



DNN-DTIs: Improved drug-target interactions prediction using XGBoost feature selection and deep neural network



Cheng Chen ^{a,b,c,1}, Han Shi ^{d,1}, Zhiwen Jiang ^{a,b}, Adil Salhi ^e, Ruixin Chen ^{a,b}, Xuefeng Cui ^c, Bin Yu ^{a,b,f,*}

^a College of Mathematics and Physics, Qingdao University of Science and Technology, Qingdao, 266061, China

^b Artificial Intelligence and Biomedical Big Data Research Center, Qingdao University of Science and Technology, Qingdao, 266061, China

^c School of Computer Science and Technology, Shandong University, Qingdao, 266237, China

^d Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai, 200032, China

^e Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology (KAUST), Thuwal, 23955, Saudi Arabia

^f Key Laboratory of Computational Science and Application of Hainan Province, Haikou, 571158, China

ARTICLE INFO

Keywords:

Drug-target interactions
Multi-information fusion
XGBoost
Deep neural network

ABSTRACT

Analysis and prediction of drug-target interactions (DTIs) play an important role in understanding drug mechanisms, as well as drug repositioning and design. Machine learning (ML)-based methods for DTIs prediction can mitigate the shortcomings of time-consuming and labor-intensive experimental approaches, while providing new ideas and insights for drug design. We propose a novel pipeline for predicting drug-target interactions, called DNN-DTIs. First, the target information is characterized by a number of features, namely, pseudo-amino acid composition, pseudo position-specific scoring matrix, conjoint triad composition, transition and distribution, Moreau-Broto autocorrelation, and structural features. The drug compounds are subsequently encoded using substructure fingerprints. Next, eXtreme gradient boosting (XGBoost) is used to determine the subset of non-redundant features of importance. The optimal balanced set of sample vectors is obtained by applying the synthetic minority oversampling technique (SMOTE). Finally, a DTIs predictor, DNN-DTIs, is developed based on a deep neural network (DNN) via a layer-by-layer learning scheme. Experimental results indicate that DNN-DTIs achieves better performance than other state-of-the-art predictors with ACC values of 98.78%, 98.60%, 97.98%, 98.24% and 98.00% on Enzyme, Ion Channels (IC), GPCR, Nuclear Receptors (NR) and Kuang's datasets. Therefore, the accurate prediction performance of DNN-DTIs makes it a favored choice for contributing to the study of DTIs, especially drug repositioning.

1. Introduction

The study of drug-target interactions (DTIs) is very helpful for the exploration of potentially novel therapeutic effects of existing drugs (drug repositioning), the prediction of unexplored adverse reactions to drugs, as well as the general understanding of pharmacology. Recent developments in molecular biology, are allowing researchers to gain a more comprehensive understanding of the biological and chemical processes involved in DTIs, which has been a major driver in the development of a number of public databases [1–3], such as Super Target [4], KEGG [5], DrugBank [6], and TTD [7]. Nevertheless, most traditional methods for DTI identification are cumbersome, expensive,

and time-consuming. Artificial intelligence offers an alternative and more effective solution, which has the potential to provide new ideas for repositioning drugs, recommending candidate drugs, and detecting drug side effects. In recent years, the prediction of DTIs has attracted increasing attention, especially methods based on computational biology [8,9], which provide potential new knowledge on drug candidates and specific targets.

Conventional approaches for predicting DTIs are roughly divided into three categories: ligand-based methods [10], docking methods [11], and chemogenomic methods [12]. However, ligand-based methodology largely depends on prior information of known ligands [13], and docking-based methods require the 3D structure [14]. This is why

* Corresponding author. College of Mathematics and Physics, Qingdao University of Science and Technology, Qingdao, 266061, China.

E-mail address: yubin@qust.edu.cn (B. Yu).

¹ These authors contributed equally to this work.

chemogenomic methods are becoming increasingly important, as they incorporate numerous types of information including sequence-based, structure-based, evolution-based, drug molecular fingerprint-based and biochemistry-based information. In order to facilitate the task of drug developers, Ezzat et al. [12] reviewed the major chemogenomic methods, including neighborhood model, bipartite local models, network diffusion methods, matrix factorization methods, and feature-based classification methods. In 2020, Bagherian et al. [15] also reviewed machine learning (ML)-based DTIs prediction methodology. Compared with Ezzat et al. Bagherian et al. divided chemogenomic methods into a taxonomy of approaches, such as deep learning (DL)-based and hybrid-based methods.

In the field of DTIs prediction, researchers have made a lot of progress and valuable results have been obtained based on traditional ML [16]. Yamanishi et al. [17] transformed the task of DTIs prediction into a supervised bipartite graph learning method by integrating chemical and genomic information. Subsequently, Yamanishi et al. [18] learned pharmacological effects from the similarity of chemical structures, and inferred unknown DTIs from these pharmacological effects and genomic sequence information. Similarity based methods such as this, have been extensively employed to predict DTIs. For example, Shi et al. [19] enhanced the similarity measure and designed WNN-GIP and KBFM2K by dealing with the missing interaction problem. In 2018, A similarity-based DDR method was proposed by Olayan et al. [20]. The nonlinear similarity fusion method was used to extract the similarity matrix information of drugs and targets, and different graph-based feature vectors were extracted from the DTIs heterogeneous graph. Subsequently, random forest (RF) was employed to verify the effectiveness of the DDR method.

Matrix factorization-based methods have also been widely used. Ezzat et al. [21] proposed two matrix factorization methods to predict DTIs between “new drugs” and “new targets” by adding edges through intermediate interaction likelihood scores. As another example, Xia et al. [22] used the matrix decomposition method of low-rank weight estimation to predict DTIs, which they called SPLCMF. They obtained low-dimensional vectors from the drug-target network based on a least-squares policy, and the soft weight method was employed to reflect the hidden information during the training process.

Using effective feature extraction or embedding strategies can achieve reliable DTIs prediction. Zhang et al. [23] used “PaDEL Descriptor” to describe the drug features, which they concatenated with the target feature vectors obtained using physicochemical properties, in order to construct the initial embedding vectors. Subsequently, random projection was employed to generate effective low-dimensional vectors from the original space. Li et al. [24] proposed a multi-view low-rank embedding for DTIs prediction, which characterized drugs and targets by fusing structural and chemical information. Shi et al. [25] proposed a prediction pipeline, LRF-DTIs, using FP2 molecular fingerprint to encode the drug. Optimal vectors were obtained using feature selection and SMOTE was used for sample balancing.

In the past few years, DL has been applied in the field of biomedicine at an unprecedented speed, and many DL-based frameworks have been used to deal with the problem of DTIs prediction [26–31]. A convolutional neural network (CNN)-based method is proposed by Özürk et al. [32], which uses only sequence information and performed DTIs prediction on the Davis and KIBA dataset. Rayhan et al. [33] proposed the FRnet-DTI, a method that used autoencoder and CNN for feature extraction and classification. Lee et al. [34] used a CNN to extract the local distribution patterns of amino acid sequences, and trained an effective DL model using a large-scale dataset of DTIs and non-DTIs. Zeng et al. [35] used cascade deep forest combining with arbitrary-order neighboring algorithms to predict DTIs, thus a network-based computing method called AOPEDF was proposed. The effective feature vectors contain chemical, genomic, phenotypic, and network profiles information. Zhao et al. [36] proposed a new graph convolutional network (GCN)-based method, which extracted feature

information of drug-target pairs through GCN.

Although much research has been done based on ML and DL for drug-target interactions prediction, there is still much room for improvement in the following four areas. First, the number of candidate drug-target pairs that do not interact is much larger than the number of drug-target pairs that do interact, leading to a serious imbalance problem. Second, more and more feature information on drugs and targets is being made available, including biological sequence information, network topology, physicochemical and structural information. Methods which can comprehensively and effectively integrate such heterogeneous feature information are necessary. Third, even though computational methods have achieved good prediction results for DTIs, this does not necessarily mean that such methods are effective at predicting interactions between unknown or independent DTIs. Therefore, it is of great necessity to evaluate the trained models based on a comprehensive DTIs network. Fourth, deep learning approaches, such as DNN and CNN models, have a myriad of hyperparameters that need to be adjusted in order to obtain optimal prediction performance. It is therefore our position that designing an effective deep learning-based pipeline which takes these concerns into account is required for the accurate and effective prediction of DTIs.

Inspired by the above discussion, the DNN-DTIs predictor, based on deep learning and multi-feature fusion, is proposed for DTIs prediction. First, information based on sequence, structure, as well as evolutionary information is derived for the target. Meanwhile, molecular fingerprints are utilized to encode drug compounds. For the prediction of DTIs, we first fuse PseAAC, CT, CTD, NMBroto and secondary structure to extract target information. The tree-based XGBoost is employed to determine the optimal feature subset in terms of gain, and thereafter, the SMOTE algorithm is used to generate artificial samples for addressing the data imbalance problem. It is worth mentioning that we are the first to perform feature selection based on XGBoost in order to predict DTIs. Finally, a DNN-based framework with multiple hidden layers is constructed. After the optimization of different feature extraction algorithms, feature selection methods, and the classifier algorithm, the DNN-DTIs model is constructed using optimal parameters for each. Results on Enzyme, IC, GPCR, NR, and Kuang’s dataset showcase the effectiveness of DNN-DTIs. In particular, our model demonstrates the promising drug design validation by predicting DTIs on both Network1 and Network2 successfully, showing improvements on both performance and topological property.

2. Materials and methods

2.1. Datasets

In this study, seven drug-target datasets are collected, four of which are from the gold standard dataset (GSD), constructed by Yamanishi et al. [17]. The datasets are divided into four categories: Enzyme, IC, GPCR and NR, and are used to perform parameter optimization. The specific distribution of each dataset can be shown in [Supplementary Material Table S1](#). We also used the dataset constructed by Kuang et al. [37] to further evaluate DNN-DTIs. The number of targets and drugs were 809 and 786, respectively, and the number of interacting pairs is 3681. Finally, we utilize two DTI networks called Network1 and Network2 to verify the effectiveness of DNN-DTIs. Network1 was obtained from Shi et al. [25], which was in turn collected from Refs. [5,6], and [38]. Network2 was from Xia et al. [22]. The number of interacting pairs for each network was 33 and 150, respectively.

2.2. Protein feature extraction methods

2.2.1. Composition, transformation and distribution

Composition (C), transition (T) and distribution (D) global protein descriptors were proposed by Dubchak et al. [39] in 1995. CTD describes the positions and both the global and local physicochemical

information of amino acids. C represents the composition-based grouped amino acid information, T represents the dipeptide-based frequency information, and D represents frequency and position information in the first, 25%, 50%, 75%, and the last occurrence of the grouped protein sequence. In this paper, thirteen physicochemical properties are used to group amino acids (Table S2). Taking hydrophobicity as an example, the amino acid characteristic signals are converted into three groups (PO (polar), NE (neutral) and HY (hydrophobic)). The calculation formulas for C, T, and D are shown in equations (1)–(3):

$$C(r) = \frac{N(r)}{N} \quad r = \{PO, NE, HY\} \quad (1)$$

$$T(r, s) = \frac{N(r, s) + N(s, r)}{N - 1} \quad r, s = \{(PO, NE), (NE, HY), (HY, PO)\} \quad (2)$$

$$D(r) = \left(\frac{L(r, 1)}{N}, \frac{L(r, 2)}{N}, \frac{L(r, 3)}{N}, \frac{L(r, 4)}{N}, \frac{L(r, 5)}{N} \right) \quad r = \{PO, NE, HY\} \quad (3)$$

where $N(r)$ is the size of the r -th group in the amino acid. N is the sequence length. $N(r, s)$ is the occurrence frequency of the dipeptide from r -th group to s -th group. $L(r, 1), L(r, 2), \dots, L(r, 5)$ indicate the position information of the r -th group of amino acid at the first, 25%, 50%, 75% and 100%.

2.2.2. Conjoint triad

The conjoint triad (CT) was proposed by Shen et al. [40] in 2007. This method considers not only the composition pattern but also the sequence order knowledge. First, 20 amino acids are grouped into seven categories. Then, the adjacent three residues are regarded as a whole block, and the frequency of conjoint triad can be calculated using equation (4):

$$d_i = \frac{f_i - \min\{f_1, f_2, \dots, f_{343}\}}{\max\{f_1, f_2, \dots, f_{343}\}} \quad (4)$$

2.2.3. Pseudo amino acid composition

Pseudo amino acid composition (PseAAC) [41] is a feature encoding method which combines amino acid composition, physicochemical properties, and statistical information (equation (5) and (6)).

$$P = [p_1, p_2, \dots, p_{20}, p_{20+1}, \dots, p_{20+\lambda}]^T \quad (5)$$

$$p_u = \begin{cases} \frac{f_u}{\sum_{u=1}^{20} f_u + w \sum_{k=1}^{\lambda} \tau_k}, & 1 \leq u \leq 20 \\ \frac{w \tau_{u-20}}{\sum_{u=1}^{20} f_u + w \sum_{k=1}^{\lambda} \tau_k}, & 20 + 1 \leq u \leq 20 + \lambda \end{cases} \quad (6)$$

where p_1, p_2, \dots, p_{20} represent the amino acid composition information, and $p_{20+1}, \dots, p_{20+\lambda}$ represent the sequence order information. $\lambda (\lambda < N)$ reflects the sequence-related factors at different levels. w represents the weight factor.

2.2.4. Pseudo position-specific scoring matrix

Pseudo position-specific scoring matrix (PsePSSM) is obtained through the PSSM matrix [42]. The PSI-BLAST program [43] is used in the Swiss-Prot database for comparison. After normalization, the PSSM is converted into PsePSSM feature vectors with the same length and biological significance through equations (8) and (9):

$$\bar{E}_j = \frac{1}{L} \sum_{i=1}^N E_{i \rightarrow j} \quad (j = 1, 2, \dots, 20; 0 \leq \xi < N) \quad (8)$$

$$G_j^\xi = \frac{1}{L - \xi} \sum_{i=1}^{N-\xi} [E_{i \rightarrow j} - E_{(i+\xi) \rightarrow j}]^2 \quad (9)$$

where \bar{E}_j represents the average score at j -th position. G_j^ξ can illustrate the order evolutional information. ξ represents the interval.

2.2.5. Normalized Moreau-Broto autocorrelation

The Normalized Moreau-Broto (NMBroto) is a type of topology autocorrelation descriptor (equation (10) and (11)) proposed by Horne [44]. It is defined using the distribution of amino acid physicochemical properties.

$$AC(d) = \sum_{i=1}^{N-d} P_i P_{i+d} \quad d = 1, 2, \dots, nlag \quad (10)$$

$$NMBroto(d) = \frac{AC(d)}{N - d} \quad d = 1, 2, \dots, nlag \quad (11)$$

where, d is called the lag of autocorrelation (already known as the interval of amino acids), P_i and P_{i+d} are the amino acid property value at positions i and $i + d$, $nlag (nlag < N)$ is the maximum value of lag. This paper uses 8 types of physicochemical properties in the AAIndex database [45].

2.2.6. Structural feature

This paper uses SPIDER3 [46] to generate an SPD file (<http://spark.s-lab.org>), which includes accessible surface area (ASA), secondary structure (SS), torsion angle (TA) and structural probability (SP), etc. ASA is a $N \times 1$ dimensional vector containing the accessible surface area values of all amino acid residues, and SS is also a $N \times 1$ dimensional vector, including three types: α -helix (H), β -fold (E) and random coil (C). TA is an $N \times 8$ matrix, including the sine and cosine values of ϕ, ψ, τ, θ . SP is an $N \times 3$ matrix. The obtained feature vectors can be analyzed via equations 12–17 proposed by Rayhan et al. [47].

(1) Secondary Structure Composition

$$SS-Composition(i) = \frac{1}{N} \sum_{j=1}^N c_{ij}, \quad 1 \leq i \leq 3 \quad (12)$$

where N is the length of the protein sequence, $c_{ij} = \begin{cases} 1, & SS_j = f_i \\ 0, & \text{else} \end{cases}$ $f_i = H \text{ or } E \text{ or } C$.

(2) Accessible surface area composition

$$ASA-Composition = \frac{1}{N} \sum_{i=1}^N ASA(i) \quad (13)$$

(3) Torsional angles composition

$$TA-Composition(k) = \frac{1}{N} \sum_{i=1}^N T_{i,k} \quad (1 \leq k \leq 8) \quad (14)$$

(4) Torsional angles bigram

$$TA - \text{bigram}(k, l) = \frac{1}{N} \sum_{i=1}^{N-1} T_{i,k} T_{i+1,l} (1 \leq k \leq 8, 1 \leq l \leq 8) \quad (15)$$

(5) Structural probabilities bigram

$$SP - \text{bigram}(k, l) = \frac{1}{N} \sum_{i=1}^{N-1} P_{i,k} P_{i+1,l} (1 \leq k \leq 3, 1 \leq l \leq 3) \quad (16)$$

(6) Torsional angles auto-covariance

$$TA - \text{Auto-Covariance}(k, j) = \frac{1}{N} \sum_{i=1}^{N-k} T_{i,j} T_{i+k,j} (1 \leq j \leq 8, 1 \leq k \leq DF) \quad (17)$$

where DF is the distance factor whose value is 10 [48].

(7) Structural Probabilities Auto-Covariance

$$SP - \text{Auto-Covariance}(k, j) = \frac{1}{N} \sum_{i=1}^{N-k} P_{i,j} P_{i+k,j} (1 \leq j \leq 3, 1 \leq k \leq DF) \quad (18)$$

To sum up, for each feature extraction of the target, the dimension of the obtained vectors are 273 (CTD), 343 (CT), $20 + \lambda$ (PseAAC), $20 + 20 \times \xi$ (PsePSSM), $8 \times nlag$ (NMBroto), 195 (structure feature).

2.3. Drug molecular substructure fingerprint

Molecular substructure fingerprints provide a good representation of drug compounds, especially binary substructure vectors. By detecting the presence of specific molecular structure fragments in the molecular structure, the drug can be converted into a binary feature vector. This paper utilizes the PaDEL-Descriptor [49] to calculate molecular fingerprint descriptors, describing the structure information by 881 chemical substructures collected from the PubChem database. Therefore, if the substructure exists in the drug molecule, the vector is set to 1 at the corresponding position. If the substructure does not exist, the value at this position is set to 0. These 881 chemical structure descriptions can be downloaded from (<https://pubchem.ncbi.nlm.nih.gov>), and the fingerprint attribute is "PUBCHEM_CACTVS_SUBGRAPHKEYS".

2.4. XGBoost feature selection

XGBoost feature selection [50] is employed to calculate the feature importance score through information gain (Equation (19)) in order to perform dimensionality reduction.

$$\text{gain} = \frac{1}{2} \left[\frac{\left(\sum_{i \in I_L} g_i \right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i \right)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{\left(\sum_{i \in I} g_i \right)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma \quad (19)$$

where $I = I_L \cup I_R$, and I_L is the number of samples in the left node space. λ and γ are regularization parameters. The higher the feature gain is, the more effective and important the feature is. XGBoost can be obtained from <https://github.com/dmlc/xgboost>.

2.5. Synthetic minority oversampling technique

Synthetic minority oversampling technique (SMOTE) [51] is used for synthesizing new samples to balance the data, and deal with the imbalanced prediction problem (The number of DTI samples is larger than that of non-DTI samples). SMOTE can be obtained using equation (20). First, the k nearest neighbors are determined by calculating the Euclidean distance (In this paper, we set k at 5). The random neighbor x_n

can be determined through the sample imbalance ratio N . Then the new samples can be obtained by combining the original samples according to equation (20).

$$x_{\text{new}} = x + \text{rand}(0, 1) \times |x - x_n| \quad (20)$$

2.6. Deep neural network

Deep neural network (DNN) was presented by Hinton et al. [52] using the idea of human brain learning. The input layer receives raw sample vectors. Each value multiplies by the weights at each node, and the output values can be obtained through a nonlinear activation function. During training, the weights and biases can be adjusted using stochastic gradient descent (Fig. 1 (A)). The mathematical representations are shown in equation (21).

$$a^l = \delta(Z^l) = \delta(w^l a^{l-1} + b^l) \quad (21)$$

where $a^l = 1, 2, \dots, N$, represent the input data of the layer, w^l is the connection weight matrix, b^l is the bias of the layer, and δ represents the activation function of the l -th layer. DNN-DTIs uses the cross-entropy loss function for optimization:

$$L = -\frac{1}{N} \sum_{i=1}^N \left(y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i) \right) \quad (22)$$

where N is the sample index, y_i and \hat{y}_i represent the true and predicted label, respectively.

ReLU (Rectified linear unit) has one-sided suppression characteristics, thus, all negative values are changed to 0, while positive values are unchanged (equation (23)).

$$\delta(z) = \max(0, z) \quad (23)$$

For this DL framework, the input layer consists of extracted features, followed by four hidden layers (the numbers of neurons are 600, 300, 150, and 75, respectively). The activation function used is ReLU, and the two neurons in the output layer represent the output probabilities of DTI and non-DTI, respectively. We use the Adam algorithm for optimization.

2.7. Performance evaluation and model construction

In order to evaluate the effectiveness of the model, this paper uses 5-fold cross-validation (5-fold CV) for evaluation. Sensitivity (SE), Specificity (SP), Matthew's correlation coefficient (MCC) and accuracy (ACC) are used to evaluate DNN-DTIs [53,54], which are described in equations (24)–(27).

$$SE = \frac{TP}{TP + FN} \quad (24)$$

$$SP = \frac{TN}{TN + FP} \quad (25)$$

$$MCC = \frac{(TP \times TN - FP \times FN)}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \quad (26)$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (27)$$

where TP , FP , TN and FN are true positive, false positive, true negative and false negative, respectively. ROC and PR curves are also important indicators and the areas under the curves are called AUC and AUPR, respectively. The larger the value, the better the performance of the model.

Our novel DTI prediction method is called DNN-DTIs, and is shown in Fig. 1. The reported experiments were conducted on Intel Xeon (E5-2650) with 32 GB of RAM and the model was written in the Python programming language. The code and datasets for DNN-DTIs are

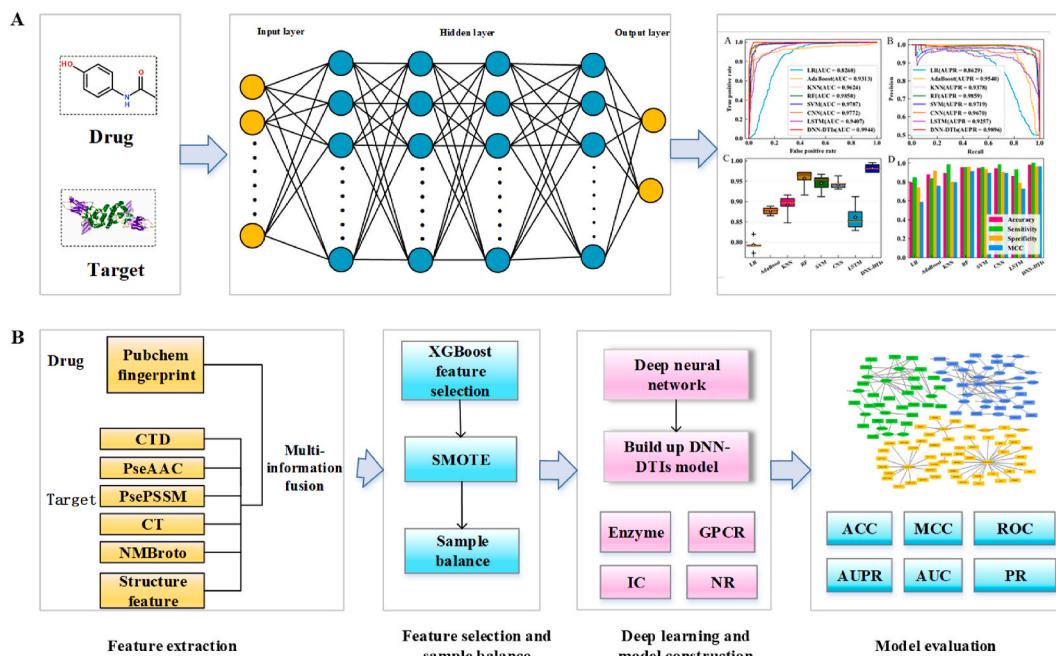


Fig. 1. DTIs prediction method based on a deep neural network. (A) represents the structure of the deep neural network and its inputs and outputs. The input should be a drug and a protein representing the target, and the output is the prediction label and evaluation indicators. (B) represents the pipeline of DNN-DTIs. This paper fuses CTD, PseAAC, PsePSSM, CT, NMBroto and structure feature to represent target information. The drug information is represented using PaDEL fingerprint. Then, XGBoost and SMOTE are used to perform the tasks of feature selection and sample balancing. Datasets: Enzyme, GPCR, IC and NR, the DNN-DTIs are used for model construction. Finally, the DTI networks are predicted and plotted.

available at <https://github.com/QUST-AIBBDRC/DNN-DTIs>. The computational process of DNN-DTIs can be described as:

- 1) Feature extraction. Physicochemical information, local and global information of sequence composition, and structural information of protein sequences are obtained through CTD, PseAAC, PsePSSM, CT, Moreau-Broto and structure feature. The chemical structure information of the drug can be characterized using molecular substructure fingerprints.
- 2) Feature selection and sample balancing. For eliminating noise, the importance score of each feature is calculated by XGBoost, and then, the essential low-dimensional features are obtained. SMOTE is employed to balance the samples of DTIs and non-DTIs, generating a balanced and reliable DTIs dataset.
- 3) Deep learning and model construction. The DTIs prediction model DNN-DTIs, based on a deep neural network is constructed and trained through parameter optimization on the described datasets. Compared with CNN, long short-term memory neural network (LSTM) and traditional ML methods, DNN-DTIs obtains better performance on Enzyme, IC, GPCR, NR, and Kuang's dataset.
- 4) Model evaluation. In addition to the evaluation process described above, the model is also evaluated using Network1 and Network2 which are not part of the original datasets. Furthermore, the predicted DTI networks are used to explore the biomedical significance of novel predicted links.

3. Results and discussion

3.1. Parameter optimization of DNN-DTIs model

Feature extraction consists of converting protein character signals into numerical signals, and fusing physicochemical, local and global sequence, structural and evolutional information of the target. In this paper, the parameter λ of PseAAC, ξ of PsePSSM, and $nlag$ of Moreau-Broto autocorrelation need to be selected for optimal performance.

Considering the shortest sequence length is 83 in gold standard datasets, this paper sets different values to optimize the parameters based on DNN (one input layer, four hidden layers and one output layer). The results on GSD via 5-fold CV are shown in Tables S3–S5.

For PseAAC, the highest accuracy is reached when $\lambda = 30$, $\lambda = 70$, $\lambda = 70$ and $\lambda = 80$ on Enzyme, IC, GPCR and NR, respectively (Table S3). Table S4 shows that the highest accuracy is achieved when $\xi = 0$, $\xi = 6$, $\xi = 6$ and $\xi = 8$ on Enzyme, IC, GPCR and NR, respectively. From Table S5, the average prediction accuracy of GSD is the greatest when $nlag = 6$. According to Tables S3–S5, the performance according to different single feature change, especially, the accuracies on the Enzyme dataset vary greatly, while the results for IC, GPCR and NR shows little variance. Finally, the optimal parameters of PseAAC, PsePSSM and Moreau-Broto autocorrelation are $\lambda = 30$, $\xi = 2$ and $nlag = 6$, respectively.

3.2. Comparison results of feature fusion versus single feature

Extracting effective information from protein sequences is critical for DTIs prediction. In this paper, the CTD and CT are adopted in order to capture amino acid sequence composition, and then PseAAC and Moreau-Broto are employed to capture the physicochemical properties, while PsePSSM is used to extract evolutionary information. Finally, we fuse these multiple features in order to comprehensively and effectively

Table 1
Comparison of single feature and feature fusion.

	Enzyme (%)	IC (%)	GPCR (%)	NR (%)
CTD	95.36	96.92	95.66	93.33
CT	97.16	97.56	95.32	92.50
PseAAC	94.87	94.91	93.82	94.52
PsePSSM	96.53	94.18	92.26	91.30
NMBroto	95.05	94.41	92.72	90.74
Structure feature	95.90	95.12	94.86	94.17
Fusion	98.01	98.34	95.27	96.30

characterize DTIs. According to section 3.1, the number of dimensions after feature fusion is 1850. The comparison results of feature fusion versus single feature through 5-fold Cross Validation are shown in Table 1. The ROC and PR curves of the individual features and multi-feature fusion are shown in Fig. S1. The vectors of each feature encoding algorithm are processed using SMOTE. The DNN is utilized as a classifier, and Fusion represents the results for multi-feature fusion.

Table 1 shows that using different special extraction methods results in different prediction accuracies for DTIs. In particular, the prediction result is shown to be better when fusing multiple features than when using single feature. With the exception of the GPCR dataset (CT and CTD), Fusion performance exceeds all other methods. This indicates that feature fusion can better characterize drugs and targets. Fig. S1 indicates that the ROC for Fusion has the largest coverage area, with AUC values at 0.9969, 0.9965, 0.9836, and 0.9858 on four GSD, respectively. Therefore, the prediction performance using multiple feature fusion is shown to be better and more robust.

3.3. XGBoost feature selection

When the XGBoost algorithm for feature selection is used, the dimension used has an important impact on the prediction results. Fusing CTD, CT, PseAAC, PsePSSM, NMBroto and structure feature results in a raw feature space with $\lambda = 30$, $\xi = 2$ and $nlag = 6$. We utilize XGBoost and the SMOTE algorithm to select features and balance the dataset, and we use a deep neural network is used for the prediction task. In order to determine the optimal and most effective features, we set the size of the feature subset to 100, 200, 300, 400, 500, 600, 700, 800, 900, and 1000 in turn. The main results of DNN-DTIs can be found in Tables S6–S9, and the ROC and PR curves are shown in Fig. S1 with various dimensions. Fig. S2 and Tables S6–S9 show that when employing XGBoost for feature selection in order to reduce the input dimension, the optimal dimension obtained for best performance is 300. The optimized features can provide high-level feature information representing DTIs for deep learning. When the dimension is set at 1000, the prediction results on the four gold standard datasets are the worst. When the size of the feature subset increases, the prediction accuracy rate decreases gradually. Therefore, the 300 optimal features thus obtained, were chosen to best characterize the DTIs and provide effective input representation for the DNN.

XGBoost, a tree-based feature selection method, can mine the nonlinear relationships between the DTI features and the labels, and consequently preserve the effective feature subset. We also tested the effect of choosing other feature selection methods on the prediction of DTIs. For information gain [55] (IG), GINI Index [56] (GINI), and max-relevance-max-distance [57] (MRMD), the number of retained optimal features is set to 300. For LASSO [58], the penalty parameter alpha is set to 0.01, and for the elastic net [59] (EN), the penalty parameters alpha and l1_ratio are set to 0.1 and 0.05. The prediction results and evaluation on Enzyme, IC, GPCR and NR are shown in Table 2 and Fig. 2. The optimal dimensions of different feature selections can be found in Table S10. We also obtained the comparison results of different feature selection methods in the dataset constructed by Kuang et al. [37] (Table S11), and the ROC and PR curves are shown in Fig. S3.

As shown in Table 2, the XGBoost feature selection methods all exceed the model prediction performance of information gain (IG), Gini coefficient (GINI), maximum correlation maximum distance (MRMD), LASSO and elastic net (EN). For Enzyme, XGBoost is 23.97% higher than IG (98.78% VS 74.81%) on ACC and 9.62% higher than LASSO (98.78% VS 89.16%). For the IC dataset, XGBoost's ACC values are 29.45%, 8.6%, 6.99%, 13.14%, and 10.27% higher than IG, GINI, MRMD, LASSO, and EN (98.6% VS 69.15%, 90%, 91.61%, 85.46%, 88.33%), and MCC value is 13.3% higher than MRMD (0.9722 VS 0.8392). For GPCR, the ACC value of XGBoost is 22.86%, 22.77%, 3.95%, 14.57%, 13.98% higher than IG, GINI, MRMD, LASSO and EN (97.98% VS 75.12%, 75.21%, 94.03%, 83.41% and 84%). In terms of MCC value, XGBoost is

Table 2

Comparison results of XGBoost and other feature selection methods on the gold standard dataset.

Datasets	Feature selection	ACC (%)	SE (%)	SP (%)	MCC
IC	IG	74.81	80.74	68.89	0.5005
	GINI	97.09	98.84	95.34	0.9424
	MRMD	94.56	98.63	90.49	0.8944
	LASSO	89.16	85.84	92.47	0.7851
	EN	92.50	95.55	89.45	0.8578
	XGBoost	98.78	99.71	97.85	0.9759
	IG	69.15	84.76	53.53	0.4032
	GINI	90.00	97.54	82.46	0.8104
	MRMD	91.61	97.31	85.91	0.8392
	LASSO	85.46	87.75	83.17	0.7128
GPCR	EN	88.33	93.76	82.90	0.7714
	XGBoost	98.60	99.62	97.58	0.9722
	IG	75.12	82.57	67.66	0.5083
	GINI	75.21	84.54	65.88	0.5138
	MRMD	94.03	98.61	89.45	0.8848
	LASSO	83.41	90.60	76.22	0.6765
NR	EN	84.00	91.71	76.30	0.6889
	XGBoost	97.98	99.90	96.06	0.9603
	IG	69.81	80.00	59.63	0.4066
	GINI	71.57	76.11	67.04	0.4344
	MRMD	93.06	97.96	88.15	0.8659
	LASSO	94.72	98.89	90.56	0.8978
XGBoost	EN	94.81	99.26	90.37	0.9005
	XGBoost	98.24	100.00	96.48	0.9656

45.2% higher than IG (0.9603 VS 0.5083). For the NR, XGBoost's ACC values are 28.43%, 26.67%, 5.18%, 3.52%, and 3.43% higher than IG, GINI, MRMD, LASSO, and EN (98.24% VS 69.81%, 71.57%, 93.06%, 94.72%, 94.81%). For MCC, the XGBoost outperform LASSO (0.9656 VS 0.8978) and EN (0.9656 VS 0.9005).

Fig. 2 (A) shows that for Enzyme, the AUC of XGBoost is 18.95%, 0.35%, 1.69%, 4.62%, 1.93% (0.9983 VS 0.8088, 0.9948, 0.9814, 0.9521, 0.979) higher than IG, GINI, MRMD, LASSO and EN. It can be seen from Fig. 2 (B), On AUPR, XGBoost is 21.61% better than IG (0.9972 VS 0.7811) and 4.57% better than LASSO (0.9972 VS 0.9515). For IC, XGBoost's AUC values are 25.19%, 3.62%, 2.36%, 7.02%, and 5.2% higher than IG, GINI, MRMD, LASSO, and EN (0.9964 VS 0.7445, 0.9602, 0.9728, 0.9262, 0.9444). Fig. 2 (E) indicates that on the GPCR dataset, the AUC values of XGBoost is 17.75%, 17.04%, 1.96%, 8.41%, 8.29% (0.994 VS 0.8165, 0.8236, 0.9744, 0.9099, 0.9111). Fig. 2 (G) indicates that for NR, the AUC value of XGBoost is 24.52%, 20.94%, 2.44%, 0.96%, 1.36% higher than IG, GINI, MRMD, LASSO, and EN (0.9944 VS 0.7492, 0.785, 0.97, 0.9848, 0.9808). XGBoost's AUPR values are 27.66%, 21.67%, 3.25%, 1.17%, 1.25% higher than other feature selection approaches (0.9896 VS 0.713, 0.7729, 0.9571, 0.9779, 0.9771).

From the above analysis, XGBoost feature selection methods all exceed the predicted performance of IG, GINI, MRMD, LASSO, and EN. XGBoost can calculate the variable score based on the tree structure and boosting algorithm to determine the important feature subset with a large contribution rate, which removes noise and redundancy and all effects that are not related to DTIs prediction. In other words, XGBoost can reduce the computational complexity as well as improve the operating efficiency. Other feature selection methods have their pros and cons, but XGBoost has been shown to be superior to information theory-based feature selection methods (IG, GINI, and MRMD) and regularization-based feature selection methods (LASSO and EN) for the prediction of DTIs. Therefore, XGBoost feature selection method is used in this paper to perform dimensionality reduction.

To analyze the effectiveness of tree-based XGBoost, we also use LightGBM [60], random forest (RF) [61], and extremely randomized tree (ET) [62] to select the optimal feature subset (Optimal subset sizes are all 300). The experimental results are shown in Table S12. We can see that XGBoost has the best performance again. The ACC difference between LightGBM and XGBoost is small, due to the fact that XGBoost

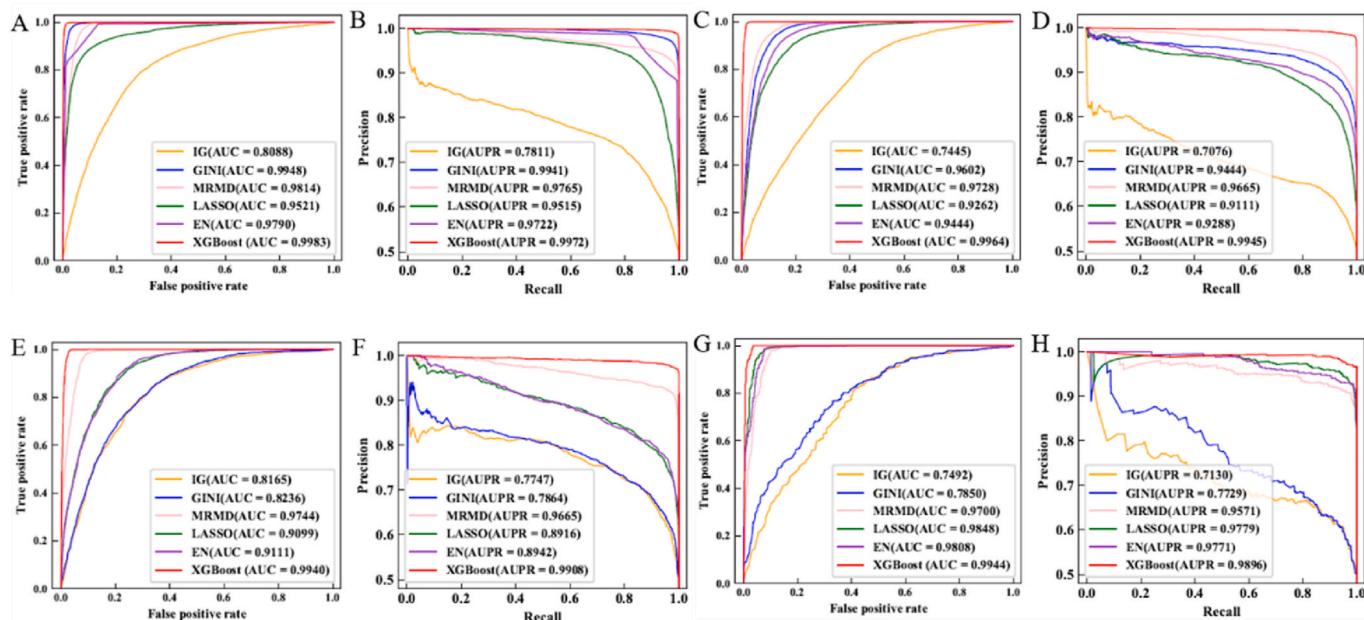


Fig. 2. ROC and PR curves of the gold standard datasets under IG, GINI, MRMID, LASSO, EN and XGBoost feature selection methods. (A–B), (C–D), (E–F) and (G–H) are ROC and PR curves of Enzyme, IC, GPCR and NR, respectively.

and LightGBM are both boosting-based algorithms. RF and ET are, on the other hand, bagging-based methods. According to importance analysis, we found that PsePSSM feature is the most important. In other words, after feature selection, the ratio of the number of selected PsePSSM features to the number of raw PsePSSM features is the highest. This means that evolution-based information of the target plays an important role in DTIs prediction.

3.4. Comparison results of DNN versus other classifiers

For studying the contribution of deep neural networks to predicting DTIs, we performed the following comparative analysis, in which we evaluate seven other classifiers alongside DNN-DTIs. The same low-dimensional feature information is input into the classifiers. These classifiers are logistic regression [63] (LR), AdaBoost [64] (the number of base decision trees is 400), K nearest neighbor algorithm [65] (KNN, neighbor size is set to 1), RF [61] (the value of n_estimators is set to 400), support vector machine [66] (SVM, using radial basis kernel function, the C value is 5), convolutional neural network [67] (CNN, incorporating a convolutional layer, a pooling layer, with ReLU activation function, and a kernel size of 10), long-short term memory neural network [68] (LSTM, with ReLU as activation function). The hyper-parameter optimization can be seen in supplementary material (Tables S13–S17). The main prediction results for GSD are shown in Tables 3–6. Furthermore, comparison results with other ML and DL methods, in terms of ROC and PR curves, box chart, and histogram are shown in Figs. 3–6.

Table 3
The main results of different classifiers for Enzyme dataset.

Methods	ACC (%)	SE (%)	SP (%)	MCC	AUC
LR	82.08	83.75	80.40	0.6420	0.8895
AdaBoost	90.77	90.27	91.26	0.8154	0.9658
KNN	95.65	99.90	91.41	0.9164	0.9565
RF	94.76	93.71	95.8	0.8953	0.9871
SVM	99.07	99.35	98.79	0.9814	0.9967
CNN	95.10	97.63	92.56	0.9033	0.9864
LSTM	94.28	96.24	92.32	0.8864	0.9819
DNN-DTIs	98.78	99.71	97.85	0.9759	0.9983

Table 4
The main results of different classifiers for IC dataset.

Methods	ACC (%)	SE (%)	SP (%)	MCC	AUC
LR	83.62	87.64	79.61	0.6747	0.8873
AdaBoost	89.84	88.47	91.2	0.7971	0.9619
KNN	93.90	98.66	89.15	0.8821	0.9390
RF	93.82	94.26	93.38	0.8765	0.9822
SVM	98.40	98.95	97.84	0.9681	0.9952
CNN	94.24	96.48	91.99	0.8861	0.9782
LSTM	89.55	94.77	84.33	0.7955	0.9546
DNN-DTIs	98.60	99.62	97.58	0.9722	0.9964

Table 5
The main results of different classifiers for GPCR dataset.

Methods	ACC (%)	SE (%)	SP (%)	MCC	AUC
LR	79.90	83.67	76.12	0.5999	0.8676
AdaBoost	90.13	88.58	91.68	0.8030	0.9590
KNN	93.67	98.85	88.5	0.8783	0.9367
RF	93.69	95.04	92.34	0.8741	0.9832
SVM	97.74	98.87	96.61	0.9553	0.9928
CNN	93.73	96.51	90.94	0.8763	0.9756
LSTM	85.52	91.08	79.97	0.7154	0.9273
DNN-DTIs	97.98	99.90	96.06	0.9603	0.9940

Table 6
The main results of different classifiers for NR dataset.

Methods	ACC (%)	SE (%)	SP (%)	MCC	AUC
LR	79.54	85.56	73.52	0.5972	0.8268
AdaBoost	89.26	86.3	92.22	0.7871	0.9339
KNN	92.87	97.04	88.7	0.8608	0.9287
RF	95.56	95.19	95.93	0.9115	0.9860
SVM	97.87	98.7	97.04	0.9581	0.9911
CNN	94.35	98.33	90.37	0.8906	0.9772
LSTM	86.11	92.96	79.26	0.7293	0.9407
DNN-DTIs	98.24	100.00	96.48	0.9656	0.9944

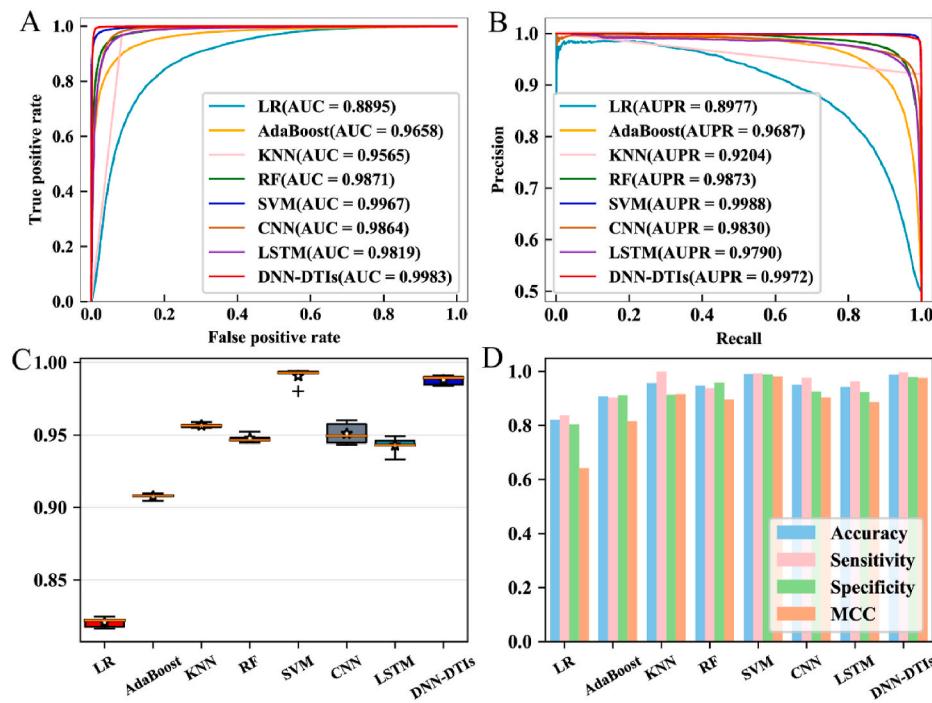


Fig. 3. ROC curves (A), PR curves (B), box plots (C) and histograms (D) of LR, AdaBoost, KNN, RF, SVM, CNN, LSTM and DNN-DTIs on Enzyme dataset.

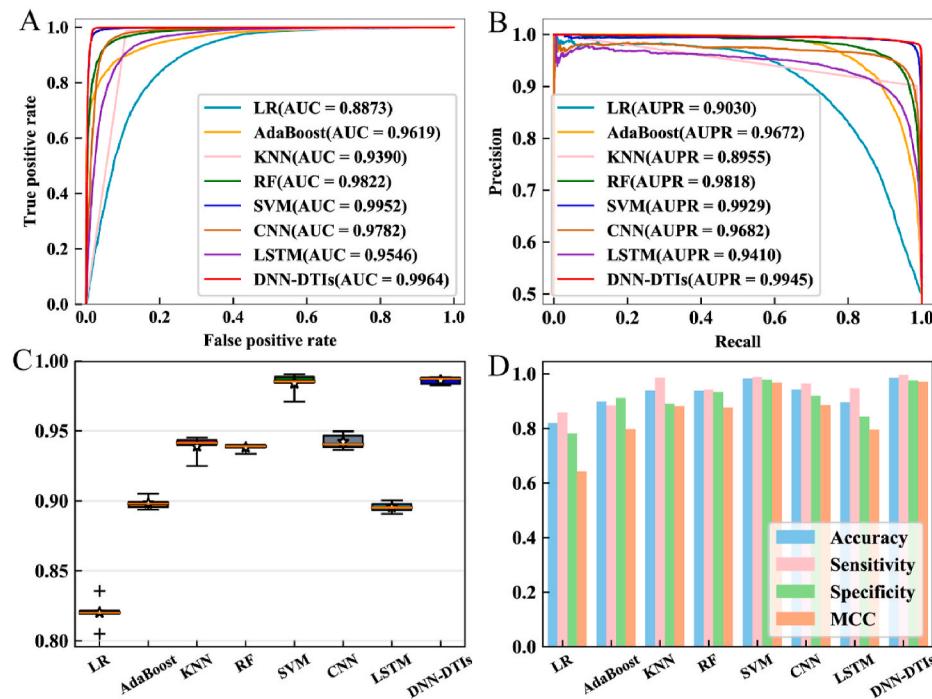


Fig. 4. ROC curves (A), PR curves (B), box plots (C) and histograms (D) of LR, AdaBoost, KNN, RF, SVM, CNN, LSTM and DNN-DTIs on IC dataset.

From Table 3 to Table 6, we can see that the prediction results of DNN-DTIs are better than the other 7 classifiers. Table 3 shows that for the Enzyme dataset, the ACC values of DNN-DTIs are 16.7%, 8.01%, 3.13%, 4.02%, 3.68% and 4.5% higher than LR, AdaBoost, KNN, RF, CNN, and LSTM, respectively (98.78% VS 82.08%, 90.77%, 95.65%, 94.76%, 95.10% and 94.28%). Table 4 shows that for the IC, the ACC values of DNN-DTIs are 14.98%, 8.76%, 4.7%, 4.78%, 0.2%, 4.36%, 4.98% and 9.05% higher than LR, AdaBoost, KNN, RF, SVM, CNN, and LSTM, respectively (98.60% VS 83.62%, 89.84%, 93.90%,

93.82%, 98.40%, 94.24% and 89.55%). Table 5 shows that for GPCR, the ACC values of DNN-DTIs are 18.08%, 7.85%, 4.31%, 4.29%, 0.24%, 4.25%, 3.8% and 12.46% higher than LR, AdaBoost, KNN, RF, SVM, CNN, and LSTM, respectively (97.98% VS 79.90%, 90.13%, 93.67%, 93.69%, 97.74%, 93.73%, 85.52%). Table 6 indicates that for NR, the ACC values of DNN-DTIs are 18.7%, 8.98%, 5.37%, 2.68%, 0.37%, 3.89% and 12.13% higher than LR, AdaBoost, KNN, RF, SVM, CNN and LSTM, respectively. (98.24% VS 79.54%, 89.26%, 92.87%, 95.56%, 97.87%, 94.35% and 86.11%).

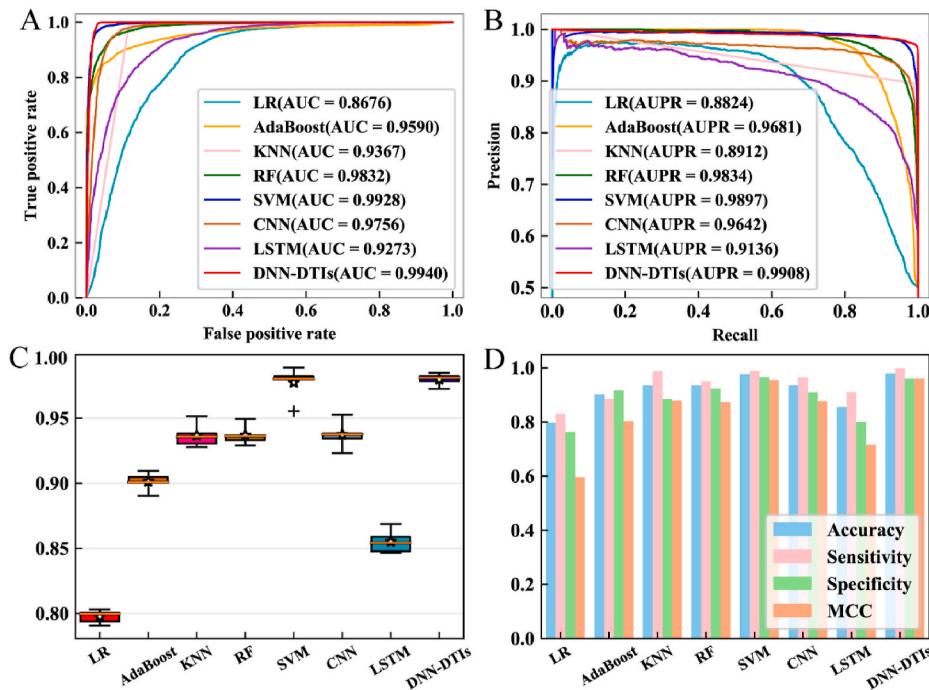


Fig. 5. ROC curves (A), PR curves (B), box plots (C) and histograms (D) of LR, AdaBoost, KNN, RF, SVM, CNN, LSTM and DNN-DTIs on GPCR dataset.

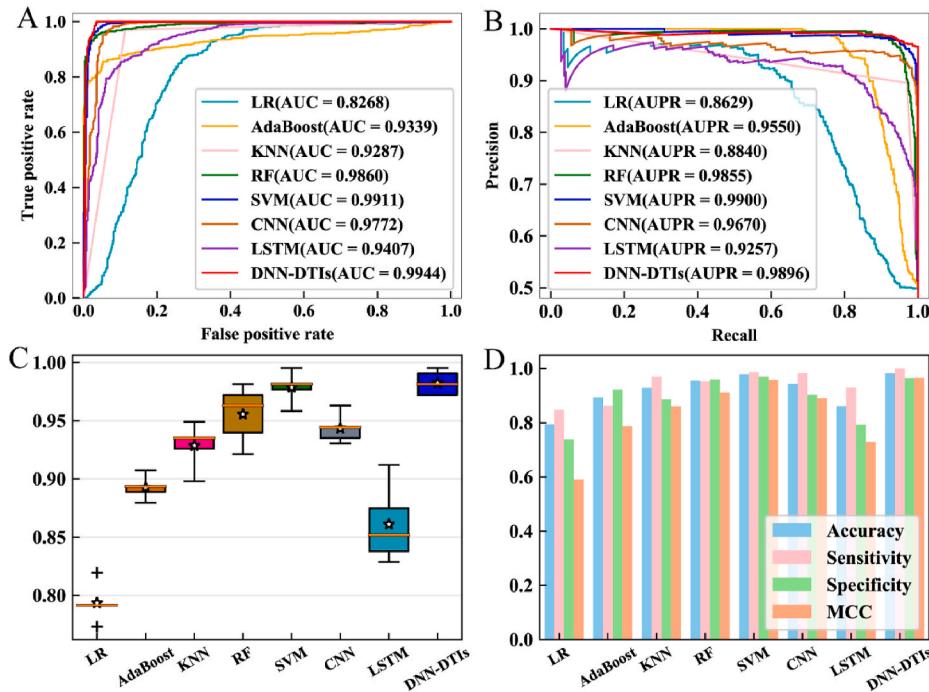


Fig. 6. ROC curves (A), PR curves (B), box plots (C) and histograms (D) of LR, AdaBoost, KNN, RF, SVM, CNN, LSTM and DNN-DTIs on NR dataset.

Fig. 3 indicates that for Enzyme, from the ROC curve, PR curve, box chart and bar chart, DNN-DTIs achieves better performance than that of LR, AdaBoost, KNN, RF, SVM, CNN and LSTM. The AUC value of DNN-DTIs is 0.9983, which is 1.19% better than CNN (0.9983 VS 0.9864) and 1.64% better than LSTM (0.9983 VS 0.9819). It can also be seen from the box plot and histogram that DNN-DTIs is superior in terms of prediction performance and robustness. **Fig. 4** indicates that for IC, the AUC values of DNN-DTIs are 10.91%, 3.45%, 5.74%, 1.42%, 0.12%, 1.82%, and 4.18% higher than LR, AdaBoost, KNN, RF, SVM, CNN, and LSTM, respectively (0.9964 VS 0.8873, 0.9619, 0.9390, 0.9822, 0.9952,

0.9782, 0.9546). The AUPR value of DNN-DTIs is also better than other classifier algorithms, and the box graph and histogram of GPCR can also illustrate its superiority. **Fig. 6** shows that for NR, the AUC values of DNN-DTIs are 16.76%, 6.05%,

6.57%, 0.84%, 0.33%, 1.72%, 5.37% higher than LR, AdaBoost, KNN, RF, SVM, CNN, and LSTM, respectively. (0.9944 VS 0.8268, 0.9339, 0.9287, 0.9860, 0.9911, 0.9772, and 0.9407). From **Fig. 6 (B)** we can see that the area of the PR curve of DNN-DTIs is higher than other classifiers, in particular, it is 2.26% higher than CNN (0.9896 VS 0.9670). **Fig. 6(C)** and **(D)** also demonstrate that the prediction

performance is better than the other classifiers in terms of the main indicators.

This paper also obtained the results of DNN, LR, AdaBoost, KNN, RF, SVM, CNN, and LSTM on the dataset constructed by Kuang et al. The ACC, SE, SP, MCC and AUC values are shown in Table S18. ROC curve, PR curve, box plot and histogram results are shown in Fig. S4. From the above analysis, it is evident that DNN can effectively establish the nonlinear relationship between drug fingerprint information and target information and categories. This paper also uses a two-tailed T-test as a statistical test for significance. At significance level 0.05, the P-values for DNN-DTIs alongside the other different classifier algorithms on ACC are shown in Table S19. It can be concluded that the classic classifier algorithms (not based on deep learning) are not best suited to dealing with the inherent high-level feature information of DTIs, and hence, their prediction performance is hindered. Deep learning, on the other hand, can identify and predict DTIs and non-DTIs through hierarchical learning. Furthermore, compared with other DL based models (CNN and LSTM), DNN performs best for DTIs prediction.

3.5. Comparison with existing DTIs prediction methods

Considering the importance of DTIs, a large number of DTI prediction methods have been developed, including PDTPS of Meng et al. [69], PSSM + DVM of Li et al. [70], GIP + RLS of Laarhoven et al. [71], BLM-NII of Mei et al. [72], NLRMF of Liu et al. [73], SAR of Cao et al. [74], LRF-DTIs of Shi et al. [25] and Kuang et al. [37]. To ensure the objectiveness of the comparison results, we list the evaluation indicators

Table 7
Comparison of different DTIs prediction methods on GSD and Kuang dataset.

Datasets	Methods	AUC	SE (%)	SP (%)	ACC (%)
Enzyme	PDTPS ^a	N/A	97.44	98.02	97.73
	PSSM + DVM ^b	0.9288	92.90	93.19	93.16
	GIP + RLS ^c	0.9780	88.00	99.90	N/A
	BLM-NII ^d	0.9880	N/A	N/A	N/A
	NRLMF ^e	0.9870	N/A	N/A	N/A
	SAR ^f	0.9486	90.10	90.64	90.31
	LRF-DTIs ^g	0.9982	97.66	98.51	98.09
	DNN-DTIs ^h	0.9983	99.71	97.85	98.78
IC	PDTPS ^a	N/A	93.32	92.95	93.12
	PSSM + DVM ^b	0.9171	92.65	90.72	91.73
	GIP + RLS ^c	0.9840	29.00	100.00	N/A
	BLM-NII ^d	0.9900	N/A	N/A	N/A
	NRLMF ^e	0.9890	N/A	N/A	N/A
	SAR ^f	0.9428	89.38	88.20	88.91
	LRF-DTIs ^g	0.9965	96.71	97.93	97.32
	DNN-DTIs ^h	0.9964	99.62	97.58	98.60
GPCR	PDTPS ^a	N/A	84.89	88.35	86.77
	PSSM + DVM ^b	0.8856	89.27	89.45	89.37
	GIP + RLS ^c	0.9540	33.10	100.00	N/A
	BLM-NII ^d	0.9840	N/A	N/A	N/A
	NRLMF ^e	0.9690	N/A	N/A	N/A
	SAR ^f	0.8902	82.54	85.49	84.68
	LRF-DTIs ^g	0.9918	95.26	96.11	95.69
	DNN-DTIs ^h	0.9940	99.90	96.06	97.98
NR	PDTPS ^a	N/A	92.63	84.44	87.78
	PSSM + DVM ^b	0.9300	96.62	87.78	92.22
	GIP + RLS ^c	0.9220	15.60	100.00	N/A
	BLM-NII ^d	0.9810	N/A	N/A	N/A
	NRLMF ^e	0.9500	N/A	N/A	N/A
	SAR ^f	0.8822	82.35	84.72	83.74
	LRF-DTIs ^g	0.9559	93.85	95.81	94.88
	DNN-DTIs ^h	0.9944	100	96.48	98.24
Kuang	Kuang et al ⁱ	0.9350	N/A	N/A	N/A
	DNN-DTIs ^h	0.9957	98.73	97.27	98.00

N/A means not available.

(a) is from literature [69], using 5-fold CV. (b) is from Ref. [70], using 5-fold CV. (c) is from Ref. [71] and uses leave-one-out CV. (d) is from Ref. [72], using 10-fold CV. (e) comes from the literature [73], using 10-fold CV. (f) is from Ref. [74], using 5-fold CV. (g) is from Ref. [25], using 5-fold CV. (h) uses 5-fold CV to evaluate the DNN-DTIs. (i) is from Kuang et al. [37].

for DNN-DTIs and others on the same gold standard datasets of Enzyme, IC, GPCR, NR and Kuang et al.'s (Table 7).

From Table 7, overall, the performance and indicators of DNN-DTIs are the best. For Enzyme, the AUC values of DNN-DTIs are 6.95%, 2.03%, 1.03%, 1.13%, 4.97%, and 0.01% higher than PSSM + DVM, GIP + RLS, BLM-NII, NRLMF, SAR, and LRF-DTIs, respectively (0.9983 VS 0.9288, 0.978, 0.988, 0.987, 0.9486 and 0.9982). The ACC value of DNN-DTIs is 5.62% higher than PSSM + DVM (98.78% VS 93.16%) and 0.69% higher than LRF-DTIs (98.78% VS 98.09%). For IC, the AUC values of DNN-DTIs are 7.93%, 1.24%, 0.64%, 0.74%, and 5.36% higher than PSSM + DVM, GIP + RLS, BLM-NII, NRLMF, and SAR, respectively (0.9964 VS 0.9171, 0.9840, 0.9900, 0.989, 0.9428 and 0.9965). The ACC value of DNN-DTIs is 9.69% better than SAR (98.60% VS 88.91%) and 1.28% higher than LRF-DTIs (98.60% VS 97.32%). For GPCR, The AUC values of DNN-DTIs are 10.84%, 4%, 1%, 2.5%, 10.38%, and 0.22% higher than PSSM + DVM, GIP + RLS, BLM-NII, NRLMF, SAR, and LRF-DTIs (0.994 VS 0.8856, 0.954, 0.984, 0.969, 0.8902, 0.9918). The ACC value of DNN-DTIs is 2.29% higher than LRF-DTIs (97.98% VS 95.69%). For NR, The AUC values of DNN-DTIs are 6.44%, 7.24%, 1.34%, 4.44%, 11.22%, and 3.85% higher than PSSM + DVM, GIP + RLS, BLM-NII, NRLMF, SAR, and LRF-DTIs (0.9944 VS 0.93, 0.922, 0.981, 0.95, 0.8822 and 0.9559). From the perspective of ACC indicators, the value of DNN-DTIs is 10.46% higher than PDTPS (98.24% VS 87.78%), 6.02% higher than PSSM + DVM (98.24% VS 92.22%), and 3.36% higher than LRF-DTIs (98.24% VS 94.88%). In terms of AUC, DNN-DTIs are also superior to the method proposed by Kuang et al. (0.9957 VS 0.9350).

As shown in Table 7, DNN-DTIs achieves the best performance, compared with PSSM + DVM, GIP + RLS, BLM-NII, NRLMF, SAR, LRF-DTIs and Kuang et al. There are a number of reasons which can explain this outcome. First, the fusion of a comprehensive and effective set of features provides appropriate information for the prediction of DTIs, as opposed to the use of single feature information. Furthermore, the tree-based feature selection method used, XGBoost, can retain better DTIs representation without reducing model complexity. In addition, the sample balancing algorithm, SMOTE, which is employed to synthesize artificial samples for the minority class, provides an optimized and balanced set of input feature vectors. Finally, of particular importance, the layer-by-layer learning ability of DNN can effectively establish the non-linear relationship between drug-target interactions and labels. This allows for the extraction of important and relevant high-level feature information.

3.6. Prediction of the drug-target interaction network

Predicting the DTIs network can contribute to the study of its topological properties, as well as the identification of new DTIs. DTIs network analysis can also illustrate the potential biological significance of newly predicted interaction links. We use drugs and targets as nodes on the graph, where edges indicate interactions. This paper uses Network1 and Network2 to demonstrate and verify the utility of DNN-DTIs. First, the raw features from PseAAC, PsePSSM, CTD, CT, NMBroto, structure and substructure fingerprint are concatenated. Then, the optimized features are obtained through XGBoost and SMOTE as previously elaborated. Finally, DNN-DTIs is trained on the NR dataset in order to predict Network1 (Fig. 7). The prediction score of each DTI in Network1 can be found in Table S20. As shown in Fig. 8, Network2 is composed of three subnetworks, where Enzyme is represented in orange, IC is represented in green, and GPCR is represented in blue. The gold standard datasets of Enzyme, IC, and GPCR are used as training sets to predict the three subnetworks. This training set is used to eliminate the drug-target interaction pairs that appear in Network2, and the obtained prediction results can be found in Fig. 8.

Fig. 7 demonstrates that the trained DNN-DTIs model can effectively predict the DTIs in Network1, and only one pair of DTI D03675-hsa2099 is not predicted successfully (The accuracy is 96.97% (32/33)). The

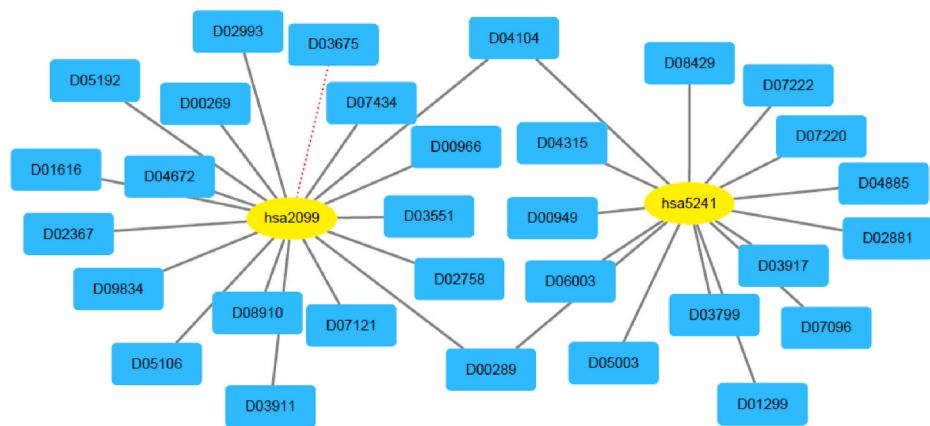


Fig. 7. The prediction results of the drug-target interactions Network1. The circular box represents the target, and the rectangular box represents the drug. The solid black edge in the figure indicates that the DTI is predicted successfully, while the red dashed edge indicates that the DTI is not successfully predicted. It can be seen that for 33 pairs of DTIs, only 1 pair failed to be predicted.

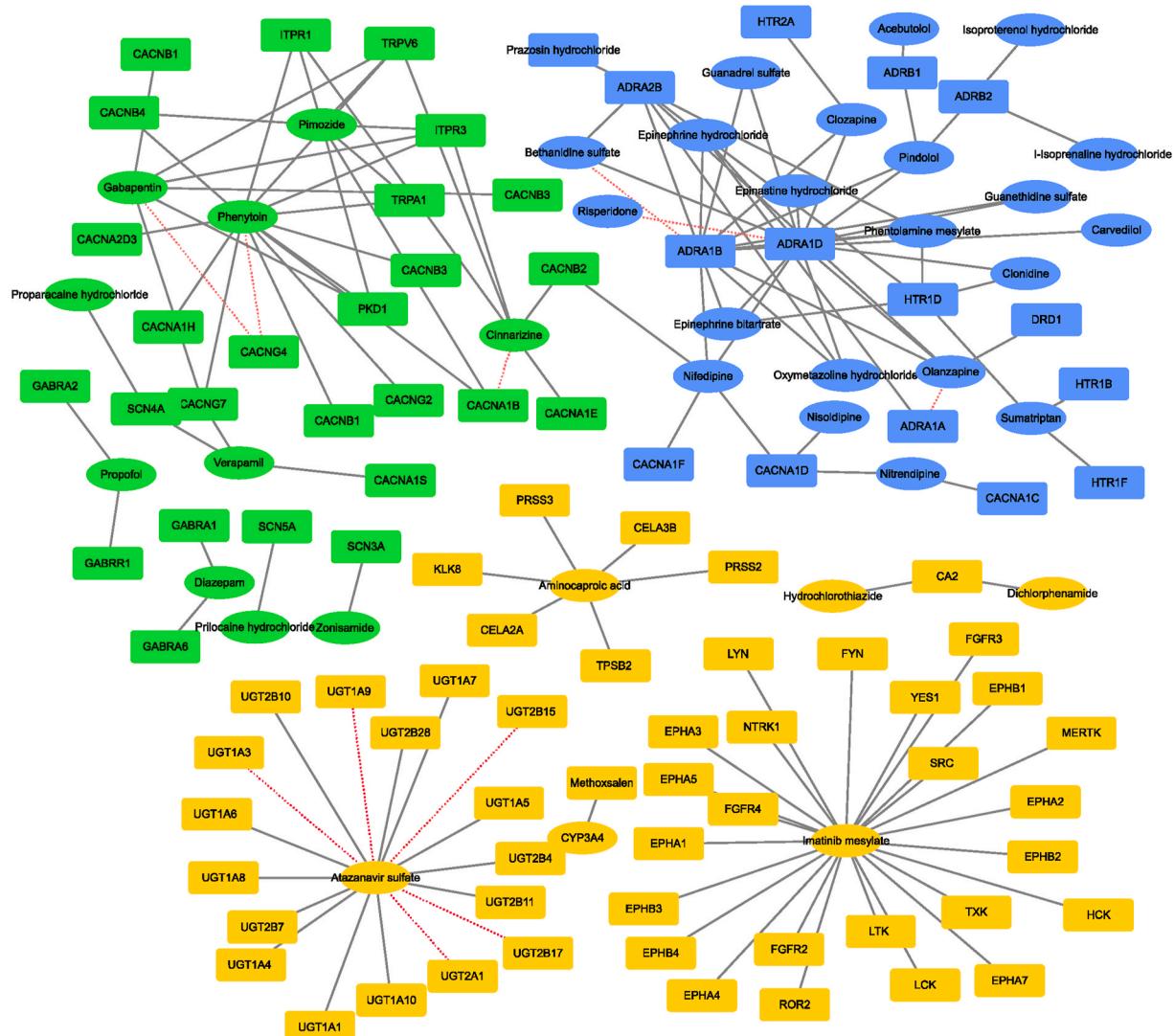


Fig. 8. Prediction results of drug-target interactions Network2. The circular box represents the drug, the rectangular box represents the target, the orange part represents subnetwork Enzyme, the green part represents subnetwork IC, and the blue part represents subnetwork GPCR. The solid black line indicates the DTI is successfully predicted, and the red dashed edge indicates that the DTI is not successfully predicted. It can be seen that for 150 pairs of DTIs, only 11 pairs of DTIs are not predicted successfully.

accuracy of Shi et al. [25] is 85.29% (29/33), and DNN-DTIs achieve better performance than LRF-DTIs. The drug name, target name, and prediction scores for Network1 are shown in Table S20. From Fig. 8, for Enzyme, UGT2A1-Atazanavir sulfate, UGT1A9-Atazanavir sulfate, UGT1A3-Atazanavir sulfate, UGT2B17-Atazanavir sulfate and UGT2B15-Atazanavir sulfate are not predicted successfully (see orange subnetwork in Fig. 8). For GPCR, ADRA1D-Risperidone, ADRA1B-Bethanidine sulfate and ADRA1A-Olanzapine are not predicted successfully (see blue subnetwork in Fig. 8). Three pairs of drug-target interactions in the IC dataset: CACNG4-Gabapentin, CACNG4-Phenytoin, and CACNA1B-Cinnarizine are not predicted successfully (see the green subnetwork in Fig. 8).

Studies have shown that the combination of the drug Etonogestrel with the receptor can reduce the secretion of luteinizing hormone and can be used to make female contraceptives [75]. The DNN-DTIs prediction method can successfully predict the drug-target interaction D04104-hsa2099 (Etonogestrel-Estrogen receptor). It can be seen that Pimozide in the IC dataset is a core drug, and the interaction between Pimozide with PKD1 can be successfully predicted. PKD1 is a gene closely related to polycystic kidney disease. The drug Aglepristone is a progesterone receptor antagonist [76], which can treat various progesterone dependent physiological or pathological states, and has certain effects on diabetes mellitus. Alfatradiol can treat female hair loss [77], obtaining a suitable target can help treat hair loss and diabetes. DNN-DTIs can successfully predict D07096-hsa5241 (Aglepristone-Progesterone receptor) and D07121-hsa2099 (Alfatradiol-Estrogen receptor). According to the Enzyme subnetwork, Imatinib mesylate is a small molecule kinase inhibitor which is used to treat cancer-related diseases [78]. DNN-DTIs can predict the interaction between Imatinib mesylate and surrounding targets. The tazanavir sulfate-UGT1A10 interaction is successfully predicted. When tazanavir sulfate binds to the receptor, it can play a therapeutic effect on inflammation and insulin resistance [79]. Pindolol is a 5-HT1A receptor antagonist, which can be used for the design of antidepressants after binding to receptor targets [80]. DNN-DTIs can effectively predict the Pindolol-ADRB1 interaction (the prediction score is 0.9956). Studies have shown that voltage-gated sodium channel is Phenytoin's main target for anti-arrhythmic drug, and has been used in the treatment of many diseases [81]. DNN-DTIs can successfully predict the Phenytoin-TRPA1 interaction.

In the above, we have demonstrated how our model can predict reliable drug-target interactions from candidate pairs. For Netwrk1, 32 out of the 33 interacting drug target pairs in the nuclear receptor dataset are successfully predicted. For Network2, out of 150 pairs of DTIs, only 11 pairs are not predicted successfully. Nowadays, there are mostly ligands and targets. DNN-DTIs could detect drug-target pairs that are not yet validated. For example, if we regard D00269-hsa2099 (<https://www.genome.jp/entry/hsa:2099>) as an unknown interaction pair, DNN-DTIs could predict this interaction successfully, using only features that are agnostic to any established relationship between the drug and the target, or the lack thereof. In other words, considering potential new interactions, DNN-DTIs can help researchers determine which pairs are best candidates for interaction, which can in turn help fulfill “new uses of old medicine” clinically. The code for our model DNN-DTIs can be downloaded from <https://github.com/QUST-AIBBDRC/DNN-DTIs>. We believe DNN-DTIs can be used according to the needs of the researcher for the purposes of drug-target interaction prediction and analysis.

4. Conclusion

Prediction of DTIs can be used for many important applications, such as drug repositioning, drug design, etc. This paper proposes a novel DTIs prediction pipeline, DNN-DTIs, based on deep learning. In the process of developing our model, we aim to emphasize the following main contributions: (1) Multi-information fusion provides useful embedding vectors, which simultaneously encompass multiple types of features, including sequence-based, structure-based, fingerprint-based and

evolution-based information. (2) XGBoost can select top-ranking features without losing prediction accuracy. Comparison with IG, GINI, MRMD, LASSO and EN, XGBoost was performed in order to analyze and establish feature importance. (3) DNN-DTIs model can obtain high-level features of biological significance from raw DTIs information through hierarchical learning. DNN-DTIs is shown to be appropriate and effective for the task of DTIs prediction both on the training and testing datasets. (4) DNN-DTIs can be used to predict drug-target interaction networks successfully with high accuracy, which can in turn provide new ideas about potential novel links and new approaches for drug design, human disease prevention, etc.

Graph convolutional neural networks (GCNN) could be used to mine high-level network topology feature information. In the future, we seek to combine a GCNN model with residual neural networks, or a generative adversarial network (GAN), to develop a prediction model for drug-target interactions prediction, which will mine very important and useful DTI information.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 61863010), the Key Research and Development Program of Shandong Province of China (No. 2019GGX101001), and the Key Laboratory Open Foundation of Hainan Province (No. JSKX2020001).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.combiomed.2021.104676>.

References

- [1] F. Cheng, W. Li, Z. Wu, X. Wang, C. Zhang, J. Li, G. Liu, Y. Tang, Prediction of polypharmacological profiles of drugs by the integration of chemical, side effect, and therapeutic space, *J. Chem. Inf. Model.* 53 (53) (2013) 753–762.
- [2] M.J. Keiser, V. Setola, J.J. Irwin, C. Laggner, A.I. Abbas, S.J. Hufeisen, N.H. Jensen, M.B. Kuijer, R.C. Matos, T.B. Tran, R. Whaley, R.A. Glennon, J. Hert, K.L. H. Thomas, D.D. Edwards, B.K. Shoichet, B.L. Roth, Predicting new molecular targets for known drugs, *Nature* 462 (462) (2009) 175–181.
- [3] Z. Wu, F. Cheng, J. Li, W. Li, G. Liu, Y. Tang, SDTNBL: an integrated network and chemoinformatics tool for systematic prediction of drug-target interactions and drug repositioning, *Briefings Bioinf.* 18 (18) (2017) 333–347.
- [4] S. Guenther, M. Kuhn, M. Dunkel, M. Campillos, C. Senger, E. Petsalaki, J. Ahmed, E.G. Urdiales, A. Gewiss, L.J. Jensen, R. Schneider, R. Skoblo, R.B. Russell, P. E. Bourne, P. Bork, R. Preissner, SuperTarget and Matador: resources for exploring drug-target relationships, *Nucleic Acids Res.* 36 (36) (2008) D919–D922.
- [5] M. Kanehisa, S. Goto, M. Hattori, K.F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, M. Hirakawa, From genomics to chemical genomics: new developments in KEGG, *Nucleic Acids Res.* 34 (34) (2006) D354–D357.
- [6] D.S. Wishart, C. Knox, A.C. Guo, D. Cheng, S. Shrivastava, D. Tzur, B. Gautam, M. Hassanali, DrugBank: a knowledgebase for drugs, drug actions and drug targets, *Nucleic Acids Res.* 36 (36) (2008) D901–D906.
- [7] X. Chen, Z.L. Ji, Y.Z. Chen, TTD: therapeutic target database, *Nucleic Acids Res.* 30 (30) (2002) 412–415.
- [8] X. Chen, C.C. Yan, X. Zhang, X. Zhang, F. Dai, J. Yin, Y. Zhang, Drug-target interaction prediction: databases, web servers and computational models, *Briefings Bioinf.* 17 (17) (2016) 696–712.
- [9] S. D’Souza, K.V. Prema, B. Seetharaman, Machine learning models for drug-target interactions: current knowledge and future directions, *Drug Discov. Today* 25 (25) (2020) 748–756.
- [10] M.J. Keiser, B.L. Roth, B.N. Armbruster, P. Ernsberger, J.J. Irwin, B.K. Shoichet, Relating protein pharmacology by ligand chemistry, *Nat. Biotechnol.* 25 (25) (2007) 197–206.
- [11] G. Pujadas, M. Vaque, A. Ardrevol, C. Blade, M.J. Salvado, M. Blay, J. Fernandez-Larrea, L. Arola, G. Pujadas, Protein-ligand docking: a review of recent advances and future perspectives, *Curr. Pharmaceut. Anal.* 4 (4) (2008) 1–19.

- [12] A. Ezzat, M. Wu, X.L. Li, C.K. Kwoh, Computational prediction of drug-target interactions using chemogenomic approaches: an empirical survey, *Briefings Bioinf.* 20 (20) (2019) 1337–1357.
- [13] L. Jacob, J.P. Vert, Protein-ligand interaction prediction: an improved chemogenomics approach, *Bioinformatics* 24 (24) (2008) 2149–2156.
- [14] S.J. Opella, Structure determination of membrane proteins by nuclear magnetic resonance spectroscopy, *Annu. Rev. Anal. Chem.* 6 (6) (2013) 305–328.
- [15] M. Bagherian, E. Sabeti, K. Wang, M.A. Sartor, Z. Nikolovska-Coleska, K. Najarian, Machine learning approaches and databases for prediction of drug-target interaction: a survey paper, *Briefings Bioinf.* (2020), <https://doi.org/10.1093/bib/bbz157>.
- [16] Y. Ding, J. Tang, F. Guo, Identification of drug–target interactions via dual laplacian regularized least squares with multiple kernel fusion, *Knowl.-Based Syst.* 204 (204) (2020) 106254.
- [17] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, M. Kanehisa, Prediction of drug–target interaction networks from the integration of chemical and genomic spaces, *Bioinformatics* 24 (24) (2008) i232–i240.
- [18] Y. Yamanishi, M. Kotera, M. Kanehisa, S. Goto, Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework, *Bioinformatics* 26 (26) (2010) i246–i254.
- [19] J.Y. Shi, S.-M. Yiu, Y. Li, H.C.M. Leung, F.Y.L. Chin, Predicting drug–target interaction for new drugs using enhanced similarity measures and super-target clustering, *Methods* 83 (83) (2015) 98–104.
- [20] R.S. Olayan, H. Ashoor, V.B. Bajic, DDR: efficient computational method to predict drug–target interactions using graph mining and machine learning approaches, *Bioinformatics* 34 (34) (2018) 1164–1173.
- [21] A. Ezzat, P. Zhao, M. Wu, X.L. Li, C.K. Kwoh, Drug-target interaction prediction with graph regularized matrix factorization, *IEEE ACM T. Comput. Bi.* 14 (14) (2017) 646–656.
- [22] L.Y. Xia, Z.Y. Yang, H. Zhang, Y. Liang, Improved prediction of drug–target interactions using self-paced learning with collaborative matrix factorization, *J. Chem. Inf. Model.* 59 (59) (2019) 3340–3351.
- [23] J. Zhang, M. Zhu, P. Chen, B. Wang, DrugRPE: random projection ensemble approach to drug–target interaction prediction, *Neurocomputing* 228 (228) (2017) 256–262.
- [24] L. Li, M. Cai, Drug target prediction by multi-view low rank embedding, *IEEE ACM T. Comput. Bi.* 16 (16) (2019) 1712–1721.
- [25] H. Shi, S. Liu, J. Chen, X. Li, Q. Ma, B. Yu, Predicting drug–target interactions using Lasso with random forest based on evolutionary information and chemical structure, *Genomics* 111 (111) (2019) 1839–1852.
- [26] Y. Chu, A.C. Kaushik, X. Wang, W. Wang, Y. Zhang, X. Shan, D.R. Salahub, Y. Xiong, D.Q. Wei, DTi-CDF: a cascade deep forest model towards the prediction of drug–target interactions based on hybrid features, *Briefings Bioinf.* 18 (22) (2021) 451–462.
- [27] P. Kumar, S. Bankapur, N. Patil, An enhanced protein secondary structure prediction using deep learning framework on hybrid profile based features, *Appl. Soft Comput.* 86 (86) (2020) 105926.
- [28] N.Q.K. Le, D.T. Do, T.N.K. Hung, L.H.T. Lam, T.T. Huynh, N.T.K. Nguyen, A computational framework based on ensemble deep neural networks for essential genes identification, *Int. J. Mol. Sci.* 21 (21) (2020) 9070.
- [29] N.Q.K. Le, Q.T. Ho, T.T.D. Nguyen, Y.Y. Ou, A transformer architecture based on BERT and 2D convolutional neural network to identify DNA enhancers from sequence information,, *Briefings Bioinf.* (2021), <https://doi.org/10.1093/bib/bbab005>.
- [30] N.Q.K. Le, E.K.Y. Yapp, N. Nagasundaram, H.Y. Yeh, Classifying promoters by interpreting the hidden information of DNA sequences via deep learning and combination of continuous FastText N-Grams, *Front. Bioeng. Biotechnol.* 7 (7) (2019) 305.
- [31] M.A. Thafar, R.S. Olayan, H. Ashoor, S. Albaradei, V.B. Bajic, X. Gao, T. Gojobori, M. Essack, DTiGEMS+: drug–target interaction prediction using graph embedding, graph mining, and similarity-based techniques, *J. Cheminf.* 12 (12) (2020) 44.
- [32] H. Öztürk, A. Ozgur, E. Ozkirimli, DeepDTA: deep drug–target binding affinity prediction, *Bioinformatics* 34 (34) (2018) 821–829.
- [33] F. Rayhan, S. Ahmed, Z. Mousavian, D.M. Farid, S. Shatabda, FRnet-DT: Deep Convolutional Neural Networks with Evolutionary and Structural Features for Drug–Target Interaction, 2018 arXiv:1806.07174.
- [34] I. Lee, J. Keum, H. Nam, DeepConv-DT: prediction of drug–target interactions via deep learning with convolution on protein sequences, *PLoS Comput. Biol.* 15 (15) (2019) e1007129.
- [35] X. Zeng, S. Zhu, Y. Hou, P. Zhang, L. Li, J. Li, L.F. Huang, S.J. Lewis, R. Nussinov, F. Cheng, Network-based prediction of drug–target interactions using an arbitrary-order proximity embedded deep forest, *Bioinformatics* 36 (36) (2020) 2805–2812.
- [36] T. Zhao, Y. Hu, L.R. Valsdottir, T. Zang, J. Peng, Identifying drug–target interactions based on graph convolutional network and deep neural network, *Briefings Bioinf.* (2020), <https://doi.org/10.1093/bib/bbaa044>.
- [37] Q. Kuang, X. Xu, R. Li, Y. Dong, Y. Li, Z. Huang, Y. Li, M. Li, An eigenvalue transformation technique for predicting drug–target interaction, *Sci. Rep.* 5 (5) (2015) 13867.
- [38] A.P. Bento, A. Gaulton, A. Hersey, L.J. Bellis, J. Chambers, M. Davies, F.A. Krueger, Y. Light, L. Mak, S. McGlinchey, M. Nowotka, G. Papadatos, R. Santos, J. P. Overington, The ChEMBL bioactivity database: an update,, *Nucleic Acids Res.* 42 (42) (2014) D1083–D1090.
- [39] I. Dubchak, I. Muchnik, S.R. Holbrook, S.H. Kim, Prediction of protein-folding class using global description of amino acid sequence, *P. Natl. Acad. Sci. USA* 92 (92) (1995) 8700–8704.
- [40] J. Shen, J. Zhang, X. Luo, W. Zhu, K. Yu, K. Chen, Y. Li, H. Jiang, Predicting protein–protein interactions based only on sequences information, *P. Natl. Acad. Sci. USA* 104 (104) (2007) 4337–4341.
- [41] K.C. Chou, Prediction of protein cellular attributes using pseudo-amino acid composition, *Proteins* 43 (43) (2001) 246–255.
- [42] H.B. Shen, K.C. Chou, Nuc-PLoc: a new web-server for predicting protein subnuclear localization by fusing PseAA composition and PsePSSM, *Protein Eng. Des. Sel.* 20 (20) (2007) 561–567.
- [43] S.F. Altschul, T.L. Madden, A.A. Schaffer, J.H. Zhang, Z. Zhang, W. Miller, D. J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* 25 (25) (1997) 3389–3402.
- [44] D.S. Horne, Prediction of protein helix content from an autocorrelation analysis of sequence hydrophobicities, *Biopolymers* 27 (27) (1988) 451–477.
- [45] S. Kawashima, H. Ogata, M. Kanehisa, AAindex: amino acid index database, *Nucleic Acids Res.* 27 (27) (1999) 368–369.
- [46] R. Heffernan, Y. Yang, K. Paliwal, Y. Zhou, Capturing non-local interactions by long short-term memory bidirectional recurrent neural networks for improving prediction of protein secondary structure, backbone angles, contact numbers and solvent accessibility, *Bioinformatics* 33 (33) (2017) 2842–2849.
- [47] F. Rayhan, S. Ahmed, S. Shatabda, D.M. Farid, Z. Mousavian, A. Dehzangi, M. S. Rahman, iDTI-ESBoost: identification of drug target interaction using evolutionary and structural features with boosting, *Sci. Rep.* 7 (7) (2017) 17731.
- [48] G. Taherzadeh, Y. Zhou, A.W.C. Liew, Y. Yang, Sequence-based prediction of protein–carbohydrate binding sites using support vector machines, *J. Chem. Inf. Model.* 56 (56) (2016) 2115–2122.
- [49] C.W. Yap, PaDEL-Descriptor: an open source software to calculate molecular descriptors and fingerprints, *J. Comput. Chem.* 32 (32) (2011) 1466–1474.
- [50] T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794.
- [51] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, *J. Artif. Intell. Res.* 16 (16) (2002) 321–357.
- [52] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, *Science* 313 (313) (2006) 504–507.
- [53] X. Wang, B. Yu, A. Ma, C. Chen, B. Liu, Q. Ma, Protein–protein interaction sites prediction by ensemble random forests with synthetic minority oversampling technique, *Bioinformatics* 35 (35) (2019) 2395–2402.
- [54] B. Yu, W. Qiu, C. Chen, A. Ma, J. Jiang, H. Zhou, Q. Ma, SubMito-XGBoost: predicting protein submitochondrial localization by fusing multiple feature information and eXtreme gradient boosting, *Bioinformatics* 36 (36) (2020) 1074–1081.
- [55] C.K. Lee, G.G. Lee, Information gain and divergence-based feature selection for machine learning-based text categorization, *Inf. Process. Manag.* 42 (42) (2006) 155–165.
- [56] Y. Wang, M. Tseng, Attribute selection for product configurator design based on Gini index, *Int. J. Prod. Res.* 52 (52) (2014) 6136–6145.
- [57] Q. Zou, J. Zeng, L. Cao, R. Ji, A novel features ranking metric with application to scalable visual and bioinformatics data classification, *Neurocomputing* 173 (173) (2016) 346–354.
- [58] R. Tibshirani, Regression shrinkage and selection via the Lasso, *J. Roy. Stat. Soc. B* 58 (58) (1996) 267–288.
- [59] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, *J. Roy. Stat. Soc. B* 67 (67) (2005) 301–320.
- [60] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.Y. Liu, LightGBM: a highly efficient gradient boosting decision tree, in: *31st Conference Neural Information Processing Systems* pp. 3146–3154.
- [61] L. Breiman, Random forests, *Mach. Learn.* 45 (45) (2001) 5–32.
- [62] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, *Mach. Learn.* 63 (63) (2006) 3–42.
- [63] R.E. Fan, K.W. Chang, C.J. Hsieh, X.R. Wang, C.J. Lin, LIBLINEAR: a library for large linear classification, *J. Mach. Learn. Res.* 9 (9) (2008) 1871–1874.
- [64] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, *J. Comput. Syst. Sci.* 55 (55) (1997) 119–139.
- [65] F. Nigsch, A. Bender, B. van Buuren, J. Tissen, E. Nigsch, J.B.O. Mitchell, Melting point prediction employing k-nearest neighbor algorithms and genetic parameter optimization, *J. Chem. Inf. Model.* 46 (46) (2006) 2412–2422.
- [66] C. Cortes, V. Vapnik, Support vector networks, *Mach. Learn.* 20 (20) (1995) 273–297.
- [67] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, *Commun. ACM* 60 (60) (2017) 84–90.
- [68] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.* 9 (9) (1997) 1735–1780.
- [69] F.R. Meng, Z.H. You, X. Chen, Y. Zhou, J.Y. An, Prediction of drug–target interaction networks from the integration of protein sequences and drug chemical structures, *Molecules* 22 (22) (2017) 1119.
- [70] Z. Li, P. Han, Z.-H. You, X. Li, Y. Zhang, H. Yu, R. Nie, X. Chen, In silico prediction of drug–target interaction networks based on drug chemical structure and protein sequences, *Sci. Rep.* 7 (7) (2017) 11174.
- [71] T.V. Laarhoven, S.B. Nabuurs, E. Marchiori, Gaussian interaction profile kernels for predicting drug–target interaction, *Bioinformatics* 27 (27) (2011) 3036–3043.
- [72] J.P. Mei, C.K. Kwoh, P. Yang, X.L. Li, J. Zheng, Drug–target interaction prediction by learning from local information and neighbors, *Bioinformatics* 29 (29) (2013) 238–245.
- [73] Y. Liu, M. Wu, C. Miao, P. Zhao, X.L. Li, Neighborhood regularized logistic matrix factorization for drug–target interaction prediction, *PLoS Comput. Biol.* 12 (12) (2016) e1004760.

- [74] D.S. Cao, S. Liu, Q.S. Xu, H.M. Lu, J.H. Huang, Q.N. Hu, Y.Z. Liang, Large-scale prediction of drug-target interactions using protein sequences and drug topological structures, *Anal. Chim. Acta* 752 (752) (2012) 1–10.
- [75] M. Binkowska, J. Woron, Progestogens in menopausal hormone therapy, *Prz. menopauzalny* 14 (14) (2015) 134–143.
- [76] A. Gogny, F. Fieni, Aglepristone: a review on its clinical use in animals, *Theriogenology* 85 (85) (2016) 555–566.
- [77] U. Blume-Peytavi, C. Kunte, A. Krisp, N. Garcia Bartels, U. Ellwanger, R. Hoffmann, Comparison of the efficacy and safety of topical minoxidil and topical alfatradiol in the treatment of androgenetic alopecia in women, *J. Dtsch. Dermatol. Ges* 5 (5) (2007) 391–395.
- [78] B. Gobin, G. Moriceau, B. Ory, C. Charrier, R. Brion, F. Blanchard, F. Redini, D. Heymann, Imatinib mesylate exerts anti-proliferative effects on osteosarcoma cells and inhibits the tumour growth in immunocompetent murine models, *PloS One* 9 (9) (2014) e90795.
- [79] M. Auclair, A.C. Guenantin, S. Fellahi, M. Garcia, J. Capeau, HIV antiretroviral drugs, dolutegravir, maraviroc and ritonavir-boosted atazanavir use different pathways to affect inflammation, senescence and insulin sensitivity in human coronary endothelial cells, *PloS One* 15 (15) (2020) e0226924.
- [80] F. Artigas, A. Adell, P. Celada, Pindolol augmentation of antidepressant response, *Curr. Drug Targets* 7 (7) (2006) 139–147.
- [81] J.M.K. Hesselink, D.J. Kopsky, Phenytoin: 80 years young, from epilepsy to breast cancer, a remarkable molecule with multiple modes of action, *J. Neurol* 264 (264) (2017) 1617–1621.