. Medicinal Biotechnology for Disease Modeling, Clinical Therapy, and Drug Discovery and Development. Int to Biotech Ent 2019;89–128: http://dx.doi.org/10.1007/978-3-030-22141-6_5
5. Klebe G. Virtual ligand screening: strategies, perspectives and limitations. Drug Discov Today. 2006; 11(13-14): 580-594. https://doi.org/10.1016/j.drudis.2006.05.015.
6. Dickson M, Gagnon JP. Key factors in the rising cost of new drug discovery and development. Nat Rev Drug Discov. 2004; 3(5): 417-429. https://doi.org/10.1038/nrd1389.
7. Kapetanovic I. Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach. Chem Biol Interact. 2008; 171: 165-176. https://doi.org/10.1016/j.cbi.2007.11.011.
8. Sachdev K, Gupta MK. A comprehensive review of feature based methods for drug target interaction prediction. J Biomed Inform. 2019; 93: 103159. https://doi.org/10.1016/j.jbi.2019.103159.
9. Pliakos K, Vens C. Drug-target interaction prediction with tree-ensemble learning and output space reconstruction. BMC Bioinform. 2020; 21: 49. https://doi.org/10.1186/s12859-020-3380-6.
10. Shin B, Park S, Kang K, Ho JC. Self-Attention Based Molecule Representation for Predicting Drug-Target Interaction. Proceedings of the Machine Learning for Healthcare Conference, MLHC 2019, Ann Arbor, MI, USA, 2019. PMLR 2019; 106: 230-248. https://doi.org/10.48550/arXiv.1908.06760.
11. Wang L, You ZH, Xing Chen, Shi-Xiong Xia, Feng Liu, Xin Yan, et al. A Computational-Based Method for Predicting Drug-Target Interactions by Using Stacked Autoencoder Deep Neural Network. J Comput Biol. 2018; 25(4): 361-373. https://doi.org/10.1089/cmb.2017.0135
12. Beck BR, Shin B, Choi Y, Park S, Kang K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. Comput

Struct Biotechnol J. 2020; 18: 784-790. https://doi.org/10.1016/j.csbj.2020.03.025
13. Nguyen T, Le H, Quinn TP, Nguyen T, Le TD, Venkatesh S. GraphDTA: Predicting drug-target binding affinity with graph neural networks. J. Bioinform. 2021; 37(7): 1140-1147. https://doi.org/10.1093/bioinformatics/btaa921
14. Abdul Raheem K. Ali, Dhannoon N. Ban, Automating Drug Discovery using Machine Learning, Curr Drug Discov Technol 2023; 20 : e070623217776. https://dx.doi.org/10.2174/1570163820666230607163313
15. Ye Q, Zhang X, Lin X. Drug-target interaction prediction via multiple classification strategies. BMC Bioinform 2022 Jan 20; 22(Suppl 12): 461. https://doi.org/10.1186/s12859-021-04366-3
16. de Souza JG, Fernandes MAC, de Melo Barbosa R. A Novel Deep Neural Network Technique for Drug-Target Interaction. Pharmaceutics. 2022; 14(3): 625. https://doi.org/10.3390/pharmaceutics14030625.
17. Shim J, Hong ZY, Sohn I, Hwang C. Prediction of drug-target binding affinity using similarity-based convolutional neural network. Sci Rep. 2021; 11(1): 4416. https://doi.org/10.1038/s41598-021-83679-y
18. Lee I, Nam H. Sequence-based prediction of protein binding regions and drug-target interactions. J Cheminform. 2022;14(1):5. https://doi.org/10.1186/s13321-022-00584-w
19. Öztürk H, Özgür A, Ozkirimli E. DeepDTA: deep drug-target binding affinity prediction. Bioinformatics. 2018; 34(17): i821-i829. https://doi.org/10.1093/bioinformatics/bty593
20. Mukherjee S, Ghosh M, Basuchowdhuri P. Deep Graph Convolutional Network and LSTM based approach for predicting drug-target binding affinity. Proceedings of the 2022 SIAM International Conference on Data Mining (SDM). 2022 ;729–37. http://dx.doi.org/10.1137/1.9781611977172.82
21. Shao K, Zhang Z, He S, Bo XC. DTIGCCN: Prediction of drug-target interactions based on GCN and CNN. EEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI) 2020. https://doi.org/10.1109/ictai50040.2020.00060.
22. Tsubaki M, Tomii K, Sese J. Compound-protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. J. Bioinform. 2019; 35(2): 309-318. https://doi.org/10.1093/bioinformatics/bty535.
23. Ranjan A, Shukla S, Datta D, Misra R. Generating novel molecule for target protein (SARS-CoV-2) using drug-target interaction based on graph neural network. Netw Model Anal Health Inform Bioinform. 2022; 11(1): 6. https://doi.org/10.1007/s13721-021-00351-1
24. Wen M, Zhang Z, Niu S, Sha H, Yang R, Yun Y, et al . Deep-Learning-Based Drug–Target Interaction Prediction. J Proteome Res. 2017; 16(4): 1401-1409. https://doi.org/10.1021/acs.jproteome.6b00618.s001
25. Wen T, Altman RB. Graph Convolutional Neural Networks for Predicting Drug-Target Interactions. J Chem Inf Model. 2019; 59(10): 4131-4149. https://doi.org/10.1021/acs.jcim.9b00628
26. Zhao T, Hu Y, Valsdottir LR, Zang T, Peng J. Identifying drug-target interactions based on graph convolutional network and deep neural network. Brief Bioinform. https://doi.org/10.1093/bib/bbaa044
27. Davis M.I, Hunt J.P, Herrgard S, Ciceri P, Wodicka L.M, Pallares G, et al. Comprehensive analysis of kinase inhibitor selectivity. Nat Biotechnol. 2011(29): 1046–1051. https://doi.org/10.1038/nbt.1990.
28. Wotaifi Tahseen, Dhannoon Ban. An Effective Hybrid Deep Neural Network for Arabic Fake News Detection. Baghdad Sci J. 2023, 20. https://doi.org/10.21123/bsj.2023.7427

2024, 21(11): 3605-3616
https://doi.org/10.21123/bsj.2024.9212
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

نماذج التعلم العميق للتفاعلات بين الأدوية والبروتينات

علي كريم عبد الرحيم[1,2]، بان نديم ذنون [3]

[1] قسم البرمجيات، كلية تكنولوجيا المعلومات ، جامعة بابل ، بابل ، العراق .
[2] قسم العلوم القرأنية ،جامعة وارث الانبياء ، كربلاء ، العراق .
[3] قسم علوم الحاسوب، كلية العلوم، جامعة النهرين، بغداد، العراق.

الخلاصة

يعتبر منهج تطوير الأدوية بطيئًا ومكلفًا ويحتمل الفشل - يختبر العلماء ملايين المركبات ، لكن القليل منها فقط ينجح في إجراء الاختبارات قبل السريرية أو السريرية. يعتبر التعلم الآلي وهو فرع من الذكاء الاصطناعي ، موضوعًا سريع التوسع تستخدمه العديد من شركات الأدوية بشكل متزايد. يمكن أن يساعد دمج تقنيات التعلم الآلي في عملية تطوير الأدوية في أتمتة عمليات المعالجة والتحليل المتكررة للبيانات ، ويمكن استخدام تقنيات ML في عدة مراحل من تطوير الأدوية ، بما في ذلك التفاعلات الدوائية المستهدفة. وهي عنصر مهم في عملية تطوير الدواء. عندما يرتبط عقار (جزيء كيميائي) بهدف (بروتينات أو أحماض نووية) ، يُقال إنه يرتبط ، ويغير سلوكه / وظيفته البيولوجية ، ويعيده إلى طبيعته. يُعد تنبؤ DTI جزءًا أساسيًا من عملية DD لأنه قد يؤدي إلى تسريع التكاليف وتقليلها ، ولكنه يمثل تحديًا ومكلفًا لأن المقايسات التجريبية لا تستغرق وقتًا طويلاً فحسب ، بل تكون باهظة الثمن أيضًا. أظهرت الأساليب القائمة على التعلم العميق نتائج مشجعة في التنبؤ بالتفاعلات المستهدفة للأدوية (DTI) في السنوات الأخيرة. في هذه الورقة ، نقترح هيكلين جديدين للتعلم العميق للتنبؤ بالتفاعلات المستهدفة للدواء (DTI). يستخدم النموذج الأول الشبكات العصبية لتمرير الرسائل (MPNN) لتشفير الأدوية والوحدات المتكررة ذات البوابات ثنائية الاتجاه (Bi-GRU) لتشفير البروتين. بينما يستخدم النموذج الثاني Bi-GRU لتشفير الأدوية وترميز البروتين. تم تدريب النموذجين وتقييمهما على مجموعة بيانات معيارية. توضح نتائجنا أن نماذجنا تتفوق في الأداء على أحدث أساليب التنبؤ DTI وهي نهج واعد للتنبؤ ب DTI بدقة عالية.

الكلمات المفتاحية: Bi-GRU ، التعلم العميق ، تفاعلات الهدف الدوائي ، اكتشاف الأدوية ، النماذج الحسابية للتنبؤ.