DBGRU-SE: Predicting drug-drug interactions based on double BiGRU and squeeze-and-excitation attention mechanism

Article *in* Briefings in Bioinformatics · May 2023

DOI: 10.1093/bib/bbad184

CITATIONS
0

READS
155

6 authors, including:

Hongli Gao
Qingdao University of Science and Technology
1 PUBLICATION **0** CITATIONS
SEE PROFILE

Xin Liao
Qingdao University of Science and Technology
1 PUBLICATION **0** CITATIONS
SEE PROFILE

Baoxing Ning
Qingdao University of Science and Technology
1 PUBLICATION **0** CITATIONS
SEE PROFILE

Haiming Gu
Qingdao University of Science and Technology
47 PUBLICATIONS **1,548** CITATIONS
SEE PROFILE

OXFORD

Briefings in Bioinformatics, 2023, 1–9
https://doi.org/10.1093/bib/bbad184
Problem Solving Protocol

DBGRU-SE: predicting drug–drug interactions based on double BiGRU and squeeze-and-excitation attention mechanism

Mingxiang Zhang, Hongli Gao, Xin Liao, Baoxing Ning, Haiming Gu and Bin Yu

Corresponding authors: Bin Yu, College of Information Science and Technology, School of Data Science, Qingdao University of Science and Technology, Qingdao 266061, China, and School of Data Science, University of Science and Technology of China, Hefei 230027, China. E-mail: yubin@qust.edu.cn; Haiming Gu, College of Mathematics and Physics, Qingdao University of Science and Technology, Qingdao 266061, China. E-mail: ghm@qust.edu.cn

Abstract

The prediction of drug–drug interactions (DDIs) is essential for the development and repositioning of new drugs. Meanwhile, they play a vital role in the fields of biopharmaceuticals, disease diagnosis and pharmacological treatment. This article proposes a new method called DBGRU-SE for predicting DDIs. Firstly, FP3 fingerprints, MACCS fingerprints, Pubchem fingerprints and 1D and 2D molecular descriptors are used to extract the feature information of the drugs. Secondly, Group Lasso is used to remove redundant features. Then, SMOTE-ENN is applied to balance the data to obtain the best feature vectors. Finally, the best feature vectors are fed into the classifier combining BiGRU and squeeze-and-excitation (SE) attention mechanisms to predict DDIs. After applying five-fold cross-validation, The ACC values of DBGRU-SE model on the two datasets are 97.51 and 94.98%, and the AUC are 99.60 and 98.85%, respectively. The results showed that DBGRU-SE had good predictive performance for drug–drug interactions.

Keywords: drug–drug interactions, Group Lasso, SMOTE-ENN, double BiGRU, squeeze-and-excitation attention mechanism

INTRODUCTION

Drug–drug interactions (DDIs) prediction is one of the most important research elements in bioinformatics, which plays an essential role in the diagnosis of diseases and the study of pharmacological mechanisms. Drug–drug interactions (DDIs) are interactions that occur between two or more dosing processes, which include pharmacological changes and toxic side effects [1]. Therefore, the study of DDIs prediction has great practical importance. However, traditional laboratory methods use time-consuming and labor-intensive biological or pharmacological methods [2]. Therefore, it is essential to predict drug–drug interactions.

A number of methods have been used for the prediction of DDIs. The basic approach to predicting DDIs is the traditional laboratory-based approach [3], which is carried out by biological and pharmacological tools. It is difficult to experiment with various combinations of drugs in trials because of the multiple possible combinations, thus identifying adverse effects that were not previously identified. However, traditional methods are time-consuming and cost-effective, and their ability to identify DDIs is very limited [4]. Therefore, deep learning-based methods have been widely used to predict DDIs in recent years.

The primary task of predicting DDIs based on deep learning methods is to extract the feature information of drugs. The accurate extraction of various feature information is a key step in predicting DDIs. DrugBank [5], FAERS [6], TWOSIDES [7] are some of the most common databases. The common feature extraction methods include drug chemical information and molecular structure information. Deep learning-based methods obtain representative feature vectors for predicting DDI by constructing different drug data and deep neural network models. Compared with traditional classification algorithms, deep learning algorithms generally yield better prediction results. The study shows that methods of integrating drug features and drug molecular structures can yield better results.

Although existing deep learning-based methods have achieved good results for the prediction of DDIs, there is still room for further improvement. Firstly, single feature extraction methods are not able to comprehensively characterize drug information, while fusion of multiple feature information can lead to redundant features and noise information. Secondly, the problem of positive and negative sample imbalance causes a significant impact on the performance of the classifier, which leads to biased prediction results. Finally, the predictive power of existing classifiers for predicting DDIs has some room for improvement. Inspired by this, we use the following three strategies: (1) We extract multiple feature vector information of the drug and perform multi-information

Mingxiang Zhang is a master student at the Qingdao University of Science and Technology, China. Her research interests include bioinformatics and deep learning.
Hongli Gao is a master student at the Qingdao University of Science and Technology, China. Her research interests include bioinformatics and deep learning.
Xin Liao is a master student at the Qingdao University of Science and Technology, China. His research interests include bioinformatics and deep learning.
Baoxing Ning is a master student at the Qingdao University of Science and Technology, China. Her research interests include bioinformatics and deep learning.
Haiming Gu is a professor at the Qingdao University of Science and Technology, China. His research interests include artificial intelligence, machine learning and algorithms.
Bin Yu is a professor at the Qingdao University of Science and Technology, China. His research interests include bioinformatics, artificial intelligence and biomedical image processing.
Received: January 27, 2023. **Revised:** April 3, 2023. **Accepted:** April 23, 2023

fusion. (2) In order to reduce the information redundancy caused by feature fusion, Group Lasso is used to reduce the dimensionality of the fused features. SMOTE-ENN is also applied to deal with the imbalance between positive and negative samples. (3) The classifier combining double BiGRU and SE attention mechanisms is used for the prediction of DDIs. The main contributions of this work are as follows:

- We designed a new prediction framework DBGRU-SE to predict DDIs. Molecular structure information of drugs is extracted using FP3 fingerprints, MACCS fingerprints, Pubchem fingerprints, and 1D and 2D molecular descriptors. The multi-information fusion features are better than single features, which improves the prediction accuracy of the model.
- In order to extract effective features, Group Lasso method is used to reduce the dimensionality of the fused features and retain the best feature subset. SMOTE-ENN method is also used to balance the positive and negative sample data to reduce the impact of data imbalance on the prediction results.
- We employed a classifier combining double BiGRU and SE attention mechanisms to predict DDIs. DBGRU-SE has good stability and can effectively predict drug–drug interactions.

DDIs is an important focus of drug safety monitoring. It aims to identify and quantify the risks associated with the use of drugs to better understand the causative mechanisms. DBGRU-SE can improve predictive power, and diverse DDIs can guide further studies of fine-scale biological consequences or events, such as the effect of DDIs on drug metabolism, risk or severity of adverse reactions, serum concentrations and antihypertensive activity of drugs. Predicting DDIs can help to understand the mechanisms behind adverse reactions, guide combination drug use and develop timely responses to adverse drug reactions.

RELATED WORK

Deep learning-based methods are widely used to predict DDIs recently. Chen et al. [8] proposed MUFFIN, a new model based on medicinal chemistry and biomedical KG, which can fuse structural information from the drug with KG for drug characterization. Feng et al. [9] proposed DNNs based on deep neural networks and graph convolutional networks (GCNs) to extract the structural features of drugs based on the characteristics of DDI networks. Shi et al. [10] obtained more features by directly processing the molecular map of the drug and proposed SSI-DDI, which combines chemical substructures and GCNs. Lin et al. [11] combined Siamese Network (SN), Convolutional Neural Network (CN), Autoencoder (AE) and self-attention mechanism to build MDF-SA-DDI, which integrates the fusion of multi-source drugs and multisource feature.

Many recent approaches to predicting drug–drug interactions follow the paradigms of graph learning. Ma et al. [12] proposed a bipartite graph neural network called DGNN-DDI, which is designed as a directed message-passing neural network (SA-DMPNN) with substructure attention mechanism to self-adaptively extract substructures. Feng et al. [13] proposed a new graph representation learning model SGRL-DDI for multi-task prediction of DDIs. The model can capture the task-joint information by integrating relation graph convolutional networks with Balance and Status patterns. He et al. [14] proposed a new deep learning architecture, 3DGT-DDI, a model composed of a 3D graph neural network and pre-trained text attention mechanism. This model uses 3D molecular graph structure and position information to enhance the predictive power of model. Feng et al. [15] proposed a new model DWAT-DDI for predicting asymmetric interactions between drugs. The model characterizes asymmetric DDIs by source role embedding, target role embedding and subrule embedding. Ryu et al. [16] proposed a method of DeepDDI based on DNN, which uses the structural similarity of drug pairs as feature vectors and uses PCA for dimensionality reduction. Guo et al. [17] proposed the deep learning model MSResG, which combined multi-source features of drugs with graph autoencoders for prediction. Su et al. [18] proposed an attention-based learning framework for KG representation, DDKG. the embedding of drugs was first initialized with an encoder-decoder layer to improve the performance of prediction. Hong et al. [19] proposed a link-aware graph-attentive network LaGAT for drug interaction prediction tasks. It can generate different attentional pathways for the same drug entity in different DDIs, providing interpretability to the results predicted by the KG model.

Table 1. Information of the dataset

Datasets	Drugs	Interactions	No-link
Dataset1	545	47,947	100,737
Dataset2	566	21,094	138,801

MATERIALS AND METHODS

Datasets

The selection of datasets is essential for the construction of prediction models. This paper chose two datasets that have been used in previous studies. We used a dataset collected and built by Zhang et al. [20] from Drugbank [21] as the first dataset (Dataset1), which contains 548 drugs and 48 584 drug interactions. The second dataset (Dataset2) was collected and constructed from DrugBank by Shi et al. [22], which contains drugs with chemical structures or without off-label side effects documented in OFFSIDES [23], a total of 568 drugs and 21 351 drug interactions. Drug mol format files are necessary for feature extraction, and several drugs did not have mol format, therefore they were deleted. Information on the dataset is shown in Table 1.

Feature extraction

Drug compound molecules are popular research subjects in medicine. The numerical representation of drug compound molecules is the prerequisite for implementing simulation calculations. The accuracy of the drug compound molecule depends directly on the accuracy of the final prediction. The numerical representation of molecules of pharmaceutical compounds is known as molecular descriptors. Molecular descriptor is a numerical indicator calculated based on the molecular structure to measure the properties of a molecule in a particular aspect, which can include physical and chemical properties [24]. Molecular descriptors are mainly divided into quantitative and qualitative categories. The quantitative molecular descriptor is a basic molecular weight expressed quantitatively, which is used to describe the physicochemical properties, molecular fields, molecular composition and molecular shapes of molecules [25]. Qualitative molecular descriptors are termed molecular fingerprints, which can be characterized by some kind of code to describe the structure and properties of a molecule [26].

Molecular descriptors can be classified as 1D, 2D and 3D according to the dimensions of the desired molecular structure. The calculation of molecular fingerprints for common compounds can be achieved by using only 2D structures, which improves the efficiency of molecular descriptor calculations and reduces the workload of screening. Molecular fingerprints can accurately represent the molecular structure of a drug compared to traditional descriptors. Therefore, choosing appropriate molecular fingerprints can facilitate the rapid screening of suitable drugs and minimize the risk of drug production.

To better describe the drug, different features are chosen to describe the drug according to its characteristics and the research problem. In this paper, FP3 fingerprints, MACCS fingerprints, Pubchem fingerprints, and 1D and 2D molecular descriptors are selected to represent the properties of drugs. FP3 is a fingerprint method created from a set of SMARTS patterns defining functional groups. MACCS is a fingerprint based on substructure keys, which contains most atomic properties, properties of different topologies of chemical bonds, and atomic neighborhoods. Pubchem is a fingerprint-based substructure key that can be used for fingerprinting similarity searches. It has 881 distinct noncoding substructures that span a wide spectrum of substructures and characteristics. 1D and 2D molecular descriptors. 1D and 2D molecular descriptors encode chemical composition and topology, respectively. 1D molecular descriptors are more commonly employed in representing compounds utilizing SMILES to describe atoms and inter-atomic bonds. 2D molecular descriptors are calculated with 2D molecular figures or structural fragments. The complementarity between the four features can enhance the characterization of drug properties to some extent. All drug pairs can be represented by an 1128-dimensional feature vector. The process of feature extraction is shown in Supplementary Si1–Si2.

Group Lasso

Feature fusion can produce more comprehensive drug information, and it can also produce redundant data. Lasso [27] is a L1-regularization based compressive estimating and linear regression method that accomplishes feature sparsity and feature selection by making the learned feature weights to be zero. Lasso tends to select a single variable based on a single input variable rather than on a group of input variables, which usually results in the selection of more factors than necessary. Group Lasso is proposed by Yuan and Lin [28], which is a method for selecting group variables. Group Lasso regression algorithm can group all variables and penalize the L2 norm of each set, resulting in the coefficients of the entire group being reduced to zero.

Considering the estimation problem of the parameters θ and σ^2 of the linear model

$$Y = X\theta + k, E(k) = 0, Cov(k) = \sigma^2 I \tag{1}$$

- where Y represents the observation vector, X represents the design matrix, θ represents the unknown parameter vector, k is the random error, and σ^2 is the variance of the error. The purpose of equation (1) is to select significant factors for accurate prediction. For a vector $\eta \in R^d$ and a symmetric positive definite matrix $d \times d$, Group Lasso estimator is defined as follows:

$$\|\eta\|_A = \left(\eta^T A \eta\right)^{\frac{1}{2}} \tag{2}$$

$$\hat{\theta}^{GLasso} = \arg\min_{\theta} \left\{ \left\| Y - \sum_{j=1}^{J} X_j \theta_j \right\|_2^2 + \lambda \sum_{j=1}^{J} \|\theta_j\|_2 \right\} \tag{3}$$

- where $\lambda \geq 0$ represents the adjustment parameter. The group of variables is not included in the model if all of the group variables' coefficients are zero. However, if the group variables' coefficients are not zero, the model will include all of the variables from that group.

SMOTE combining edited nearest neighbors

The dataset has a positive and negative sample imbalance, which may impair model's performance and predictive output. In order to reduce the impact of data imbalance, SMOTE combining edited nearest neighbors (SMOTE-ENN) method is used to balance the data. Chawla et al. [29] presented SMOTE method, which was used to synthetic the same number of new samples for each minority class sample [30, 31] SMOTE method works on the principle that each minority class sample is analyzed, and then new samples are manually synthesized and added to the dataset based on the minority class samples. First, the distance of each sample in the minority class a_i and all the samples of the minority class is calculated according to the Euclid distance, and k-nearest neighbor minority samples $a_{i(n)}$ of the sample are found. Second, the sampling ratio is determined by the unbalanced proportion of samples, and several samples are randomly selected in the-nearest neighbor of minority class samples. Finally, each randomly selected adjacent point $b_{i(n)}$ and the original sample a_i establish new sample values c_i respectively. The calculation process of c_i is shown in Equation (4):

$$c_i = a_i + \theta_i \left(b_i - a_i\right) \tag{4}$$

SMOTE algorithm may cause overfitting problems. The above problems need to be solved by data cleaning methods. ENN [32] is a common data-cleaning technique. The basic idea of ENN [33] method is to remove the neighboring samples in the majority class samples that are different from its own class. SMOTE-ENN [34] is a sampling method for clearing noise samples that combines oversampling and under-sampling.

Double bi-directional gated recurrent unit (BiGRU)

Gated recurrent unit (GRU) neural network is a deformation of recurrent neural network (RNN). RNN [35] is a model specifically designed to process sequence information. LSTM [36] is a special structure of RNN, and GRU is a variant of LSTM. Compared with ordinary RNN, LSTM with storage unit can store previous information for a long time. The prior transmission state and current input information are controlled by gated state in the LSTM. LSTM is made up of forget gate, input gate and output gate. GRU neural network is a simplified version of LSTM proposed by Cho et al. [37]. GRU neural networks are composed of the reset gates and the update gates. Reset gates define the previously stored and new input patterns, and update gates determine the amount of previously stored information up to the current time. Two gating vectors determine the output of the gating loop unit. Both gating mechanisms preserve long sequences of information, which are not deleted over time or irrelevant to the prediction.

The input and output structure of GRU is the same as that of generic RNN. There is a current input x_t and the concealed state of

a prior node h_{t-1}, which provides information about the preceding node. GRU can obtain the output of currently hidden node and send it to the hidden state of the next node by combining x_t and h_{t-1}.

$$Z_t = \sigma\left(x_t W^{(z)} + h_{t-1} U^{(z)} + b^{(z)}\right) \tag{5}$$

$$R_t = \sigma\left(x_t W^{(r)} + h_{t-1} U^{(r)} + b^{(r)}\right) \tag{6}$$

$$h_t' = \tanh\left\{x_t W^{(h)} + \left(h_t * R_t\right) U^{(h)} + b^{(h)}\right\} \tag{7}$$

$$h_t = (1 - Z_t) * h_{t-1} + Z_t * h_t' \tag{8}$$

where, R_t and Z_t represent the reset gate and update gate, respectively; σ is sigmoid nonlinear activation function; $b^{(r)}$ and $b^{(h)}$ represent the bias vector; $*$ represents element multiplication; $W^{(z)}$, $W^{(r)}$, $W^{(h)}$ and $U^{(z)}$, $U^{(r)}$, $U^{(h)}$ represent the input and output weight matrices corresponding to the current moment of neurons, respectively. The structure of GRU can effectively overcome the influence of short-term memory, thus improving the stability of the system. One-way GRU transmission only propagates one direction from forward to backward, which results in the loss of critical information. BiGRU can enable the information from positive propagation to the network and then use that information to fully utilize essential features backward, which permits the network to obtain more features. BiGRU obtains two different hidden layer states using forward and backward calculations, and then the vector obtained by concatenating the two vectors is used as the final encoded representation.

In order to consider the information of the input feature vectors, this paper employs double BiGRU to consider the information of the input feature vectors.

Squeeze-and-excitation attention mechanism

Attention mechanism [38] has made important breakthroughs in many fields such as image and natural language processing in recent years, which has been shown to be beneficial to improve the performance of the model. Attention model in practice can be regarded as attention distribution. The fundamental idea of attention mechanism aims to break the restriction that traditional encoders and decoders depend on fix-length vectors internally to encode and decode.

Squeeze-and-excitation (SE) attention mechanism [39] is a relatively common attention mechanism in deep learning. The three components of SE attention mechanism are Squeeze operation, Excitation operation and Scale operation. Squeeze operation is used to globally average the input feature layers for pooling to obtain the global receptive field of the vector to some extent. Excitation operation makes two complete linkages of the input features. The first fully connected layer has fewer neurons, and the second fully connected layer has the same number of neurons as the input feature layer, which returns the feature dimension to the original dimension. Then a Sigmoid function is used to obtain the normalized weights between 0 and 1. Scale operation regards the output weight of Excitation as the importance of each feature channel, which is multiplied by the previous features to complete the rescaling of the original features.

Performance evaluation

The cross-validation method is a common approach to evaluate the generalization ability of machine learning models. The k-fold cross-validation method is a practical method to divide the data sample into smaller subsets from a statistical perspective. The 5-fold cross-validation is used to evaluate the predictive performance of the model in this paper. First, the dataset is partitioned into five equal-sized, disjoint subsets at random. Then, the test samples are taken from one of the five subsets, and the training samples are taken from the remaining four subsets. The cycle is repeated five times using different subsets as the test set, and the test subsets and training subsets are not duplicated. Finally, the final validation result is presented as the average of all five results.

This study used the following evaluation metrics: Accuracy (ACC), Precision, Sensitivity (SE), Specificity (SP), F1-score and Matthew's correlation coefficient (MCC) [40–42].

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \tag{9}$$

$$SE = \frac{TP + TN}{TP + FP + TN + FN} \tag{10}$$

$$SP = \frac{TN}{TP + FN} \tag{11}$$

$$MCC = \frac{(TP + TN) - (FP + FN)}{\sqrt{(TN + FN) \times (TN + FP) \times (TP + FP) \times (TP + FN)}} \tag{12}$$

$$F_1Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{13}$$

$$precision = \frac{TP}{TP + FP} \tag{14}$$

where true positive (TP) indicates the drug pairs of interactions properly predicted by the model, true negative (TN) indicates the number of correctly classified non-interacting drug pairs, false positive (FP) indicates the number of misclassified non-interacting drug pairs, and false negative (FN) indicates the drug pairs with interactions misclassified by the model. This paper also uses both AUC (Area Under ROC Curve) [43] and AUPR (Precision Versus Recall Curve) [44] metrics to measure and evaluate the model. ROC curve presents the True positive rate (TPR) on the vertical axis and the False positive rate (FPR) on the horizontal axis. ROC and PR curves can objectively and effectively validate the experimental results.

The workflow of DBGRU-SE prediction method

The method proposed in this work to predict drug–drug interactions is called DBGRU-SE, and its workflow is shown in Figure 1. Windows Server 2012 R2 computer with Intel (R) Xeon (TM) CPU E5–2650 @ 2.30GHz, 32GB of RAM and Python 3.8 programming environment is used for the experiment. The source codes and datasets are available at https://github.com/YuBinLab-QUST/DBGRU-SE/.

The major process of DBGRU-SE is as follows.

Step 1. Data preparation. Download mol format files according to drug names in DrugBank and KEGG databases.

Step 2. Feature extraction. The drug information in the dataset is transformed into a numerical vector using four feature extraction methods, and the four features of the drug are fused to yield a vector with 1128 dimensions.

Step 3. Feature selection and data imbalance processing. The positive and negative sample datasets are constructed based on the drug–drug interaction matrix. Then, Group Lasso is used to reduce the dimensionality of the feature vector, and SMOTE-ENN is applied to handle positive and negative sample imbalance, which played an important role in the improvement of the model prediction accuracy.

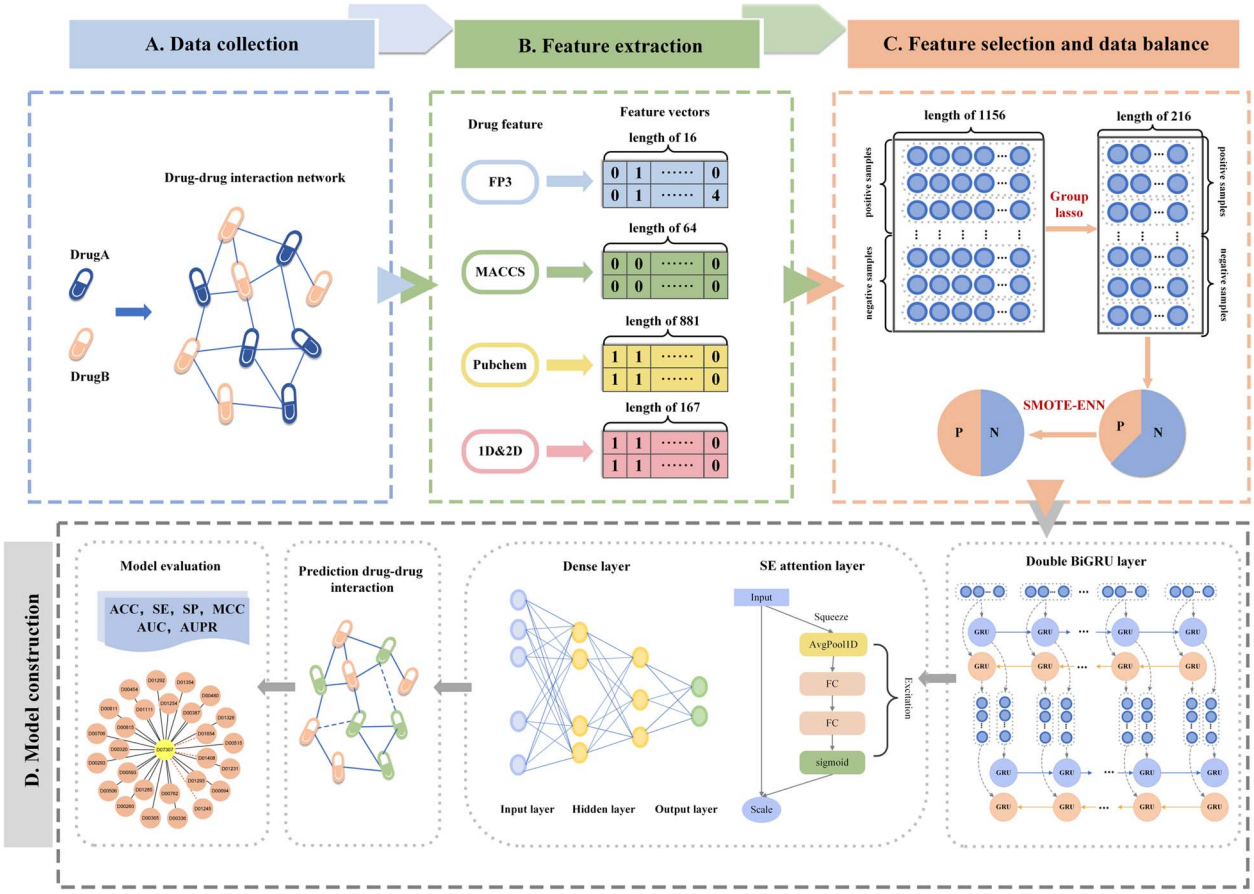

Figure 1. The overall framework flow chart of DBGRU-SE.

Step 4. Classification algorithm. The classifier combines double BiGRU and SE attention mechanism, and finally input to the three-layer dense layer to complete the prediction of DDIs.

Step 5. Model evaluation. This paper uses two datasets for prediction based on five-fold cross-validation. AUC, Precision, ACC, SE, SP and MCC values, as well as ROC and PR curves are used to assess the model's prediction performance. The effectiveness and performance of the model is validated on an independent test set.

RESULTS AND DISCUSSION

Comparison results of feature extraction methods

Considering the different feature information of drugs, this paper firstly uses FP3 fingerprints, MACCS fingerprints, Pubchem fingerprints, and 1D and 2D molecular descriptors for feature extraction to obtain 16-dimensional, 64-dimensional, 881-dimensional, and 167-dimensional feature vectors, respectively. Then four feature information is connected end-to-end to obtain 1128-dimensional feature vector fusion. Four features and the fused features are fed into the classifier separately. Five-fold cross-validation is used to evaluate the model. The prediction results are shown in Supplementary Table S1. The detailed ROC curves and PR curves are presented in Supplementary Figure S1.

Selection of dimension reduction methods

Combining four features obtains a 2256-dimensional feature vector. Fusion feature enhances the prediction of the model to some extent, but it also generates redundant data. Choosing a suitable feature selection method is essential for the construction of the model. In order to select the appropriate dimensionality reduction method and retain the optimal feature vector, six dimension reduction methods are compared in this paper, which are Max-Relevance-Max-Distance (MRMD) [45], GINI Index (GINI) [46], Information gain (IG) [47], Random Projection (RP) [48], Extra-Trees (ET) [49] and Group Lasso. The penalty parameter of Group Lasso is 0.03, the characteristic numbers of MRMD, GINI, IG and RP are 200, and the final characteristic numbers of Group Lasso and ET are 216 and 864, respectively. Dimensionality reduction is used for the feature set of Dataset1, which is the input of classifier. The results of different dimension reduction methods are shown in Supplementary Table S2.

Analysis of imbalance data processing methods

Training set used in this paper contains 47 947 positive samples and 100 737 negative samples, and there is an obvious imbalance between positive and negative samples. To choose the appropriate data balancing method, we compare SMOTE-ENN with the following baseline methods.

(i) NearMiss [44], Random Under-Sampling (RUS) [50] and Edited Nearest Neighbors (ENN) [51] are the under-sampling technique. Among them, the version of NearMiss is set to 3, and the random seed number of RUS and ENN are 42, and the rest parameters use the default parameters of the algorithm.

(ii) One-side Selection (OSS) [52] and Adaptive Synthetic Sampling (ADASYN) [53] are the over-sampling technique. The random seed number of OSS and ADASYN are 42, and the rest parameters use the default parameters of the algorithm.

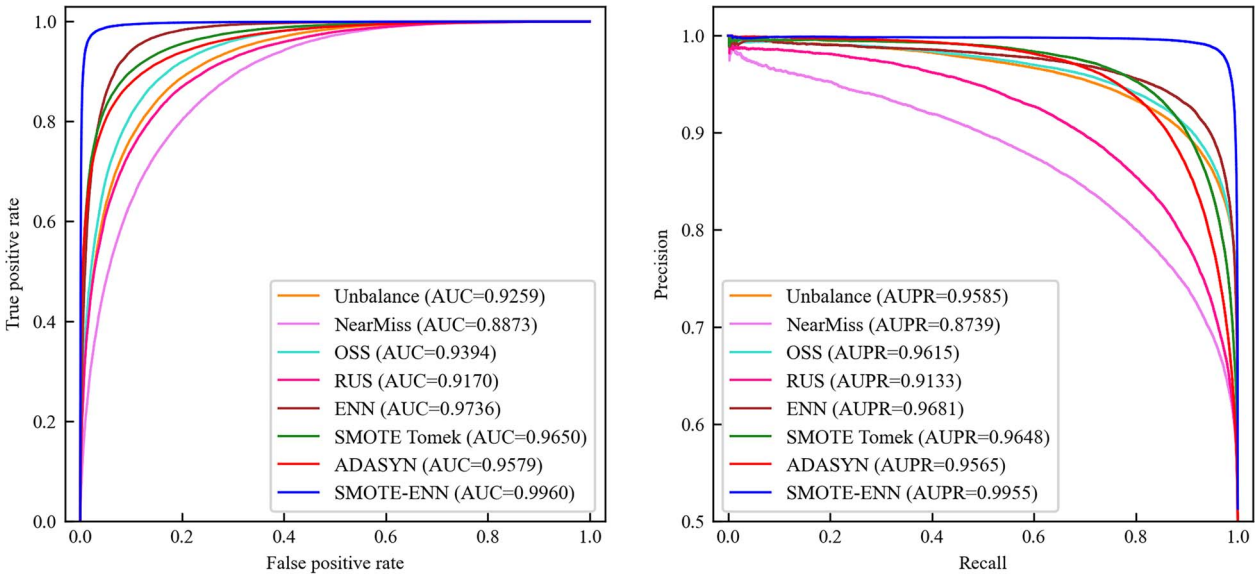

Figure 2. The ROC and PR curves of different data balance methods and unbalanced data. (A) ROC curves of different data balance methods and unbalanced data. (B) PR curves of different data balance methods and unbalanced data.

(iii) SMOTE Tomek [54] is an integrated sampling technique. The random seed number of SMOTE Tomek is 42, and the rest parameters use the default parameters of the algorithm.

The optimal feature subsets obtained by Group Lasso are processed by six data balancing methods, and input into the classifier for prediction. To verify the effectiveness of SMOTE-ENN method, we also compare the prediction results of the unbalanced data with those of the balanced treatment. The performance evaluation of all methods and the baseline are shown in Supplementary Table S3. The ROC and PR curves are as Figure 2.

As shown in Figure 2, The AUC and AUPR values of SMOTE-ENN are 99.60 and 99.55%, respectively, which are 2.24–10.87% and 2.74–12.16% higher than other methods. SMOTE-ENN first uses SMOTE to process the data imbalance, and then removes most of the k-nearest neighbor samples that are not the same as their own category according to ENN to achieve the effect of data cleaning. Therefore, this paper chooses SMOTE-ENN to balance the unbalanced dataset.

Selection of classifiers

In order to build an efficient drug–drug interaction prediction model, choosing an appropriate classifier is one of the most critical steps. This paper selects K-nearest Neighbor (KNN) [40], Naive Bayesian (NB) [55], Logistic Regression (LR) [56], Random Forest (RF) [57], Light Gradient Boosting Machine (LightGBM) [58], Multi-layer Perceptron (MLP) [59], CNN [60], DNN [45], Transformer [61] and DBGRU-SE to predict DDIs. Among them, the number of nearest neighbor points of KNN is 5. LR and NB algorithms use default parameters. RF algorithm sets the decision tree with a number of 500 and the depth of the base decision tree to 10. LightGBM sets the learning rate of 0.01 and the maximum number of iterations to 500. MLP sets the number of nodes in the hidden layer to 25 and 1, respectively. DNN contains four hidden layers, and the activation function sets ReLU. CNN consists of convolution layer, pooling layer and fully connected layer. Transformer is built as a layer of keras to fully exploit the information contained in the sub-structures of the feature vectors, generating potential feature vectors. The prediction comparison results of different classifiers are shown in Table 2. The ROC and PR curves of all classifiers are as Supplementary Figure S2.

From Table 2, DBGRU-SE achieves better performance. The ACC value of DBGRU-SE is 97.51%, which is 1.31–27.68% higher than KNN, NB, LR, RF, LightGBM, MLP, CNN, DNN and Transformer. The MCC value of DBGRU-SE reaches 0.9503, which is higher than KNN, NB, LR, RF, LightGBM, MLP, CNN, DNN and Transformer by 0.0262–0.5342. The prediction accuracy of NB algorithm is the lowest, and the ACC value is 69.83%. All other metrics of DBGRU-SE achieved significant improvements in comparison with other classification algorithms.

To further validate the effectiveness of DBGRU-SE, this paper chooses the bilateral *t*-test [62]. *t*-test can be utilized to analyze the differences between the metrics of each classifier. The p-values of classifiers KNN, NB, LR, RF, LightGBM, MLP, CNN, DNN and Transformer for ACC, MCC and AUC are as Supplementary Table S4.

Comparison with other methods

To verify the effectiveness of DBGRU-SE, this paper is tested on two datasets. On Dataset1, it is compared with five other prediction methods, including DPDDI, Weighted average ensemble method (WAM), Classifier ensemble method (CEM), NDD and ISCMF. Feng et al. [9] proposed DNNs based on deep neural networks and graph convolutional networks (GCNs) to extract the structural features of drugs based on the characteristics of DDI networks. Zhang et al. [20] integrated different models and suitable ensemble rules by using a flexible framework and proposed models for the weighted average ensemble method and classifier ensemble method. Rohani et al. [63] use integrated similarity-constrained matrix factorization (ISCMF) to predict DDIs. All models are evaluated by five-fold cross-validation. Vo et al. [64] transformed the chemical structure of drugs into useful predictive information and proposed an ensemble deep neural network EDN based on DNN, RF and XGBoost. On Dataset2, DBGRU-SE is compared with three prediction methods, including Naive-based similarity method [65], label propagation-based method [66] and semi-negative matrix decomposition-based method (DDINMF) [67]. The comparison detailed results of different methods are shown in Table 3.

As can be seen from Table 3, for Dataset1, DBGRU-SE achieves the best prediction results, and most indicators are better than

Table 2. Comparison of prediction results of nine different classifiers

Classifiers	ACC (%)	Precision (%)	SE (%)	SP (%)	MCC	F1 (%)
KNN	86.61	79.03	98.76	75.08	0.7565	87.79
NB	69.83	78.88	52.29	86.47	0.4161	62.52
LR	81.30	79.36	83.28	79.43	0.6271	81.27
RF	77.62	72.19	87.85	67.91	0.5683	79.22
LightGBM	84.59	81.32	88.58	80.80	0.6971	84.68
MLP	85.54	79.46	94.84	76.70	0.7251	86.47
CNN	95.65	95.25	95.84	95.47	0.9130	95.55
DNN	96.03	96.32	95.49	96.53	0.9205	95.90
Transformer	96.20	96.99	95.14	97.20	0.9241	96.06
DBGRU-SE	97.51	97.01	97.91	97.14	0.9503	97.46

Table 3. Performance comparison of different methods on Dataset1 and Dataset2

Datasets	Method	AUC	AUPR	Precision	ACC	F1
Dataset1	DPDDI	0.956	0.907	0.754	0.940	0.840
	WAM	0.951	0.795	0.775	0.955	0.712
	CEM	0.957	0.807	0.785	0.955	0.723
	DeepDDI	0.996	0.890	0.728	0.837	0.685
	EDN	N/A	N/A	0.838	0.938	N/A
	NDD	0.954	0.922	0.833	N/A	0.835
	ISCMF	0.899	0.864	0.988	0.851	0.885
	DBGRU-SE	0.996	0.996	0.970	0.975	0.975
Dataset2	Naive Similarity	0.779	0.342	N/A	N/A	N/A
	Label Propagation	0.776	0.327	N/A	N/A	N/A
	DDINMF	0.872	0.605	N/A	N/A	N/A
	DBGRU-SE	0.988	0.993	0.934	0.950	0.935

other methods. Specifically, the Precision index of DBGRU-SE was slightly lower than that of ISCMF. However, the AUPR, F1 and ACC values of DBGRU-SE are higher than other methods by 0.074–0.201, 0.124–0.29 and 0.09–0.263, respectively. In Dataset2, the AUPR and AUC values of DBGRU-SE have higher accuracy compared with other methods of predicting DDIs. As the results show, DBGRU-SE obtains better performance than other methods. Taken together, DBGRU-SE shows good prediction performance in both datasets.

Predicting drug–drug interaction networks

Drug–drug interaction network research is of critical importance for drug development and biomedical health. In this study, we downloaded drug information from DrugBank and KEGG drug databases to construct a new DDIs network, which includes 3 and 108 drugs interacting with them. A total of 105 new interactions were generated. Firstly, feature information of the drug is extracted from new DDI network. Secondly, Group lasso is applied to remove the redundant information to obtain the best feature subset. Finally, Dataset1 and new DDI network are input to DBGRU-SE as training and test sets to predict drug–drug interactions. The specific results are shown in Supplementary Figure S3.

CONCLUSION

This paper proposes a new method DBGRU-SE to predict DDIs. Firstly, we convert the character representation of drugs into digital signals by means of intelligent computing. FP3 fingerprints, MACCS fingerprints, Pubchem fingerprints, and 1D and 2D molecular descriptors are used for the feature extraction of drug molecules. Secondly, Group Lasso is used to process the fused high-dimensional data. The algorithm enables variable selection and removal of irrelevant features while retaining the predictive power of the model. Then, SMOTE-ENN is applied to solve the amount difference between positive and negative samples. Finally, we input the best subset of features into DBGRU-SE to predict. This algorithm not only makes use of the BiGRU network, enabling forward and backward propagation of data and retaining relevant features. It also uses the SE attention mechanism to focus on the most important features, further improving the efficiency of recognition. Compared with other DDIs prediction methods, DBSE-DDI produces better prediction results. Meanwhile, DBGRU-SE successfully predicted the drug–drug interaction network. Therefore, DBGRU-SE proposed in this paper is an effective method to predict DDI. DBGRU-SE can provide a scientific basis for clinical medical decision-making and drug development. Although DBGRU-SE has improved the accuracy of predicting DDI to a certain extent, there is still room for improvement. Firstly, we will continue to collate and build large-scale and high-quality drug–drug interaction datasets to meet the constantly updated and changing high-quality information. Secondly, DBGRU-SE uses end-to-end multidimensional feature fusion, which is not conducive to the optimal allocation of feature proportions. In the next work, we will consider introducing statistical learning optimization algorithms to assign different weights to different single features for optimal combinations. Finally, the interpretability of theoretical models is crucial in the life sciences. In future work, we will consider constructing an interpretable model for the prediction of DDIs.

Key Points

- We propose a new predictor, DBGRU-SE, to predict drug–drug interactions.

- This paper fuses four feature extraction methods. Group Lasso algorithm is used to remove irrelevant information. SMOTE-ENN algorithm is applied to imbalance the data. We combine the BiGRU and SE attention mechanisms for the prediction of drug–drug interactions.
- The results of this paper on two datasets show that DBGRU-SE achieves good predictive performance for predicting drug–drug interactions.

SUPPLEMENTARY DATA

Supplementary data are available online at http://bib.oxfordjournals.org/.

FUNDING

National Natural Science Foundation of China (62172248); Natural Science Foundation of Shandong Province of China (ZR2021MF098).

REFERENCES

1. Qiu Y, Zhang Y, Deng Y, *et al.* A comprehensive review of computational methods for drug-drug interaction detection. *IEEE/ACM Trans Comput Biol Bioinform* 2022;**19**(4):1968–85.
2. Güvenç Paltun B, Mamitsuka H, Kaski S. Improving drug response prediction by integrating multiple data sources: matrix factorization, kernel and network-based approaches. *Brief Bioinform* 2021;**22**:346–59.
3. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* 2018;**34**(13):i457–66.
4. Wang J, Liu X, Shen S, *et al.* DeepDDS: deep graph neural network with attention mechanism to predict synergistic drug combinations. *Brief Bioinform* 2022;**23**(1):bbab390.
5. Wishart DS, Knox C, Guo AC, *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* 2006;**34**(1):D668–72.
6. Hoffman KB, Dimbil M, Erdman CB, *et al.* The weber effect and the United States Food and Drug Administration's adverse event reporting system (FAERS): analysis of sixty-two drugs approved from 2006 to 2010. *Drug Saf* 2014;**37**(4):283–94.
7. Beutler B, Cerami A. Cachectin and tumour necrosis factor as two sides of the same biological coin. *Nature* 1986;**320**:584–8.
8. Chen J, Zhang L. A survey and systematic assessment of computational methods for drug response prediction. *Brief Bioinform* 2021;**22**:232–46.
9. Feng YH, Zhang SW, Shi JY. DPDDI: a deep predictor for drug-drug interactions. *BMC Bioinf* 2020;**21**(1):419.
10. Nyamabo AK, Yu H, Shi JY. SSI-DDI: substructure-substructure interactions for drug–drug interaction prediction. *Brief Bioinform* 2021;**22**(6):bbab133.
11. Lin S, Wang Y, Zhang L, *et al.* MDF-SA-DDI: predicting drug-drug interaction events based on multi-source drug fusion, multi-source feature fusion and transformer self-attention mechanism. *Brief Bioinform* 2022;**23**(1):bbab421.
12. Ma M, Lei X. A dual graph neural network for drug-drug interactions prediction based on molecular structure and interactions. *PLoS Comput Biol* 2023;**19**(1):e1010812.
13. Feng YH, Zhang SW, Feng YY, *et al.* A social theory-enhanced graph representation learning framework for multitask prediction of drug-drug interactions. *Brief Bioinform* 2023;**24**(1):bbac602.
14. He H, Chen G, Yu-Chian CC. 3DGT-DDI: 3D graph and text based neural network for drug-drug interaction prediction. *Brief Bioinform* 2022;**23**(3):bbac134.
15. Feng YY, Yu H, Feng YH, *et al.* Directed graph attention networks for predicting asymmetric drug-drug interactions. *Brief Bioinform* 2022;**23**(3):bbac151.
16. Ryu JY, Kim HU, Lee SY. Deep learning improves prediction of drug-drug and drug-food interactions. *Proc Natl Acad Sci U S A* 2018;**115**:E4304–11.
17. Guo L, Lei X, Chen M, *et al.* MSResG: using GAE and residual GCN to predict drug-drug interactions based on multi-source drug features. *Interdiscip Sci* 2023;**15**:171–88.
18. Su X, Hu L, You Z, *et al.* Attention-based knowledge graph representation learning for predicting drug-drug interactions. *Brief Bioinform* 2022;**23**(3):bbac140.
19. Hong Y, Luo P, Jin S, *et al.* LaGAT: link-aware graph attention network for drug-drug interaction prediction. *Bioinformatics* 2022;**38**:5406–12.
20. Zhang W, Chen Y, Liu F, *et al.* Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. *BMC Bioinf* 2017;**18**(1):1–12.
21. Law V, Knox C, Djoumbou Y, *et al.* DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res* 2014;**42**:D1091–7.
22. Shi JY, Huang H, Li JX, *et al.* TMFUF: a triple matrix factorization-based unified framework for predicting comprehensive drug-drug interactions of new drugs. *BMC Bioinf* 2018;**19**(14):27–37.
23. Tatonetti NP, Ye PP, Daneshjou R, *et al.* Data-driven prediction of drug effects and interactions. *Sci Transl Med* 2012;**4**(125):125ra131.
24. Lengauer T, Lemmen C, Rarey M, *et al.* Novel technologies for virtual screening. *Drug Discov Today Technol* 2004;**9**(1):27–34.
25. An X, Chen X, Yi D, *et al.* Representation of molecules for drug response prediction. *Brief Bioinform* 2022;**23**(1):bbab393.
26. Chen IJ, Hubbard RE. Lessons for fragment library design: analysis of output from multiple screening campaigns. *J Comput Aided Mol Des* 2009;**23**(8):603–20.
27. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol* 1996;**58**(1):267–88.
28. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *J R Stat Soc Series B Stat Methodol* 2006;**68**(1):49–67.
29. Chawla NV, Bowyer KW, Hall LO, *et al.* SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 2002;**16**:321–57.
30. Wang X, Yu B, Ma A, *et al.* Protein-protein interaction sites prediction by ensemble random forests with synthetic minority oversampling technique. *Bioinformatics* 2019;**35**(14):2395–402.
31. Yu B, Qiu W, Chen C, *et al.* SubMito-XGBoost: predicting protein submitochondrial localization by fusing multiple feature information and eXtreme gradient boosting. *Bioinformatics* 2020;**36**(4):1074–81.
32. He H. ENN: extended nearest neighbor method for pattern recognition. *IEEE Comput Intell Mag* 2015;**10**(3):52–60.
33. Muntasir NM, Faisal F, Jahan RI, *et al.* A comprehensive investigation of the performances of different machine learning classifiers with SMOTE-ENN oversampling technique and hyperparameter optimization for imbalanced heart failure dataset. *Sci Program* 2022;**2022**:1–17.
34. Lamari M, Azizi N, Hammami NE, *et al.* SMOTE–ENN-based data sampling and improved dynamic ensemble selection for imbalanced medical data classification. *Advances on Smart and Soft Computing* 2021;**1188**:37–49.

35. Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D* 2020;**404**:132306.
36. Hochreiter S, Schmidhuber J. LSTM can solve hard long time lag problems. *Adv Neural Inf Process Syst* 1996;**9**:473–9.
37. Cho K, Van MB, Gulcehre C, *et al.* Learning phrase representations using RNN encoder-decoder for statistical machine translation. *Proc Conf Empir Methods Nat Lang Process* 2014;1724–34.
38. Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. *Adv Neural Inf Process Syst* 2017;**30**:6000–10.
39. Hu J, Shen L, Albanie S, *et al.* Squeeze-and-excitation networks. *IEEE Trans Pattern Anal Mach Intell* 2019;**42**(8):2011–23.
40. Zhang Q, Li S, Zhang Q, *et al.* MpsLDA-ProSVM: predicting multi-label protein subcellular localization by wMLDAe dimensionality reduction and ProSVM classifier. *Chemometr Intell Lab Syst* 2021;**208**:104216.
41. Zhang Q, Liu P, Wang X, *et al.* StackPDB: predicting DNA-binding proteins based on XGB-RFE feature optimization and stacked ensemble classifier. *Appl Soft Comput* 2021;**99**:106921.
42. Zhang Q, Zhang Y, Li S, *et al.* Accurate prediction of multi-label protein subcellular localization through multi-view feature learning with RBRL classifier. *Brief Bioinform* 2021;**22**(5): bbab012.
43. Yu B, Wang X, Zhang Y, *et al.* RPI-MDLStack: predicting RNA–protein interactions through deep learning with stacking strategy and LASSO. *Appl Soft Comput* 2022;**120**:108676.
44. Wang M, Song L, Zhang Y, *et al.* Malsite-deep: prediction of protein malonylation sites through deep learning and multi-information fusion based on NearMiss-2 strategy. *Knowledge-Based Syst* 2022;**240**:108191.
45. Chen C, Shi H, Jiang Z, *et al.* DNN-DTIs: improved drug-target interactions prediction using XGBoost feature selection and deep neural network. *Comput Biol Med* 2021;**136**:104676.
46. Injadat MN, Moubayed A, Nassif AB, *et al.* Systematic ensemble model selection approach for educational data mining. *Knowledge-Based Syst* 2020;**200**:105992.
47. Shang C, Li M, Feng S, *et al.* Feature selection via maximizing global information gain for text classification. *Knowledge-Based Syst* 2013;**54**:298–309.
48. Wang Z, Jie W, Chen S, *et al.* Random projection ensemble learning with multiple empirical kernels. *Knowledge-Based Syst* 2013;**37**:388–93.
49. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn* 2006;**63**(1):3–42.
50. Zhang Y, Lam S, Yu T, *et al.* Integration of an imbalance framework with novel high-generalizable classifiers for radiomics-based distant metastases prediction of advanced nasopharyngeal carcinoma. *Knowledge-Based Syst* 2022;**235**:107649.
51. Zhang HH, Yang L, Liu Y, *et al.* Classification of Parkinson's disease utilizing multi-edit nearest-neighbor and ensemble learning algorithms with speech samples. *Biomed Eng Online* 2016;**15**:122–44.
52. Kubat M, Matwin S. Addressing the curse of imbalanced training sets: one-sided selection. *Icml* 1997;**97**(1):179.
53. Zhang T, Li Y, Wang X. Gaussian prior based adaptive synthetic sampling with non-linear sample space for imbalanced learning. *Knowledge-Based Syst* 2020;**191**:105231.
54. Batista GE, Bazzan ALC, Monard MC. Balancing training data for automated annotation of keywords: a case study. *WOB* 2003;**3**: 10–8.
55. Alizadeh SH, Hediehloo A, Harzevili NS. Multi independent latent component extension of naive Bayes classifier. *Knowledge-Based Syst* 2021;**213**:106646.
56. Yu B, Chen C, Zhou H, *et al.* GTB-PPI: predict protein–protein interactions based on L1-regularized logistic regression and gradient tree boosting. *Genomics, Proteomics Bioinf* 2020;**18**(5):582–92.
57. Shi H, Liu S, Chen J, *et al.* Predicting drug-target interactions using lasso with random forest based on evolutionary information and chemical structure. *Genomics* 2019;**111**(6):1839–52.
58. Liu Y, Yu Z, Chen C, *et al.* Prediction of protein crotonylation sites through LightGBM classifier based on SMOTE and elastic net. *Anal Biochem* 2020;**609**:113903.
59. Riedmiller M, Lernen A. Machine learning lab special lecture. *University of Freiburg* 2014;7–24.
60. Yu B, Zhang Y, Wang X, *et al.* Identification of DNA modification sites based on elastic net and bidirectional gated recurrent unit with convolutional neural network. *Biomed Signal Process Control* 2022;**75**:103566.
61. Zaikis D, Vlahavas I. TP-DDI: transformer-based pipeline for the extraction of drug-drug interactions. *Artif Intell Med* 2021; **119**:102153.
62. Zhang Y, Jiang Z, Chen C, *et al.* DeepStack-DTIs: predicting drug-target interactions using LightGBM feature selection and deep-stacked ensemble classifier. *Interdiscip Sci* 2022;**14**(2):311–30.
63. Rohani N, Eslahchi C, Katanforoush A. ISCMF: integrated similarity-constrained matrix factorization for drug-drug interaction prediction. *Netw Model Anal Health Inform Bioinform* 2020; **9**(1):1–8.
64. Vo TH, Nguyen NTK, Le NQK. Improved prediction of drug-drug interactions using ensemble deep neural networks. *Med Drug Discov* 2023;**17**:100149.
65. Vilar S, Uriarte E, Santana L, *et al.* Similarity-based modeling in large-scale prediction of drug-drug interactions. *Nat Protoc* 2014;**9**(9):2147–63.
66. Zhang P, Wang F, Hu J, *et al.* Label propagation prediction of drug-drug interactions based on clinical side effects. *Sci Rep* 2015;**5**(1): 1–10.
67. Yu H, Mao KT, Shi JY, *et al.* Predicting and understanding comprehensive drug-drug interactions via semi-nonnegative matrix factorization. *BMC Syst Biol* 2018;**12**(1):101–10.