



DTI-BERT: Identifying Drug-Target Interactions in Cellular Networking Based on BERT and Deep Learning Method

Jie Zheng, Xuan Xiao* and Wang-Ren Qiu*

Computer Department, Jing-De-Zhen Ceramic Institute, Jing-De-Zhen, China

OPEN ACCESS

Edited by:

Juexin Wang,
University of Missouri, United States

Reviewed by:

Yang Liu,
Dana-Farber Cancer Institute,
United States
Xing Chen,
China University of Mining and
Technology, China

*Correspondence:

Xuan Xiao
jdzxiaoxuan@163.com
Wang-Ren Qiu
qiuone@163.com

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

Received: 21 January 2022

Accepted: 25 April 2022

Published: 08 June 2022

Citation:

Zheng J, Xiao X and Qiu W-R (2022)
DTI-BERT: Identifying Drug-Target
Interactions in Cellular Networking
Based on BERT and Deep
Learning Method.
Front. Genet. 13:859188.
doi: 10.3389/fgene.2022.859188

Drug-target interactions (DTIs) are regarded as an essential part of genomic drug discovery, and computational prediction of DTIs can accelerate to find the lead drug for the target, which can make up for the lack of time-consuming and expensive wet-lab techniques. Currently, many computational methods predict DTIs based on sequential composition or physicochemical properties of drug and target, but further efforts are needed to improve them. In this article, we proposed a new sequence-based method for accurately identifying DTIs. For target protein, we explore using pre-trained Bidirectional Encoder Representations from Transformers (BERT) to extract sequence features, which can provide unique and valuable pattern information. For drug molecules, Discrete Wavelet Transform (DWT) is employed to generate information from drug molecular fingerprints. Then we concatenate the feature vectors of the DTIs, and input them into a feature extraction module consisting of a batch-norm layer, rectified linear activation layer and linear layer, called BRL block and a Convolutional Neural Networks module to extract DTIs features further. Subsequently, a BRL block is used as the prediction engine. After optimizing the model based on contrastive loss and cross-entropy loss, it gave prediction accuracies of the target families of G Protein-coupled receptors, ion channels, enzymes, and nuclear receptors up to 90.1, 94.7, 94.9, and 89%, which indicated that the proposed method can outperform the existing predictors. To make it as convenient as possible for researchers, the web server for the new predictor is freely accessible at: <https://bioinfo.jcu.edu.cn/dtibert> or <http://121.36.221.79/dtibert/>. The proposed method may also be a potential option for other DTIs.

Keywords: drug-target interactions, bidirectional encoder representations from transformers, BRL block, convolutional neural network, computational methods

1 INTRODUCTION

In the process of drug development, there are many important drug-related interaction directions, including drug-protein, drug-miRNA, drug-disease, drug-drug, etc. Small molecule therapeutic drugs typically exert their effects through binding to one or a few protein targets (Dubach et al., 2014; Lim et al., 2021), therefore identifying drug-protein interaction is an important part of genomic drug discovery (Yamanishi et al., 2014). Besides, several studies have indicated that although ncRNAs lack the potential to encode proteins, they play important roles in cellular functions, and their

deregulation heavily contributes to various pathological conditions. Among them, miRNAs are promising therapeutic targets for complex diseases (Wang and Chen, 2019; Yin et al., 2019; Zhou et al., 2020), it thus becomes important to understand the relationship between ncRNAs and drug targets, what's more, several databases and studies are actively promoting development (Chen et al., 2017). Drug-disease and drug-drug interaction play a crucial role in drug relocation, often serving as important information other than drug-target protein pairing and mainly based on a processing framework called a heterogeneous network. Qu et al. developed a novel computational model of HeteSim-based inference for SM-miRNA Association prediction by implementing a path-based measurement method of HeteSim on a heterogeneous network combined with known miRNA-SM associations, integrated miRNA similarity, and integrated SM similarity (Qu et al., 2019). Jin et al. combine drug features from multiple drug-related networks, and disease features from biomedical corpora with the known drug-disease association's network to predict the correlation scores between drug and disease (Qu et al., 2019). Drug-protein interactions play a key role in the field of biochemistry due to their scientific significance in drug discovery. This paper focuses on the identification of drug-protein interactions.

Drugs modulate the biological functions of proteins by interacting with target proteins, such as ion channels, nuclear receptors, enzymes, and G Protein-coupled receptors (GPCRs). For an in-depth understanding of the functions of drugs, the knowledge of their target protein is indispensable. Despite the substantial effort, only a few DTIs have been identified so far, since the experimental determination of drug-target interactions remains some defects, such as expensive, time-consuming, low accuracy, and so on (Haggarty et al., 2003). It is highly demanded to develop powerful computational tools, which are capable of detecting potential DTIs. Computational prediction of DTIs has emerged for 20 years as a research hotspot, which is not only for better understanding of the molecular mechanism of drug side effects but also for inventing new genomic drugs and identifying new targets for existing drugs (Wang et al., 2010; Kotlyar et al., 2012).

Knowledge of genomic space and chemical space is indispensable for identifying DTIs. With the coming of the post-genome era and the emergence of molecular medicine, transcriptome, and chemical compound, the rapidly increasing knowledge in the field of genomic space and chemical space enables researchers to study drug-target interaction problems (Dobson, 2004) on the basis of high-throughput experimental projects. Several different professional databases have been established, such as Drug Bank, which is consist of two parts information involving drug data and drug target information (Wishart et al., 2018); Therapeutic Target Database (TTD) provides comprehensive information about the drug resistance mutations, gene expressions, and target combinations data (Qin et al., 2014); BindingDB a public database of protein-ligand binding affinities (Liu et al., 2007); Kyoto Encyclopedia of Genes and Genomes (KEGG) including experimental knowledge on protein and their drug target, etc. These resources provide important materials for researchers to

predict drug-target interactions based on computational methods, it is time to develop more integrative approaches capable of taking genomic space, chemical space, and the available known drug-target network information into account simultaneously for the issue.

The development of identifying DTIs followed four main directions for research. Firstly, the most direct method is to use the docking simulation (Pujadas et al., 2008; Morris et al., 2009), which is a process of scoring favorable intermolecular interactions, the three-dimensional (3D) structures of proteins and chemical compounds are indispensable. With the development of techniques (e.g., X-ray crystallography, nuclear magnetic resonance), the rate of 3D protein structure determination is increasing every year, however, it is still not able to keep up with the exponential growth of sequence discovery, such as the PDB database only covers a small fraction of the ion channels and GPCRs, both are considered as the most pharmaceutically useful drug targets. Some programs and webservers provide the prediction of the protein structure, in practice, structure prediction is still relatively immature, and interaction prediction may be affected by the inaccurate structure. Secondly, based on the fact that similar molecules usually bind to similar proteins, it is most straightforward to apply the ligand-based approach (Keiser et al., 2007), for example, conducting Quantitative Structure-Activity Relationship (QSAR) studies that a new ligand can be categorized and compared to known proteins ligands. However, ligand-based approaches often present unreliable results due to available binding ligands of targets' insufficient number, and difficult to scientifically set thresholds to divide positive and negative samples (Butina et al., 2002). Thirdly, literature text mining could be used to extract DTIs from the related articles (Zhu et al., 2005), but this approach could not be used for new drugs and proteins. Fourthly, to overcome the drawbacks of the above-mentioned traditional approaches, chemogenomic approaches are universally studied directions. Chemogenomic approaches integrate information of chemical space, genomic space, and known drug-target interactions, which provide an architecture for deep learning approaches.

Chemogenomic approaches can be classified into three categories: graph-based approaches (Chen et al., 2012), network-based approaches (Alaimo et al., 2013), and learning-based approaches (Mousavian and Masoudi-Nejad, 2014). In the graph-based approach, drugs and targets are represented with graphs, in which nodes for chemical elements or amino acids and adjacency matrices for edges between nodes, adjacency matrices including atom/bond or residue/bond information (Lim et al., 2021). Drug and target graphs can be fed into Graph Neural Network (GNN); after a set of training iterations, information learned by Graph Convolutional Network (GCN) can be converted into vectors for DTIs prediction. Torng and Altman proposed a graph-convolutional framework to determine the interaction patterns (Torng and Altman, 2019). Karlov et al. used the message passing neural network to overcome the limitation of graph convolutional network by considering both nodes and edges (Karlov et al., 2020). Furthermore, the self-attention mechanism in Neural Networks is often coupled with

Graph convolutional network to predict DTIs better. But some research showed that there are difficulties in predicting the local non-covalent interactions between drugs and proteins (Li et al., 2020). Network-based approaches utilized the DTI network of identified edges between drugs and targets to identify new DTIs. Indeed, by constructing a heterogeneous network that includes information on drugs, proteins, diseases, and side-effects, the DTINet method can improve the accuracy of DTIs prediction (Luo et al., 2017), but the learning model only takes relatively simple log-bilinear functions, obtaining features may not be the inherent representations of drugs or targets for the final DTI prediction task (Wan et al., 2019). Supervised learning-based approaches are classified into similarity-based approaches and feature-based approaches (Chen et al., 2018). Similarity-based approaches generate the similarity matrixes for drugs and targets respectively, via various similarity measurement strategies such as chemical-based similarity (Haggarty et al., 2003), pharmacological-based similarity (Kim et al., 2013), therapeutic-based similarity, and drug-drug interaction similarity for drugs, and sequence-based similarity (Yamanishi et al., 2008), functional-bases similarity, protein-protein interaction similarity for targets. These similarity matrixes have been used in bipartite local models (Mei et al., 2013), matrix factorization models (Ezzat et al., 2016), and the nearest neighbor methods (Zhang et al., 2016) to predict DTIs. The feature-based approaches extract more useful information from protein sequences and drug chemical structure, via the adequate support offered by the rapid development of algorithms.

Predicting DTIs with machine learning algorithms has recently become the focus of research. There are 1-D, 2-D, and 3-D representations of drugs (Rognan, 2007). Simplified Molecular Input Line Entry System (SMILES) string is a typical 1-D representation of the drug (Öztürk et al., 2016) that are commonly used descriptors (Kombo et al., 2013; Sawada et al., 2014). For targets, the sequences of protein are encoded by the physicochemical properties of amino acids, sequential evolution information formulation and general form of pseudo amino acid composition (Li et al., 2020). Lastly, machine learning algorithms are applied for decision-making. Recently, Wang et al. used a novel bag-of-words model and discrete Fourier transform to extract target sequence feature and molecular fingerprint pattern information, respectively, and then use a distance-weighted K-nearest-neighbor algorithm as a predictor (Wang et al., 2020). This paper motivates our work, that instead of using amino acid physicochemical properties to encode words and perform clustering, we can vectorization drugs and protein by using advanced methods such as word2vec and ProtBert (Elnaggar et al., 2021), which could map every word (amino acids are regarded as words) into the latent vector space where the geometric relationship can be used to characterize the semantic relationship between the words. And based on the present situation of identifying DTIs by the way of investigating a series of recently published articles (Keiser et al., 2007; Ezzat et al., 2016; Zhang et al., 2016) as well as some review papers (Rognan, 2007; Kombo et al., 2013; Öztürk et al., 2016), we have proposed a novel feature-based computational model for predicting drug-target interactions to enhance prediction

performance. The novelty of this proposed work 1) Compared with the end-to-end predictor, we treat DTIs task more flexibly. The protein sequences are regarded as natural language and vectorized by the state-of-art ProtBert model, and drug molecular is transformed by DWT, which is commonly used in signal processing. 2) Calculating the hybrid loss function (contrastive loss and cross-entropy loss), which can make the samples of the same interaction label closer, and the distance between different labels as far as possible and help the predictor achieve higher accuracy.

2 MATERIALS AND METHODS

2.1 Benchmark Dataset

Identifying DTIs can be regarded as a supervised prediction task to predict whether a pair of counterparts interact with each other or not in the drug-target networks. In this study, the benchmark dataset was taken from (He et al., 2010). There are mainly two reasons, 1) The information about the DTIs was collected from the DrugBanks, BRENDA, SuperTarget, and KEGG BRITE databases, which included four main drug target proteins of G Protein-coupled receptors (GPCR), enzymes (Ezy), ion channels (Chl), and nuclear receptors (NR). 2) In recent years, many researchers have been proposed to predict DTIs, which are based on this benchmark dataset, and hence will facilitate the comparison under the same condition. It can be summarized as follows:

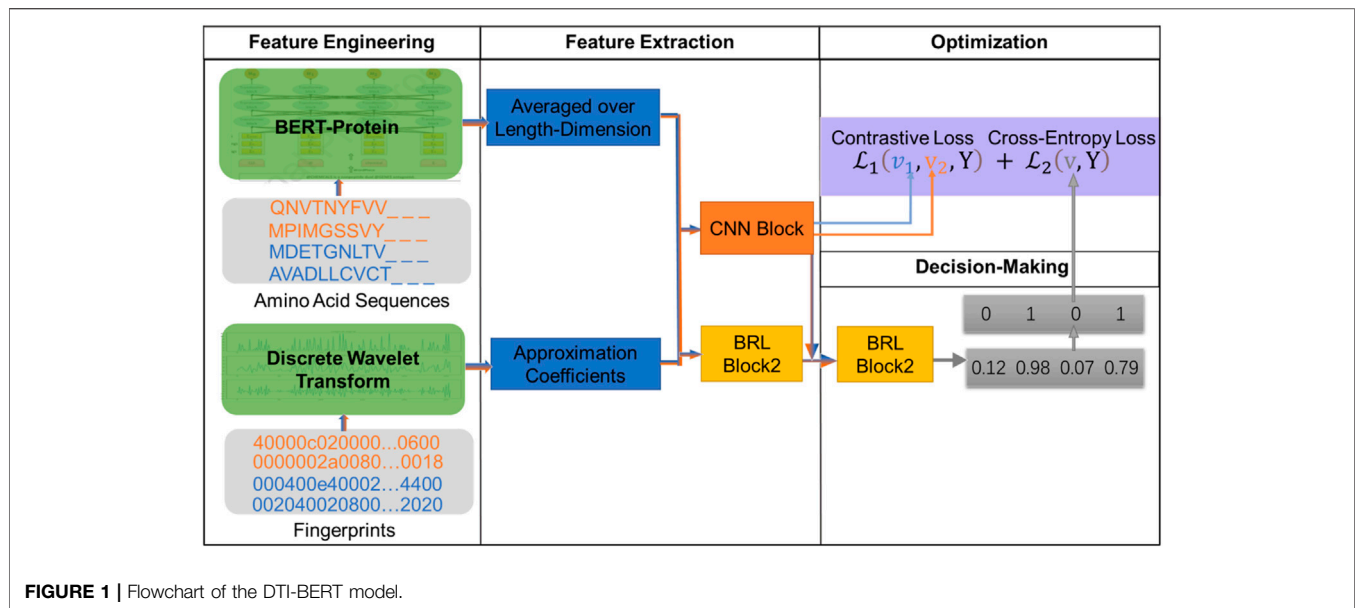
$$\begin{cases} \mathbb{S} = \mathbb{S}_{GPCR-Drug} + \mathbb{S}_{Chl-Drug} + \mathbb{S}_{Ezy-Drug} + \mathbb{S}_{NR-Drug} \\ \mathbb{S}_{GPCR-Drug} = \mathbb{S}_{GPCR-Drug}^+ (630) + \mathbb{S}_{GPCR-Drug}^- (1240) \\ \mathbb{S}_{Chl-Drug} = \mathbb{S}_{Chl-Drug}^+ (1372) + \mathbb{S}_{Chl-Drug}^- (2744) \\ \mathbb{S}_{Ezy-Drug} = \mathbb{S}_{Ezy-Drug}^+ (2719) + \mathbb{S}_{Ezy-Drug}^- (5438) \\ \mathbb{S}_{NR-Drug} = \mathbb{S}_{NR-Drug}^+ (82) + \mathbb{S}_{NR-Drug}^- (164) \end{cases} \quad (1)$$

There are 4,803 drug-target pairs in positive subsets, 2,719 for enzymes, 1,372 for ion channels, 630 for GPCRs, and 82 for nuclear receptors. Negative samples are randomly synthesized by separating each target and drug in S^+ , and none of them appear in the corresponding positive dataset. The proportion of positive samples and negative samples was set as 1:2. For comparison with previously published papers, both our positive and negative samples are consistent with He et al. (He et al., 2010)

Check390 is a dataset constructed by Hu et al. It contains 130 pairs of positive samples from the KEGG database, and 260 negative samples generated using the above method (Hu et al., 2016). Each pair in Check390 cannot be found in S .

2.2 Framework of the Constructed Model

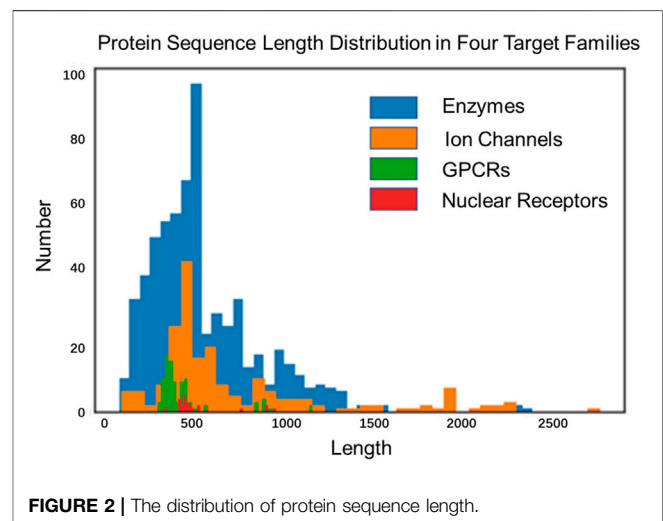
In this article, we construct a novel model for DTIs based on large-scale pre-trained Bidirectional Encoder Representations from Transformers (BERT) and the fully connected neural network-based module called the BRL block. **Figure 1** shows an overview of the DTIs model. The model has four modules: feature engineering, feature extraction, optimization, and decision-making. Firstly, in the feature engineering module, we use the auto-encoder ProtBert model, which is pre-trained on



data from UniRef100 containing 216M protein sequences, to generate embedding vectors for protein sequences. As a result, the proteins can be represented via 1024-D vectors (dimensionality of the features extracted by the ProtBERT model). Drug molecular fingerprints are represented by 128-D vectors through semi decomposition process discrete wavelet transform (DWT). Secondly, the 1152-D vectors (a concatenation of protein sequence feature and drug feature) are fed into the feature extraction model to generate interaction information through the first BRL block and CNN. Afterward, in the decision-making module, the second BRL block is used to map interaction features into a unified vector space. The optimization module contains a contrastive loss and a cross-entropy loss. The contrastive loss is used to calculate the interaction information (generated by CNN block), which can reduce the distance between samples with the same label, and increase the distance between samples with different labels, while the cross-entropy loss is computed as the loss of second BRL block, both are used to adapt weights in the module during the learning process by minimizing the total loss. At the end of model, we can obtain the interaction score (generated by a softmax layer after second BRL block, and range from 0-1), the pair is interaction if the prediction score is > 0.5 .

2.2.1 Feature Extraction From Protein

Recently, many word-embedding methods have been used for protein feature extraction, for example, Zheng et al. identified the ion channel-drug interaction using both word2vec and node2vec as molecular representation learning methods (Zheng et al., 2021). However, there are still imperfect, like in these word-embedding methods may map every word with their unique vector, therefore this representation is context-independent. With the exponential growth of textual data, major progress has been made in the pre-training language representations (Peng et al., 2019; Bianchi et al., 2021). Bidirectional Encoder



Representations from Transformers (BERT) was the first fine-tuning-based representation model (Devlin et al., 2018), which can generate different representations for the same word based on context (Devlin et al., 2018; Nozza et al., 2020).

Almost all sequence-based language models (e.g., context ELMo (Ilić et al., 2018), BERT (Devlin et al., 2018), Xlnet (Yang et al., 2019)) have been promoted the development of processing natural languages successfully, but model architectures and pre-training tasks may not be suitable for representing proteins. The primary reason is that proteins are more variable than sentences in length, and show many interactions in distant positions (due to their 3D structure). The length of English sentences is multiple, usually around 15-30 words (Brandes et al., 2022). Although the length limit of a sentence is not an issue in sentence-level NLP tasks (Dai et al., 2019; Brandes et al., 2022), however, many proteins are more than

20-times longer than nature sentences, reaching an average length of up to 600 residues in drug–the target benchmark dataset and over 20% of the sequences are longer than 1,000. The average length of GPCR, ion channel, enzyme and nuclear receptor are 470, 760, 570 and 540, the distribution of protein sequence length is shown in **Figure 2**.

For protein sequence representation, Elnaggar et al. released a model called ProtBert, which was trained on UniRef100 datasets (contained 216M protein sequences) (Elnaggar et al., 2021). In the ProtBert model, amino acids are set as single words and protein sequences as sentences. The model can deal with protein sequences up to 40k in length, and can download from: <https://github.com/agemagician/ProtTrans> (Elnaggar et al., 2021). In the current study, the protein sequence feature can be extracted by ProtBert based on transfer learning (Lee et al., 2019; Noorbakhsh et al., 2020).

The sequence expressed as an amino acid residue may be formulated in the following format:

$$G = R_1 R_2 R_3 \dots R_L \quad (2)$$

where R_1 is the first residue in the protein sequence, R_2 is the second residue, \dots , R_L is the L -th residue.

The framework of ProtBert is similar to the original Bert publication, some special encoding symbols like [CLS] and [SEP] remain in the BERT model. [CLS] means classification, is added as the first token in the Bert sequence information. When designing the model, [CLS] token was considered as the representation of subsequent text classification. [SEP] means a separator, for example, the task was sentence-pair regression, the input for BERT consists of the two sentences, that would be separated by a special [SEP] token.

We add a [CLS] token at the beginning of the protein sequence marked as R_0 , which acts as an aggregate sequence representation and is usually used for sequence classification tasks in the BERT model, and the [SEP] token at the end of the sequence, marked as R_{L+1} .

We get protein features from the last layer of ProtBert, and every amino acid can be converted to a 1024-dimensional vector B_{R_j} , and the protein can be represented as a feature matrix P_{BERT} :

$$B_{R_j} = [B_{R_j}^1 B_{R_j}^2 \dots B_{R_j}^i \dots B_{R_j}^{1024}] \quad (3)$$

$$P_{BERT} = \begin{bmatrix} B_{R_0}^1 & \dots & B_{R_0}^{1024} \\ \vdots & \ddots & \vdots \\ B_{R_{L+1}}^1 & \dots & B_{R_{L+1}}^{1024} \end{bmatrix} \quad (4)$$

It can be seen from **Eqs. 3, 4** that different protein has different size of P_{BERT} . To formulate the protein sequences with the same size mathematics formulation, the matrix was averaged (mean-pooled) over the vertical axis and a 1024-dimensional vector was obtained to be used as a representation of protein named BERT_Mean:

$$b_n = \frac{\sum_{j=0}^{j=L+1} B_{R_j}^n}{L+2} \quad (1 \leq n \leq 1024) \quad (5)$$

$$P_{PROT} = [b_0 b_2 \dots b_n \dots b_{1024}] \quad (6)$$

2.2.2 Feature Extraction From Drug Molecule

A drug is saved as an MOL file (a file format that represents a compound in the form of a graph connection table) or SMILES in the database, both formats containing information about the molecule structure, and can be retrieved from the KEGG database (<http://www.kegg.jp/kegg/>) or ChEMBL (<https://www.ebi.ac.uk/chembl/>) according to drug IDs. We can also use the MOL file or SMILES as the input of the OpenBabel tool (<http://openbabel.org/>) to generate the molecular fingerprint file, including FP2, FP3, FP4, and MACSS. FP2 is an enumeration of linear fragments or ring substructures of one to seven connected atoms in a molecule, then maps them to a 256-bit hexadecimal string through a hash function. FP3, FP4, and MACSS use predefined structures to generate fingerprints. FP2 retains more sequence information, we use FP2 as molecular input.

The FP2 molecular fingerprint is represented by a 256-bit hexadecimal string, the hexadecimal char “0~F” can be converted to the number 0–15, drug molecule is represented as S_{FP2} in the following formulation:

$$S_{FP2} = [f_1 f_2 \dots f_{256}] \quad (7)$$

In previous studies, the FP2 can be further processed using some transposition functions, and Hu et al. (Hu et al., 2016) and Wang et al. (Wang et al., 2020) have confirmed the effectiveness of applying Discrete Fourier Transform (DFT). DFT can convert molecular fingerprints into frequency-domain values, reflecting the specific characteristics of drug molecules. DFT can freely choose frequency domain or time domain according to the needs of practical applications, however, it cannot obtain information in both cases simultaneously, and we cannot know the time when a signal occurs (in our study, it means sequence position information). To solve the local non-stationary components contained in the FP2, DWT was chosen to extract drug features. Daubechies family is the wavelet basis function in DWT, which can support discrete transformation and have good orthogonality and symmetry compared to other wavelet bases. In this paper, the specified wavelet basis function is used to decompose the fingerprint vector, and the approximation coefficients are used as the wavelet coefficients of the fingerprint vector.

After the transformation of DWT with the Daubechies family, 128 approximation coefficients can be obtained to form a vector:

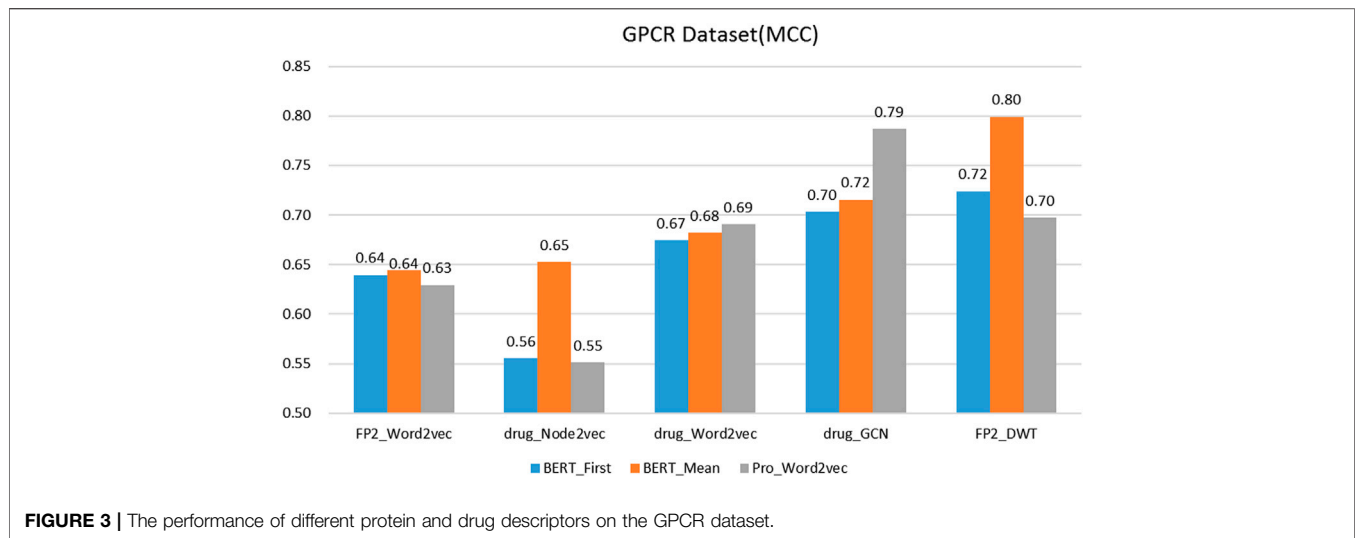
$$S_A = [a_1 a_2 \dots a_{128}] \quad (8)$$

To better characterize the drug, S_A was subjected to a standard conversion as described by the following equation:

$$d_i = \frac{a_i}{\sum_{j=1}^{128} a_j} \quad (9)$$

$$D_{DWT} = [d_1 d_2 \dots d_{128}] \quad (10)$$

And D_{DWT} a 128-dimensional vector is obtained to be used as representation of drug. Finally, through the above several steps, a drug-protein pair can be represented with an 1152-D vector given by:



$$\Phi = [\Phi_1 \quad \Phi_2 \quad \cdots \quad \Phi_i \quad \cdots \quad \Phi_{1152}] \quad (11)$$

2.3 CNN Block

The CNN block includes a convolution layer, a rectified linear unit activation (ReLU), and a max-pooling layer. Instead of using multi-channels, we applied one channel only (Peng et al., 2018). In the convolution layer, apply a convolution kernel with a window size of $h \times k$ to extract the DTIs features, then use the rectified linear unit activation function and performed max-pooling to get the most useful interaction feature from the feature matrix subsequently. Through this block, an output of input x is formulated as:

$$v = \max_{\text{pool}}(f(w \cdot x + b)) \quad (12)$$

where $w \in R^{h \times k}$, which is applied to a window of $h = 18, k = 64$ to produce a new feature; $b \in R$ is a bias term and f is a non-linear function.

2.4 BRL Block

The BRL is built as a special block in the neural network, where data is normalized and then mapped into a specific vector space. This block consists of three layers: a batch-norm layer (BN), a leaky rectified linear activation layer (Leaky ReLU), and a linear layer (Pedregosa et al., 2011).

The input data x is first Batch-normalized, which serves to increase the learning rates further, remove the dropout layer, and apply other modifications afforded by the batch normalization (Ioffe and Szegedy, 2015); then input to the Leaky ReLU activation layer, and finally linearly mapped. BRL block can mathematically be represented as:

$$\begin{aligned} X &= \text{Linear}(\text{LeakyReLU}(\text{BN}(x))) \\ &= W \times (\text{LeakyReLU}(\text{BN}(x))) + B \end{aligned} \quad (13)$$

where x is the input data, the BN transform is applied independently to each dimension of x , W is the weight of the

linear layer, and B is the bias of the linear layer. The first BRL block and CNN block are used for capturing both global and local information to represent the drug-protein pair; the second BRL block is used for predicting DTIs.

The BRL block was implemented with PyTorch (version 1.6.0), and a fully connected layer was used for the linear mapping. The parameters of the first BRL block were set as: the number of input neurons and the batch normalized dimensions dimension were both 1,152, and the number of output neurons was set to 128. The parameters of the second BRL block were set as 192 (128-D from the first BRL block and 64-D from the CNN block), and two respectively. A softmax layer is applied after the second BRL block, which is used to generate the prediction score. Other hyperparameters used default values in Pytorch. The source code for the related methods is available on a GitHub repository at: <https://github.com/Jane4747/DTI-BERT>.

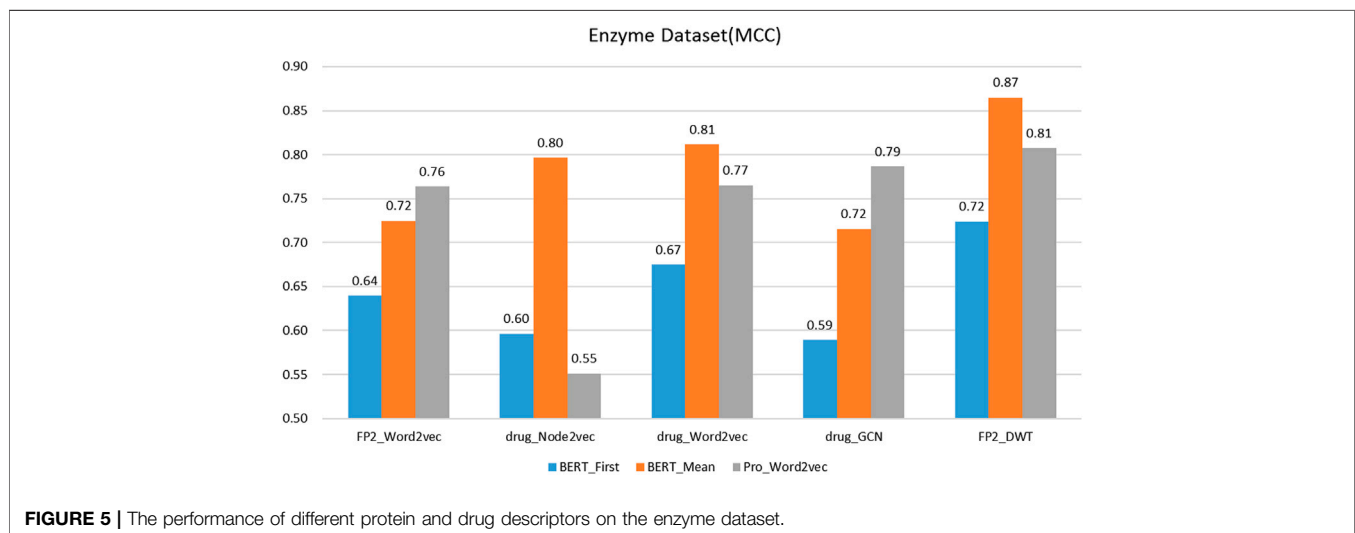
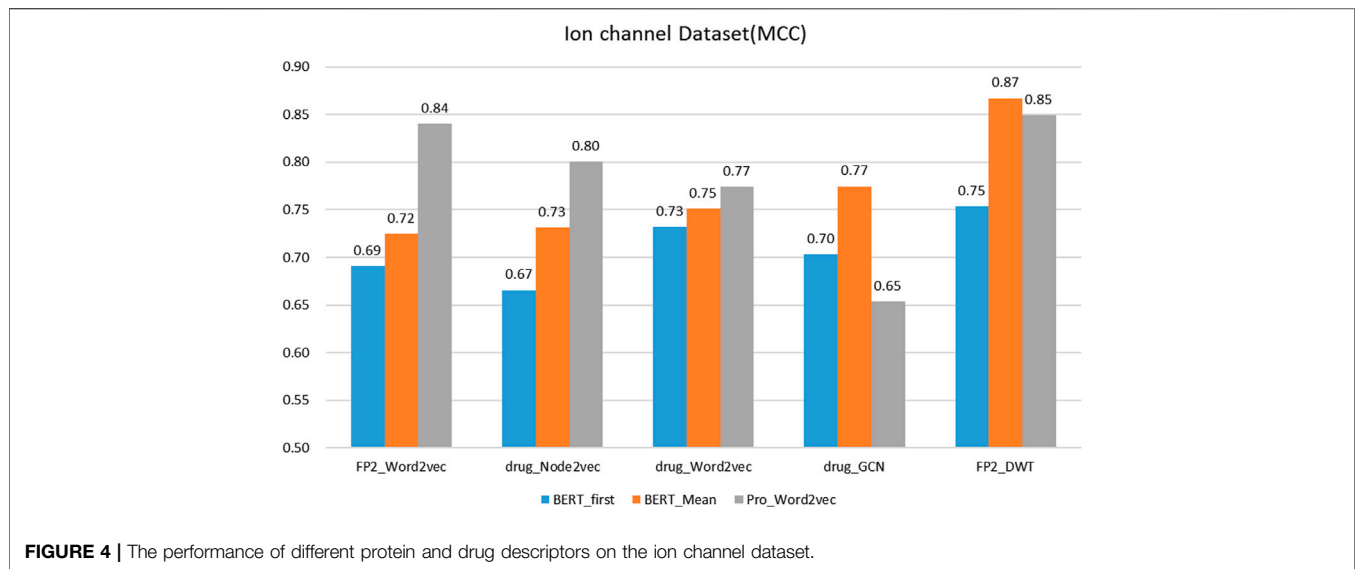
2.5 Optimization Module

In this frame, given two vectors v_1 and v_2 , input them into the same network in turn, the network will map the inputs to the new vector space where the similarity between two inputs can be evaluated by the distance measure function. Here, Euclidean distance was served as the distance measure, denoted as $D(v_1, v_2)$:

$$D(v_1, v_2) = \|v_1 - v_2\|_2 \quad (14)$$

To make the samples of the same interaction label closer, and the distance between different labels as far as possible, the contrastive loss was applied as the loss function of the CNN network:

$$\begin{aligned} \mathcal{L}_1(v_1, v_2, Y) &= \frac{1}{2}(1 - Y)D(v_1, v_2)^2 \\ &\quad + \frac{1}{2}Y\{\max(0, m - D(v_1, v_2))\}^2 \end{aligned} \quad (15)$$



where $Y = 0$ if sequences v_1 and v_2 , have the same label and $Y = 1$ if they are different, $m > 0$ is a margin. In other words, the margin defines a radius, and dissimilar pairs contribute to the loss function only if their distance is within the radius.

In this study, the second BRL block was used to convert the representation vector v to binary category outputs, the backpropagation algorithm was used to update network parameters, and the cross-entropy loss function was selected as the loss function of the second BRL block:

$$\mathcal{L}_2(v, Y) = -Y \log(D(v)) - (1 - Y) \log(1 - D(v)) \quad (16)$$

Therefore, the loss function of the DTI-BERT model is:

$$\mathcal{L}(v_1, v_2, Y, Y_1, Y_2) = \mathcal{L}_1(v_1, v_2, Y) + \mathcal{L}_2(v_1, Y_1) + \mathcal{L}_2(v_2, Y_2) \quad (17)$$

where Y_1 and Y_2 are the labels of v_1 and v_2 .

We implemented our model using Python three and Pytorch (version 1.6.0). Optimizer, training epochs and batch size are set with “Adam”, 70 and 64, respectively. In our work, the optimizing function, “Adam”, use its default parameters value. All codes and trained models can be found via <https://github.com/Jane4747/DTI-BERT>.

3 RESULTS AND DISCUSSION

3.1 Performance Metrics

The determination of a pair belongs to an interactive drug-target pair or non-interactive drug-target pair, is in the case of single-label classification. The metrics such as accuracy (ACC), sensitivity (Sn), Specificity (Sp), strength (str, the average of Sn and Sp) and Matthew’s correlation coefficient (MCC) are frequently used. The specific formulas are as follows:

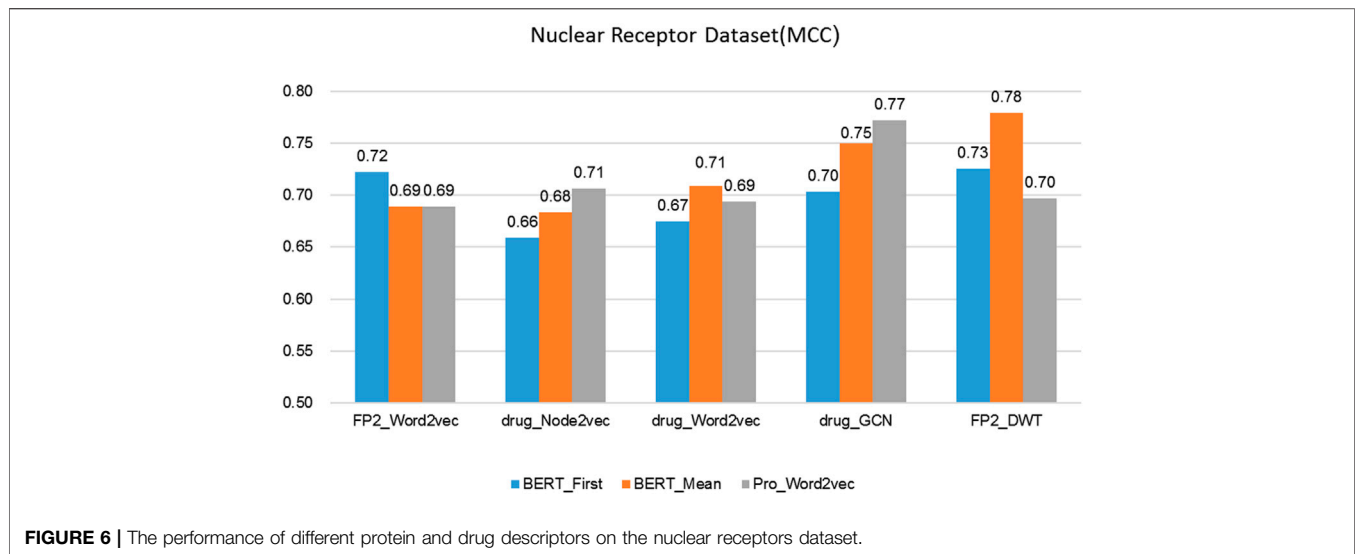


TABLE 1 | Results of comparison with several traditional machine learning methods on four datasets.

Dataset	Method	Sn(%)	Sp(%)	ACC(%)	Str (%)	MCC
GPCR	MLP	86.8	75.3	82.8	81.5	0.61
GPCR	LightGBM	87.5	80.5	86.3	84.0	0.67
GPCR	BRL + CNN	89.3	91.0	90.1	90.2	0.80
Ion channel	MLP	93.3	83.1	89.6	88.2	0.77
Ion channel	LightGBM	92.7	89.3	91.7	91.0	0.81
Ion channel	BRL + CNN	95.9	91.4	94.7	93.7	0.87
Enzyme	MLP	92.2	86.0	90.1	89.1	0.79
Enzyme	LightGBM	92.8	90.5	92.4	91.7	0.83
Enzyme	BRL + CNN	95.9	92.0	94.9	94.0	0.88
NR	MLP	84.2	76.9	79.9	80.6	0.60
NR	LightGBM	84.4	83.1	82.7	83.8	0.65
NR	BRL + CNN	92.5	85.2	89.0	88.9	0.78

The best results for each metric are in bold.

TABLE 2 | Performance comparison on four datasets inaccuracy rate.

Method	GPCRs	Ion-Channels	Enzymes	NR
He et al. (2010)	78.5	80.8	85.5	88.4
DrugRPE Zhang et al. (2017)	85.2	89.0	90.0	91.1
Hu et al. (2019)	88.4	91.9	94.3	85.7
Our method	90.1	94.7	94.9	89.0

The best results for each metric are in bold.

TABLE 3 | Performance comparison on GPCR dataset over leave-one-out cross-validation.

Method	Sn(%)	Sp(%)	ACC(%)	Str (%)	MCC
IGPCR-Drug Xiao et al. (2013)	78.3	91.4	86.9	84.9	0.71
OET-KNN Hu et al. (2016)	77.8	88.7	85.0	83.3	0.67
QuickRBF Hu et al. (2016)	74.8	92.4	86.4	83.6	0.69
SVM Hu et al. (2016)	74.2	92.7	86.4	83.6	0.69
RF Hu et al. (2016)	76.5	92.9	87.3	84.7	0.71
RF + PP Hu et al. (2016)	79.7	92.8	88.3	86.3	0.73
DWKNN(Ensemble) Wang et al. (2020)	81.1	87.1	85.1	84.1	0.67
BOW-GBDT Qiu et al. (2021)	79.8	93.1	88.5	86.3	0.74
Our method	92.2	92.0	91.9	90.1	0.84

The best results for each metric are in bold.

$$\begin{cases}
 Acc = \frac{TP + TN}{TP + TN + FP + FN} \\
 Sn = \frac{TP}{TP + FN} \\
 Sp = \frac{TP}{TP + FP} \\
 Str = \frac{Sp + Sn}{2} \\
 MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN) + (TN + FP)(TN + FN)}}
 \end{cases}
 \quad (18)$$

where TP represents the true positive, FN the false negative, TN the true negative, FP the false positive.

3.2 Comparison of Several Classic Protein and Drug Feature Extraction Methods

On the protein representation task, auto-encoder models (word2vec and BERT) with different model parameters scales were tested. For the drug representation task, a variety of algorithms in various fields, including natural language

TABLE 4 | Performance comparison on Check390.

Method	Sn(%)	Sp(%)	ACC(%)	Str (%)	MCC
IGPCR-Drug Xiao et al. (2013)	80.8	66.9	71.6	73.9	0.45
OET-KNN Hu et al. (2016)	67.7	84.2	78.7	76.9	0.52
QuickRBF Hu et al. (2016)	76.2	77.7	77.2	77.6	0.52
SVM Hu et al. (2016)	76.2	78.9	78.0	77.6	0.53
RF Hu et al. (2016)	78.5	78.1	78.2	78.3	0.54
RF + PPP Hu et al. (2016)	83.1	79.6	80.8	81.3	0.60
DWKNN Wang et al. (2020)	83.9	80.0	81.3	81.9	0.61
DWKNN(Ensemble) Wang et al. (2020)	83.1	82.7	82.8	82.9	0.63
BOW-GBDT Qiu et al. (2021)	80.0	90.0	86.7	85.0	0.70
Our method	87.1	89.4	88.4	88.3	0.76

The best results for each metric are in bold.

processing (word2vec), graph (node2vec and GCN), and signal processing (DWT) were tested.

We evaluated the BERT_Mean + DWT feature extraction method and compared it with several other classic protein and drug feature extraction methods, such as Pr ord2vec (a 64-D vector is obtained to represent the protein, it was extracted by an un-supervised word2vec model and implicated important biophysical and biochemical information (Yang et al., 2018; Zhang et al., 2020), BERT_First (the first row of P_{BERT} is obtained to represented protein, it is a 1024-D vector) (Nambiar et al., 2020), FP2_Word2vec (Jaeger et al., 2018), drug_Node2vec (Grover and Leskovec, 2016; Tetko et al., 2020), drug_Word2vec (Zhang et al., 2020; Zheng et al., 2021), drug_GCN (Chen et al., 2020). **Figures 3–6** show the Matthews correlation coefficient (MCC) for the datasets $S_{GPCR-Drug}$, $S_{ChI-Drug}$, $S_{Ezy-Drug}$, and $S_{NR-Drug}$ obtained for each approach in CNN + BRL classifier via 10-fold cross validation.

It was found that BERT_Mean for the proteins and DWT for drugs can improve the performance of the classifier greatly in four datasets. The BERT_Mean + DWT increased capacity for identifying DTIs compared to the using BERT_First, PRO_Word2vec, drug_Node2vec, drug_Word2vec, and drug_GCN, and BERT_Mean can find the most compact and informative features subsets which are deeply hidden in protein sequences. It is showed that word2vec for protein sequences and GCN for drugs in DTIs tasks, could also obtain good prediction results on three datasets ($S_{GPCR-Drug}$, $S_{Ezy-Drug}$, and $S_{NR-Drug}$), which inspires us that different protein representation methods need to consider different drug molecule representation methods, which need to be determined experimentally.

3.3 Comparison With Some Machine Learning Methods

In order to test the performance of the BRL + CNN and compare it with the existing machine learning methods, we use the same benchmark dataset (listed in **Eq. 1**) and the same BERT_Mean + DWT feature as the input of the prediction model. The proposed BRL + CNN predictor and other commonly used classifiers provided by the Scikit-learn library, like Multi-Layer Perceptron (MLP) with two hidden layers (Pedregosa et al., 2011) and gradient boosting tree-based ensemble method called LightGBM (LGB) (Ke et al., 2017), were tested via 10-

fold cross-validation, the results are listed in **Table 1**. It was found that the proposed BRL + CNN predictor in this article has better performance than other classifiers in all metrics.

3.4 Comparison With Existing Predictor

To further demonstrate the power of the DTI-BERT predictor, we compared it with some existing methods. There are some new models for identifying DTIs trained with the datasets established by He et al. (He et al., 2010). For example, Hu et al. proposed a deep learning-based method to predict DTIs by using the information of drug structures and proteins sequences (Hu et al., 2019), this CnnDIT predictor has better prediction performance in predicting DTIs, and it has its own web server. Zhang et al. proposed a random projection ensemble approach DrugRPE to predict DTIs (Zhang et al., 2017), and several random projections build an ensemble REPTress system. In general, the method of fusing multiple predictors outperforms a single predictor. To facilitate comparison, the scores of accuracies (defined in **Eq. (18)**) obtained by these three predictors (He et al., 2010; Hu et al., 2016; Zhang et al., 2017) based on the benchmark datasets used in He et al. (He et al., 2010) via the 10-fold cross-validation test were listed in **Table 2**. Comprehensively, the comparative results showed that our model is more accurate than other existing methods.

GPCRs have proved to be one of the most important target families of modern drugs. Identifying the GPR-drug interaction is an important issue in bioinformatics, and a number of researchers have proposed effective predicted methods to identify GPCR-drug interactions. Our method was also compared with the performance of different methods which predicting GPCR-drug interaction on the training dataset $S_{GPCR-Drug}$ over leave-one-out cross-validation, and validated in independent test dataset check390 (Xiao et al., 2013; Hu et al., 2016; Wang et al., 2020; Qiu et al., 2021). The results of the different methods tested on $S_{GPCR-Drug}$ over leave-one-out cross-validation were shown in **Table 3**. The results of the other eight methods were reported in (Qiu et al., 2021). From **Table 3**, we can find that the MCC values of our method were 10% higher than others.

The generalization ability of machine learning models is usually evaluated through an independent test. The D92M is the GPCR-drug interaction dataset in (Wang et al., 2020), which is applied as a training dataset, and check390 as a validation

dataset. The results of the validation test on check390 were listed in **Table 4**, which demonstrated that our method almost outperform the others across the five metrics, except for BOW-GBDT achieves the highest value of Sp (93.1%). Compared with other state-of-the-art methods, the ACC value of our method is 3.4% higher, the MCC value is 6% higher than the second one. All these results demonstrate the effectiveness of the proposed methods.

4 CONCLUSION

In this work, we developed a powerful predictor based on the sequences of proteins and FP2 of drugs. We attempted to use pre-trained BERT to present proteins in DTIs and choose a useful representation for drugs via extensive experiments, including several state-of-art drug descriptions like drug_Word2vec, drug_Node2vec, drug_GCN, FP2_Word2vec, FP2_DWT. The presenting results showed that FP2_DWT is more efficient to present drug molecules than other descriptions. Furthermore, we used the deep learning method to generate interaction information and optimized the predicting network based on contrastive loss and cross-entropy loss, which performed much better than other common machine learning models. Moreover, compared with other existing predictors, DTI-BERT has better prediction performance in different target families of GPCRs, ion channels, enzymes and nuclear receptors, without any help of prior knowledge and handcrafted feature engineering. Overall, DTI-BERT can predict drug-target interactions that achieved high accuracy and we established a prediction web-server for the convenience of the most experienced scientists.

REFERENCES

- Alaimo, S., Pulvirenti, A., Giugno, R., and Ferro, A. (2013). Drug-target Interaction Prediction through Domain-Tuned Network-Based Inference. *Bioinformatics* 29 (16), 2004–2008. doi:10.1093/bioinformatics/btt307
- Bianchi, F., Terragni, S., Hovy, D., and Assoc Computat, L. (2021). Pre-training is a Hot Topic: Contextualized Document Embeddings Improve Topic Coherence. in Joint Conference of 59th Annual Meeting of the Association-for-Computational-Linguistics (ACL)/11th International Joint Conference on Natural Language Processing (IJCNLP)/6th Workshop on Representation Learning for NLP (ReL4NLP), Aug 01–06 2021. (Electr Network), 759–766.
- Brandes, N., Ofer, D., Peleg, Y., Rappoport, N., and Linial, M. (2022). ProteinBERT: A Universal Deep-Learning Model of Protein Sequence and Function. *Bioinformatic*. 38 (8), 2102–2110. doi:10.1101/2021.05.24.445464
- Butina, D., Segall, M. D., and Frankcombe, K. (2002). Predicting ADME Properties In Silico: Methods and Models. *Drug Discov. today* 7 (11), S83–S88. doi:10.1016/s1359-6446(02)02288-2
- Chen, L., Tan, X., Wang, D., Zhong, F., Liu, X., Yang, T., et al. (2020). TransformerCPI: Improving Compound-Protein Interaction Prediction by Sequence-Based Deep Learning with Self-Attention Mechanism and Label Reversal Experiments. *Bioinformatics* 36 (16), 4406–4414. doi:10.1093/bioinformatics/btaa524
- Chen, R., Liu, X., Jin, S., Lin, J., and Liu, J. (2018). Machine Learning for Drug-Target Interaction Prediction. *Molecules* 23 (9), 2208. doi:10.3390/molecules23092208
- Chen, X., Sun, Y. Z., Zhang, D. H., Li, J. Q., Yan, G. Y., An, J. Y., et al. (2017). NRDTD: a Database for Clinically or Experimentally Supported Non-coding RNAs and Drug Targets Associations. *Database (Oxford)* 2017, bax057. doi:10.1093/database/bax057

The BERT model has very excellent general capabilities and has very outstanding feature extraction capabilities for DNA sequences (Le et al., 2021) and RNA sequences (Zhang et al., 2021). The DTIs prediction framework proposed in this paper has very good potential for predicting other drug targets as well.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <http://121.36.221.79/dtibert/download>.

AUTHOR CONTRIBUTIONS

XX conceived and designed the experiments, JZ performed the extraction of features, model construction, model training, and evaluation. JZ drafted the manuscript, XX and W-RQ supervised this project and revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the grants from the National Natural Science Foundation of China (Nos 31860312, 62162032, and 62062043), Natural Science Foundation of Jiangxi Province, China (NO. 20202BAB202007), the International Cooperation Project of the Ministry of Science and Technology, China (NO. 2018-3-3).

- Chen, X., Liu, M.-X., and Yan, G.-Y. (2012). Drug-target Interaction Prediction by Random Walk on the Heterogeneous Network. *Mol. Biosyst.* 8 (7), 1970–1978. doi:10.1039/c2mb00002d
- Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q. V., and Salakhutdinov, R. (2019). Acl: Transformer-xl: Attentive Language Models beyond a Fixed-Length Context. in 57th Annual Meeting of the Association-for-Computational-Linguistics (ACL): 2019, Florence, ITALY, Jul 28–Aug 02 2019, 2978–2988.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). *Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv preprint. *arXiv:1810.04805* 2018.
- Dobson, C. M. (2004). Chemical Space and Biology. *Nature* 432 (7019), 824–828. doi:10.1038/nature03192
- Dubach, J. M., Vinegoni, C., Mazitschek, R., Fumene Feruglio, P., Cameron, L. A., and Weissleder, R. (2014). In Vivo imaging of Specific Drug-Target Binding at Subcellular Resolution. *Nat. Commun.* 5 (1), 3946–3949. doi:10.1038/ncomms4946
- Elnaggar, A., Heinzinger, M., Dallago, C., Rihawi, G., Wang, Y., Jones, L., et al. (2021). ProtTrans: Towards Cracking the Language of Life's Code through Self-Supervised Deep Learning and High Performance Computing. in IEEE Transactions on Pattern Analysis and Machine Intelligence 2021.
- Ezzat, A., Zhao, P., Wu, M., Li, X. L., and Kwok, C. K. (2016). Drug-target Interaction Prediction with Graph Regularized Matrix Factorization. *IEEE/ACM Trans. Comput. Biol. Bioinform* 14 (3), 646–656. doi:10.1109/TCBB.2016.2530062
- Grover, A., and Leskovec, J. (2016). node2vec: Scalable Feature Learning for Networks. in 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD): 2016, San Francisco, CA, Aug 13–17 2016, 855–864. doi:10.1145/2939672.2939754
- Haggarty, S. J., Koeller, K. M., Wong, J. C., Butcher, R. A., and Schreiber, S. L. (2003). Multidimensional Chemical Genetic Analysis of Diversity-Oriented

- Synthesis-Derived Deacetylase Inhibitors Using Cell-Based Assays. *Chem. Biol.* 10 (5), 383–396. doi:10.1016/s1074-5521(03)00095-4
- He, Z., Zhang, J., Shi, X.-H., Hu, L.-L., Kong, X., Cai, Y.-D., et al. (2010). Predicting Drug-Target Interaction Networks Based on Functional Groups and Biological Features. *PLoS one* 5 (3), e9603. doi:10.1371/journal.pone.0009603
- Hu, J., Li, Y., Yang, J.-Y., Shen, H.-B., and Yu, D.-J. (2016). GPCR-drug Interactions Prediction Using Random Forest with Drug-Association-Matrix-Based Post-processing Procedure. *Comput. Biol. Chem.* 60, 59–71. doi:10.1016/j.compbiolchem.2015.11.007
- Hu, S., Zhang, C., Chen, P., Gu, P., Zhang, J., and Wang, B. (2019). Predicting Drug-Target Interactions from Drug Structure and Protein Sequence Using Novel Convolutional Neural Networks. *BMC Bioinforma.* 20 (25), 689. doi:10.1186/s12859-019-3263-x
- Ilić, S., Marrese-Taylor, E., Balazs, J. A., and Matsuo, Y. (2018). Deep Contextualized Word Representations for Detecting Sarcasm and Irony. in Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2–7.
- Ioffe, S., and Szegedy, C. (2015). “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.” in 32nd International Conference on Machine Learning: 2015, Lille, France, Jul 07–09 2015, 448–456.
- Jaeger, S., Fulle, S., and Turk, S. (2018). Mol2vec: Unsupervised Machine Learning Approach with Chemical Intuition. *J. Chem. Inf. Model.* 58 (1), 27–35. doi:10.1021/acs.jcim.7b00616
- Karlov, D. S., Sosnin, S., Fedorov, M. V., and Popov, P. (2020). graphDelta: MPNN Scoring Function for the Affinity Prediction of Protein-Ligand Complexes. *ACS omega* 5 (10), 5150–5159. doi:10.1021/acsomega.9b04162
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al. (2017). Lightgbm: A Highly Efficient Gradient Boosting Decision Tree. Proceedings of the 31st International Conference on Neural Information Processing Systems, Red Hook, NY, USA. (Long Beach, CA, USA: Curran Associates Inc.), 3149–3157.
- Keiser, M. J., Roth, B. L., Armbruster, B. N., Ernsberger, P., Irwin, J. J., and Shoichet, B. K. (2007). Relating Protein Pharmacology by Ligand Chemistry. *Nat. Biotechnol.* 25 (2), 197–206. doi:10.1038/nbt1284
- Kim, S., Jin, D., and Lee, H. (2013). Predicting Drug-Target Interactions Using Drug-Drug Interactions. *PLoS one* 8 (11), e80129. doi:10.1371/journal.pone.0080129
- Kombo, D. C., Tallapragada, K., Jain, R., Chewing, J., Mazurov, A. A., Speake, J. D., et al. (2013). 3D Molecular Descriptors Important for Clinical Success. *J. Chem. Inf. Model.* 53 (2), 327–342. doi:10.1021/ci300445e
- Kotlyar, M., Fortney, K., and Jurisica, I. (2012). Network-based Characterization of Drug-Regulated Genes, Drug Targets, and Toxicity. *Methods* 57 (4), 499–507. doi:10.1016/j.ymeth.2012.06.003
- Le, N. Q. K., Ho, Q. T., Nguyen, T. T., and Ou, Y. Y. (2021). A Transformer Architecture Based on BERT and 2D Convolutional Neural Network to Identify DNA Enhancers from Sequence Information. *Brief. Bioinform* 22 (5), bbab005. doi:10.1093/bib/bbab005
- Lee, C., Cho, K., and Kang, W. (2019). Mixout: Effective Regularization to Finetune Large-Scale Pretrained Language Models. in International Conference on Learning Representations (ICLR): 2020. (International Conference on Learning Representations).
- Li, S., Wan, F., Shu, H., Jiang, T., Zhao, D., and Zeng, J. (2020). MONN: a Multi-Objective Neural Network for Predicting Compound-Protein Interactions and Affinities. *Cell Syst.* 10 (4), 308–322. doi:10.1016/j.cels.2020.03.002
- Lim, S., Lu, Y., Cho, C. Y., Sung, I., Kim, J., Kim, Y., et al. (2021). A Review on Compound-Protein Interaction Prediction Methods: Data, Format, Representation and Model. *Comput. Struct. Biotechnol. J.* 19, 1541–1556. doi:10.1016/j.csbj.2021.03.004
- Liu, T., Lin, Y., Wen, X., Jorissen, R. N., and Gilson, M. K. (2007). BindingDB: a Web-Accessible Database of Experimentally Determined Protein-Ligand Binding Affinities. *Nucleic Acids Res.* 35 (Suppl. 1), D198–D201. doi:10.1093/nar/gkl999
- Luo, Y., Zhao, X., Zhou, J., Yang, J., Zhang, Y., Kuang, W., et al. (2017). A Network Integration Approach for Drug-Target Interaction Prediction and Computational Drug Repositioning from Heterogeneous Information. *Nat. Commun.* 8 (1), 573. doi:10.1038/s41467-017-00680-8
- Mei, J.-P., Kwok, C.-K., Yang, P., Li, X.-L., and Zheng, J. (2013). Drug-target Interaction Prediction by Learning from Local Information and Neighbors. *Bioinformatics* 29 (2), 238–245. doi:10.1093/bioinformatics/bts670
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* 30 (16), 2785–2791. doi:10.1002/jcc.21256
- Mousavian, Z., and Masoudi-Nejad, A. (2014). Drug-target Interaction Prediction via Chemogenomic Space: Learning-Based Methods. *Expert Opin. drug metabolism Toxicol.* 10 (9), 1273–1287. doi:10.1517/17425255.2014.950222
- Nambiar, A., Heflin, M., Liu, S., Maslov, S., Hopkins, M., and Ritz, A. (2020). “Transforming the Language of Life: Transformer Neural Networks for Protein Prediction Tasks,” in Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, 1–8.
- Noorbakhsh, J., Farahmand, S., Foroughi Pour, A., Namburi, S., Caruana, D., Rimm, D., et al. (2020). Deep Learning-Based Cross-Classifications Reveal Conserved Spatial Behaviors within Tumor Histological Images. *Nat. Commun.* 11 (1), 6367. doi:10.1038/s41467-020-20030-5
- Nozza, D., Bianchi, F., and Hovy, D. (2020). *What the [mask]? Making Sense of Language-specific BERT Models.* arXiv preprint. arXiv:200302912 2020.
- Öztürk, H., Ozkirimli, E., and Özgür, A. (2016). A Comparative Study of SMILES-based Compound Similarity Functions for Drug-Target Interaction Prediction. *BMC Bioinforma.* 17 (1), 1–11. doi:10.1186/s12859-016-0977-x
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Peng, Y., Rios, A., Kavuluru, R., and Lu, Z. (2018). Extracting Chemical-Protein Relations with Ensembles of SVM and Deep Learning Models. *Database: J. Biol. Databases curation* 2018, bay073. doi:10.1093/database/bay073
- Peng, Y., Yan, S., and Lu, Z. (2019). *Transfer Learning in Biomedical Natural Language Processing: An Evaluation of BERT and ELMo on Ten Benchmarking Datasets.* arXiv preprint. arXiv:190605474 2019.
- Pujadas, G., Vaque, M., Ardevol, A., Blade, C., Salvado, M., Blay, M., et al. (2008). Protein-ligand Docking: A Review of Recent Advances and Future Perspectives. *Cpa* 4 (1), 1–19. doi:10.2174/157341208783497597
- Qin, C., Zhang, C., Zhu, F., Xu, F., Chen, S. Y., Zhang, P., et al. (2014). Therapeutic Target Database Update 2014: a Resource for Targeted Therapeutics. *Nucl. Acids Res.* 42 (D1), D1118–D1123. doi:10.1093/nar/gkt1129
- Qiu, W., Lv, Z., Hong, Y., Jia, J., and Xiao, X. (2021). A GBDT Classifier Combining with Artificial Neural Network for Identifying GPCR-Drug Interaction Based on Wordbook Learning from Sequences. *Front. Cell Dev. Biol.* 8, 1789. doi:10.3389/fcell.2020.623858
- Qu, J., Chen, X., Sun, Y.-Z., Zhao, Y., Cai, S.-B., Ming, Z., et al. (2019). In Silico Prediction of Small Molecule-miRNA Associations Based on the HeteSim Algorithm. *Mol. Ther. - Nucleic Acids* 14, 274–286. doi:10.1016/j.omtn.2018.12.002
- Rognan, D. (2007). Chemogenomic Approaches to Rational Drug Design. *Br. J. Pharmacol.* 152 (1), 38–52. doi:10.1038/sj.bjp.0707307
- Sawada, R., Kotera, M., and Yamanishi, Y. (2014). Benchmarking a Wide Range of Chemical Descriptors for Drug-Target Interaction Prediction Using a Chemogenomic Approach. *Mol. Inf.* 33 (11-12), 719–731. doi:10.1002/minf.201400066
- Tetko, I. V., Karpov, P., Van Deursen, R., and Godin, G. (2020). State-of-the-art Augmented NLP Transformer Models for Direct and Single-step Retrosynthesis. *Nat. Commun.* 11 (1), 5575. doi:10.1038/s41467-020-19266-y
- Tornø, W., and Altman, R. B. (2019). Graph Convolutional Neural Networks for Predicting Drug-Target Interactions. *J. Chem. Inf. Model.* 59 (10), 4131–4149. doi:10.1021/acs.jcim.9b00628
- Wan, F., Hong, L., Xiao, A., Jiang, T., and Zeng, J. (2019). NeoDTI: Neural Integration of Neighbor Information from a Heterogeneous Network for Discovering New Drug-Target Interactions. *Bioinformatics* 35 (1), 104–111. doi:10.1093/bioinformatics/bty543
- Wang, C.-C., and Chen, X. (2019). A Unified Framework for the Prediction of Small Molecule-MicroRNA Association Based on Cross-Layer Dependency Inference on Multilayered Networks. *J. Chem. Inf. Model.* 59 (12), 5281–5293. doi:10.1021/acs.jcim.9b00667

- Wang, P., Huang, X., Qiu, W., and Xiao, X. (2020). Identifying GPCR-Drug Interaction Based on Wordbook Learning from Sequences. *BMC Bioinforma.* 21 (1), 150. doi:10.1186/s12859-020-3488-8
- Wang, Y.-C., Yang, Z.-X., Wang, Y., and Deng, N.-Y. (2010). Computationally Probing Drug-Protein Interactions via Support Vector Machine. *Lddd* 7 (5), 370–378. doi:10.2174/157018010791163433
- Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: a Major Update to the DrugBank Database for 2018. *Nucleic acids Res.* 46 (D1), D1074–D1082. doi:10.1093/nar/gkx1037
- Xiao, X., Min, J.-L., Wang, P., and Chou, K.-C. (2013). iGPCR-Drug: A Web Server for Predicting Interaction between GPCRs and Drugs in Cellular Networking. *PLoS one* 8 (8), e72234. doi:10.1371/journal.pone.0072234
- Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., and Kanehisa, M. (2008). Prediction of Drug-Target Interaction Networks from the Integration of Chemical and Genomic Spaces. *Bioinformatics* 24 (13), i232–i240. doi:10.1093/bioinformatics/btn162
- Yamanishi, Y., Kotera, M., Moriya, Y., Sawada, R., Kanehisa, M., and Goto, S. (2014). DINIES: Drug-Target Interaction Network Inference Engine Based on Supervised Analysis. *Nucleic acids Res.* 42 (W1), W39–W45. doi:10.1093/nar/gku337
- Yang, K. K., Wu, Z., Bedbrook, C. N., and Arnold, F. H. (2018). Learned Protein Embeddings for Machine Learning. *Bioinformatics* 34 (15), 2642–2648. doi:10.1093/bioinformatics/bty178
- Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., and Le, Q. V. (2019). “Xlnet: Generalized Autoregressive Pretraining for Language Understanding” in *Advances in Neural Information Processing Systems*. Editor H. Wallach, H. Larochelle, A. Beygelzimer, F. d. {e}-Buc, E. Fox, and R. Garnett (Curran Associates, Inc.) 32.
- Yin, J., Chen, X., Wang, C.-C., Zhao, Y., and Sun, Y.-Z. (2019). Prediction of Small Molecule-MicroRNA Associations by Sparse Learning and Heterogeneous Graph Inference. *Mol. Pharm.* 16 (7), 3157–3166. doi:10.1021/acs.molpharmaceut.9b00384
- Zhang, J., Zhu, M., Chen, P., and Wang, B. (2017). DrugRPE: Random Projection Ensemble Approach to Drug-Target Interaction Prediction. *Neurocomputing* 228, 256–262. doi:10.1016/j.neucom.2016.10.039
- Zhang, L., Qin, X., Liu, M., Liu, G., and Ren, Y. (2021). BERT-m7G: A Transformer Architecture Based on BERT and Stacking Ensemble to Identify RNA N7-Methylguanosine Sites from Sequence Information. *Comput. Math. Methods Med.* 2021, 7764764. doi:10.1155/2021/7764764
- Zhang, W., Zou, H., Luo, L., Liu, Q., Wu, W., and Xiao, W. (2016). Predicting Potential Side Effects of Drugs by Recommender Methods and Ensemble Learning. *Neurocomputing* 173, 979–987. doi:10.1016/j.neucom.2015.08.054
- Zhang, Y.-F., Wang, X., Kaushik, A. C., Chu, Y., Shan, X., Zhao, M.-Z., et al. (2020). SPVec: a Word2vec-Inspired Feature Representation Method for Drug-Target Interaction Prediction. *Front. Chem.* 7, 895. doi:10.3389/fchem.2019.00895
- Zheng, J., Xiao, X., and Qiu, W. R. (2021). iCDI-W2vCom: Identifying the Ion Channel-Drug Interaction in Cellular Networking Based on Word2vec and Node2vec. *Front. Genet.* 12, 738274. doi:10.3389/fgene.2021.738274
- Zhou, X., Dai, E., Song, Q., Ma, X., Meng, Q., Jiang, Y., et al. (2020). In Silico drug Repositioning Based on Drug-miRNA Associations. *Briefings Bioinforma.* 21 (2), 498–510. doi:10.1093/bib/bbz012
- Zhu, S., Okuno, Y., Tsujimoto, G., and Mamitsuka, H. (2005). A Probabilistic Model for Mining Implicit ‘chemical Compound-Gene’ Relations from Literature. *Bioinformatics* 21 (Suppl. 1_2), ii245. doi:10.1093/bioinformatics/bti1141

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zheng, Xiao and Qiu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.