



Article

Integration of Neighbor Topologies Based on Meta-Paths and Node Attributes for Predicting Drug-Related Diseases

Ping Xuan ¹, Zixuan Lu ¹, Tiangang Zhang ^{2,*}, Yong Liu ^{1,*} and Toshiya Nakaguchi ³

¹ School of Computer Science and Technology, Heilongjiang University, Harbin 150080, China; xuanping@hlju.edu.cn (P.X.); zxuanlu2270@163.com (Z.L.)

² School of Mathematical Science, Heilongjiang University, Harbin 150080, China

³ Center for Frontier Medical Engineering, Chiba University, Chiba 2638522, Japan; nakaguchi.chibau@gmail.com

* Corresponding authors: zhang@hlju.edu.cn (T.Z.); liuyong123456@hlju.edu.cn (Y.L.)

Abstract: Identifying new disease indications for existing drugs can help facilitate drug development and reduce development cost. The previous drug–disease association prediction methods focused on data about drugs and diseases from multiple sources. However, they did not deeply integrate the neighbor topological information of drug and disease nodes from various meta-path perspectives. We propose a prediction method called NAPred to encode and integrate meta-path-level neighbor topologies, multiple kinds of drug attributes, and drug-related and disease-related similarities and associations. The multiple kinds of similarities between drugs reflect the degrees of similarity between two drugs from different perspectives. Therefore, we constructed three drug–disease heterogeneous networks according to these drug similarities, respectively. A learning framework based on fully connected neural networks and a convolutional neural network with an attention mechanism is proposed to learn information of the neighbor nodes of a pair of drug and disease nodes. The multiple neighbor sets composed of different kinds of nodes were formed respectively based on meta-paths with different semantics and different scales. We established the attention mechanisms at the neighbor-scale level and at the neighbor topology level to learn enhanced neighbor feature representations and enhanced neighbor topological representations. A convolutional-autoencoder-based module is proposed to encode the attributes of the drug–disease pair in three heterogeneous networks. Extensive experimental results indicated that NAPred outperformed several state-of-the-art methods for drug–disease association prediction, and the improved recall rates demonstrated that NAPred was able to retrieve more actual drug–disease associations from the top-ranked candidates. Case studies on five drugs further demonstrated the ability of NAPred to identify potential drug-related disease candidates.

Keywords: drug–disease association prediction; neighbor topology learning based on meta-paths; pairwise node attribute encoding; multiple drug–disease heterogeneous networks; fully connected neural networks and autoencoder based on CNN



Citation: Xuan, P.; Lu, Z.; Zhang, T.; Liu, Y.; Nakaguchi, T. Integration of Neighbor Topologies Based on Meta-Paths and Node Attributes for Predicting Drug-Related Diseases.

Int. J. Mol. Sci. **2022**, *23*, 3870.

<https://doi.org/10.3390/ijms23073870>

Academic Editors: Dusanka Janezic and Lorentz Jäntschi

Received: 30 January 2022

Accepted: 15 March 2022

Published: 31 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The process of producing a new medicine is typically lengthy, expensive, and fraught with failure; it may require more than 10 y and cost between USD 0.8 billion and USD 1.5 billion on average [1–5]. Therefore, a method to reduce the time and funding costs for the development of new medicines must be identified. That approved drugs are subject to clinical trials endows them with a favorable safety profile. In contrast to developing a medicine from scratch, using indications for current drugs (drug repositioning) [6] can effectively reduce research and development costs and accelerate drug development [7–9].

Drug candidates can be further screened for wet laboratory validation using computational predictions of the relationship between licensed drugs and diseases [10,11]. Several approaches for predicting drug-related diseases that have been reported can be classified

into two categories. The first category of methods predicts the disease indications for drugs based on the integration of multiple kinds of information about the drugs and diseases. A couple of methods integrate the known drug–disease associations, the drug similarities, and the disease similarities [12,13]. They estimate the association possibilities between drugs and diseases by utilizing a logistic regression classifier and matrix decomposition with a similarity constraint. Wang et al. employed kernel functions to incorporate drug and disease similarities and applied the support vector machine approach to forecast drug–disease correlations [14]. Liang et al. applied sparse subspace learning and graph Laplacian regularization to combine multiple types of drug characteristics to predict drug indications [15]. To infer drug–disease associations, relevant data from drugs and diseases are utilized or combined in these strategies. However, the above-mentioned approaches cannot consider topological information in a network to demonstrate the potential use of a specific drug.

The second method primarily considers prediction based on the topology of the network. For example, heterogeneous network models based on diseases, drugs, and targets are used to infer drug candidates using iterative algorithms [16]. In several methods, random walk algorithms are employed to predict possible drug–disease associations; in fact, they have been employed in networks such as drug similarity, disease similarity, and integrated drug–disease heterogeneity networks [17–21]. However, because these methods do not consider the attribute information of drug and disease network nodes, they cannot learn the deep feature representation of nodes. Furthermore, these shallow-model-based approaches cannot extract potentially complicated relationships between drug and disease nodes.

Deep learning technologies have been widely utilized for the prediction of miRNA–disease associations [22] and disease-related lncRNAs [23,24]. Owing to the development of deep learning, the indications of drug candidates are identified more accurately in recent approaches by integrating multiple sources of drug- and disease-relevant information. For the prediction of drug-related diseases, models employing graph convolutional and fully connected autoencoders with attention mechanisms are used [25]. Xuan et al. [26] proposed a prediction model comprising a convolutional neural network (CNN) and a bi-directional long short-term memory (BiLSTM) network. Jiang et al. devised a module for forecasting drug–disease correlations by employing Gaussian interaction profile kernels and autoencoders [27]. Deep relationships between drugs and diseases can be extracted more easily using deep learning models. At the node pair level, however, the present deep learning approaches cannot combine and incorporate the drug–disease neighbor topology and attribute information. In addition, when capturing the neighbor topology information in three heterogeneous networks, the multi-scale meta-paths to obtain the set of neighbor nodes is important auxiliary information.

Herein, we propose and develop NAPred, a predictive model for capturing, encoding, and learning the neighbor topology and attribute representation of node pairs from diverse heterogeneous networks. The primary contributions of our proposed model are as follows:

- Three drug–disease heterogeneous networks were constructed, each with different aspects of drug similarities, to facilitate the acquisition of topological information regarding drug and disease nodes from different perspectives. To construct sets of different types of neighbors of the nodes, multi-scale meta-path sets of drug or disease nodes were established;
- We present an approach based on fully connected and convolutional neural networks with attention mechanisms for learning topological information regarding the same type of neighbors for drug and disease nodes. Multiple-neighbor feature representations extracted from drug and disease nodes were adaptively combined via a neighbor-scale-level attention mechanism;
- We developed a neighbor-topology-level attention mechanism to distinguish the contributions and then obtain the neighbor topological representations of the nodes;

this is because different types of neighbor topological features contribute differently to drug–disease association prediction;

- The attribute information of the node pairs was extracted from the three heterogeneous networks using the proposed embedding mechanism and encoded using a convolutional autoencoder (CAE). The premise of this embedding mechanism is that drug–disease pairs are more likely to be associated with each other if they exhibit similarities or associations with more typical drugs or diseases.

2. Experimental Results and Discussion

2.1. Evaluation Metrics

The performances of all prediction models were analyzed and compared using five-fold cross-validation. Positive and negative samples were those with known and unknown drug–disease associations, respectively. We used 4/5 of the positive samples, as well as 4/5 of the random negative samples formed in the training set in each fold of the cross-validation. The remaining 1/5 positive samples, as well as all negative samples were tested. The prediction correlation scores of the test samples were generated and ranked; the higher the rank of the positive sample use cases, the better was their prediction performance.

Several evaluation metrics were used in this study, i.e., the true positive rate (TPR), false positive rate (FPR), receiver operating characteristic (ROC) curve, area under the ROC curve (AUC) [28], precision–recall (PR) curve, area under the PR (AUPR) curve [29], and recall at various top-*k*. The performances of all models in the cross-validation were compared based on the average AUC and AUPR.

The AUC is an accepted appraisal metric for comparing algorithms and probabilistic estimates [30]. The TPR and FPR at various thresholds yield the ROC curve. The sample was regarded as positive if the predicted association score of a drug–disease pair exceeds a threshold θ ; otherwise, it was considered negative. The fraction of correctly (incorrectly) detected positive (negative) samples among all the positive (negative) samples is denoted as the *TPR* (*FPR*).

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{TN + FP} \quad (1)$$

where *TP* (*FN*) represents the number of positive samples correctly (incorrectly) classified as positive (negative) and *TN* (*FP*) indicates the number of negative samples correctly (incorrectly) categorized as negative (positive) [31,32].

This was due to the uneven distribution of drug–disease candidates. The AUPR curve provides more information regarding the AUC for assessing the predictive performance [29]. *precision* and *recall* were determined as follows:

$$precision = \frac{TP}{TP + FP}, recall = \frac{TP}{TP + FN} \quad (2)$$

where *precision* indicates the rate of TP samples among those anticipated to be positive and *recall* expresses the rate of positive samples accurately recognized among the total positive samples. The AUC and AUPR curve were calculated using the mean cross-validation [33]. Each fold's mean AUC and AUPR curve must be calculated, and the final score is the average of the five results.

Considering that biologists typically choose the top-ranked candidates and confirm their predictions based on wet laboratory trials, determining the actual drug–disease connections is critical. Therefore, for the projected outcomes, the recall rates of the top-*k* candidate drug–disease pairs were evaluated. The more trustworthy the prediction performance, the higher is the recall of the top-*k*.

2.2. Comparison with Other Methods

NAPred is more effective compared with six cutting-edge drug–disease association forecasting models: GFPred [25], CBPred [26], SCMFDD [13], LRSSL [15], MBiRW [18], and

HGBI [16]. In the cross-validation, the other six methods were trained or tested using the same or similar datasets as the NAPred model. The best performance was achieved by each method when the optimal parameters were used. In particular, $lr = 0.001$ for GFPred; $lr = 0.001$ and $\lambda = 0.12$ for CBPred; $\mu = \lambda = 0.01$, $\gamma = 2$, and $k = 10$ for LRSSL; $\alpha = 0.3$, $c = -11$, $d = \log(9999)$, and $l = r = 2$ for MBiRW; $k = 45\%$, $\lambda = 1$, and $\lambda = 4$ for SCMFDD; $\alpha = 0.4$ for HGBI.

For each of the 763 drugs, we calculated the AUC and AUPR curve at each fold before calculating their five-fold mean. The final results were averaged across all AUCs (or AUPR curves) for the 763 drugs. As shown in Figure 1A, in the comparison of the 763 drugs, NAPred achieved the best mean AUC value among all the methods investigated (AUC = 0.978), outperforming GFPred by 3.3%, CBPred by 5.2%, SCMFDD by 25.5%, LRSSL by 14.7%, MBiRW by 15%, and HGBI by 27.6%. The second-best model GFPred successfully learned multiple attribute representations of nodes and fully extracted topological information from multiple heterogeneous networks. This suggests that constructing heterogeneous networks on the basis of multiple drug similarities and capturing topological information improved the prediction accuracy. CBPred, LRSSL, and MBiRW extract topology information from heterogeneous networks for drug repositioning, where CBPred considers the path information between pairs of diseases, whereas MBiRW disregards the properties of the nodes. Hence, CBPred performed better, whereas MBiRW performed worse than LRSSL. SCMFDD is a matrix-decomposition-based model. The dimensionality reduction process may cause the loss of low-frequency valid information. Therefore, SCMFDD performed worse, but better than HGBI; additionally, it did not exploit the multiple similarities of the drugs. In conclusion, our NAPred achieved the best results owing to the comprehensive learning of the neighborhood topology, as well as the property information of the drug–disease pairs.

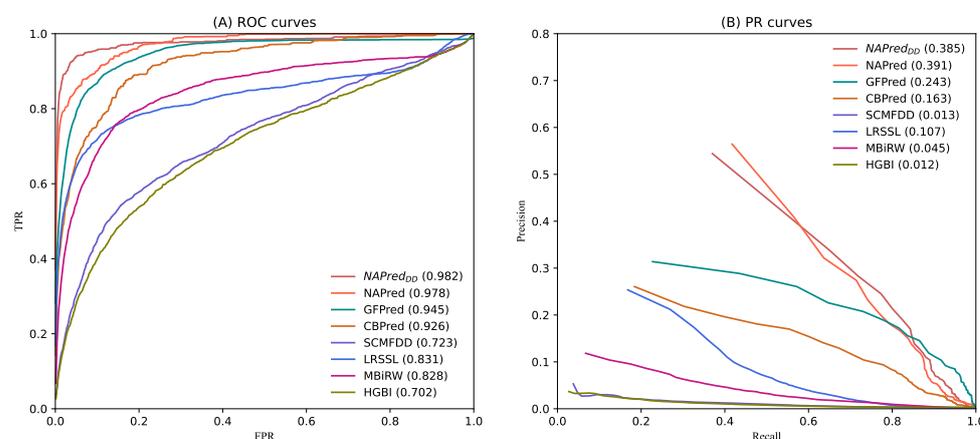


Figure 1. ROC and PR curves of all the methods of drug–disease association.

As shown in Figure 1B, our method NAPred performed better than GFPred, CBPred, LRSSL, MBiRW, SCMFDD, and HGBI by 14.8%, 22.8%, 28.4%, 34.6%, 37.8%, and 37.9%, respectively, based on the AUPR curves of 763 drugs.

In addition, to validate the robustness of our model under multiple datasets, we used the CC dataset [34] to replace drug-related data and implement another instance of our method, $NAPred_{DD}$. We utilized the A (chemistry) data, B (targets) data, and C (networks) data of CC dataset to replace the original chemical substructure, protein structural domain, and gene ontology data of the drugs. In Figure 1, the AUC and AUPR of $NAPred_{DD}$ are still higher than those of the compared methods. The experimental results demonstrated the good robustness of our model.

To evaluate the impact of cross-validation folds on NAPred performance, we also performed an additional ten-fold cross-validation. The number of training samples in the ten-fold cross-validation was larger than that in the five-fold cross-validation. As shown in

Supplementary Table S1, the AUC and AUPR for the ten-fold cross validation were 0.8% and 1.3% higher than the five-fold cross validation. NAPred achieved better performance when the training data were increased.

The Wilcoxon test was used to evaluate the ability of the 763 drugs to predict the outcomes. NAPred performed much better than the other approaches in terms of the AUCs and AUPR curves when a 0.05 p -value threshold was used (Table 1).

Table 1. The statistical results of the paired Wilcoxon test on the AUCs over all the 763 drugs by comparing NAPred and all other five methods.

	GFPred	CBPred	SCMFDD	LRSSL	MBiRW	HGBI
p -value of AUC	5.27051×10^{-25}	1.83480×10^{-33}	5.49787×10^{-65}	5.31080×10^{-47}	2.89205×10^{-62}	1.74747×10^{-81}
p -value of AUCPR	3.42304×10^{-31}	4.72506×10^{-47}	1.81013×10^{-71}	8.63715×10^{-65}	4.68094×10^{-59}	4.85712×10^{-89}

Figure 2 shows the recall rates of drug candidates for various top- k values. More real drug–disease associations can be successfully identified using a higher recall rate. The average recall rate for the 763 drugs was 86.14%, 89.19%, 93.24%, 95.54%, and 97.33% for the top-30, -60, -90, -120, and -150, respectively. Among the top-30, -90, -150, and -210, GFPred indicated the second-highest recall rate, with 81.03%, 90.20%, 94.64%, and 97.12%, respectively. CBPred obtained recall rates of 68.63%, 82.41%, 90.69%, and 94.17% in the top-30, -90, -150, and -210, respectively, with a slightly lower performance than GFPred. LRSSL demonstrated a higher recall than MBiRW for the top-30, -60, and -90. The former model achieved 66.12%, 70.73%, and 74.90% recall rates, whereas the latter obtained recall rates of 57.65%, 65.30%, and 73.71%, respectively. The recall of SCMFDD was 32.97%, 51.18%, 59.75%, and 66.13% when k was 30, 90, 150, and 210, respectively. HGBI had a slightly lower recall rate than SCMFDD, i.e., 30.62%, 46.10%, 56.34%, and 63.98% for the top-30, -90, -150, and -210, respectively.

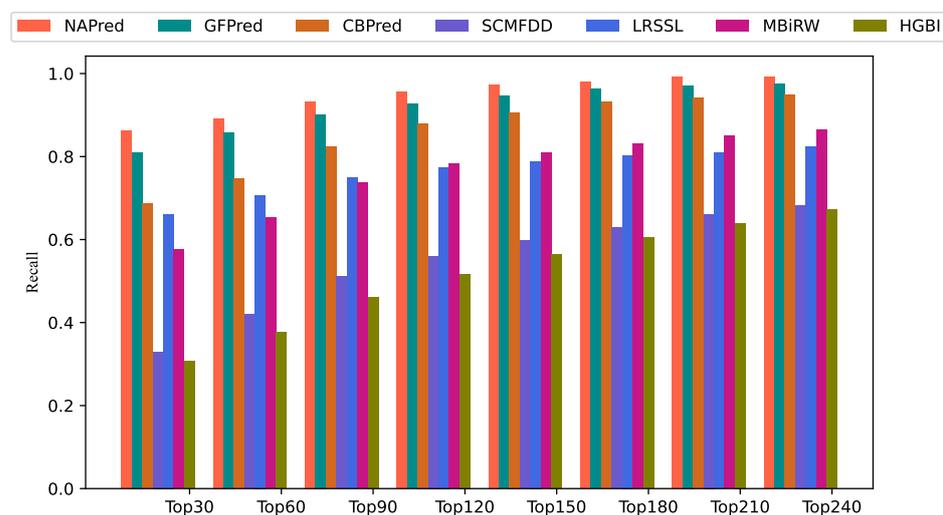


Figure 2. The average recalls of all the drugs under different top- k .

2.3. Case Studies of Five Drugs

Case studies of ampicillin, ceftriaxone, doxorubicin, erythromycin, and itraconazole were conducted to further illustrate the efficacy of NAPred in drug–disease association prediction. The association prediction scores for each drug candidate in the descending order, as well as the top-ten candidates for each of the five drugs are listed in Table 2.

The Comparative Toxicogenomics Database (CTD), which was painstakingly acquired and validated based on the literature, contains information regarding drugs and their

effects on human health [35]. DrugBank is a database containing drug-related targets, mechanisms of action, interactions, and integrated molecular information [36]. A total of 16 candidate diseases are covered by CTD, and 23 candidates are recorded in DrugBank. This indicates that the disease candidate was receiving effective treatment.

ClinicalTrials.gov, which is the world's largest searchable clinical trial database, contains data pertaining to clinical studies conducted worldwide; the National Library of Medicine in the United States contributes to its resources. As supporting material, we only used experimental records with a "completed" status. PubChem is a public database sponsored by the National Institutes of Health that includes information regarding chemicals and their biological activity, safety, and toxicity [37]. There were 23 candidate diseases supported by ClinicalTrials.gov, whereas PubChem approved 33 of the candidates. These records indicate that clinical trials established an association between the candidate disease and the relevant drug.

Besides manually validated drug–disease correlations, CTD additionally includes those derived from the literature with temporarily unverified associations. The inferred section of the CTD contains two candidates, which suggests a more plausible correlation between the diseases and their corresponding drugs. Among all 50 drug candidates, two candidates were labeled as "unconfirmed".

In addition, we conducted case studies on an additional five drugs (betamethasone, acetaminophen, etoposide, flurbiprofen, and verapamil) and list their top-ten candidate diseases in Supplementary Table S2. There were 42 candidate diseases recorded by CTD. There were 29 and 42 candidates covered by DrugBank and PubChem. ClinicalTrials contained 20 candidate diseases. This indicates that these candidates are more likely to be associated with the corresponding drugs. Only one candidate was labeled as "unconfirmed". All the above analysis indicated that NAPred had the ability to discover potential candidate drug–disease associations.

Table 2. The top-10 candidate diseases of 5 drugs.

Drug Name	Rank	Disease Name	Description	Rank	Disease Name	Description
Ampicillin	1	Staphylococcal Infections	CTD, PubChem	6	Staphylococcal Skin Infections	PubChem
	2	Pneumonia, Bacterial	ClinicalTrials	7	Streptococcal Infections	CTD, ClinicalTrials
	3	Urinary Tract Infections	CTD, DrugBank, PubChem	8	Osteomyelitis	PubChem, ClinicalTrials
	4	Wound Infection	PubChem, ClinicalTrials	9	Postoperative Complications	PubChem
	5	Proteus Infections	Inferred Candidate by 2 Literature Works	10	Bacterial Infections	CTD, DrugBank, ClinicalTrials
Ceftriaxone	1	Escherichia coli Infections	CTD, PubChem, ClinicalTrials	6	Salmonella Infections	DrugBank, PubChem, ClinicalTrials
	2	Urinary Tract Infections	DrugBank, PubChem, ClinicalTrials	7	Enterobacteriaceae Infections	PubChem, ClinicalTrials
	3	Haemophilus Infections	PubChem	8	Septicemia	DrugBank, PubChem, ClinicalTrials
	4	Gonorrhea	DrugBank, PubChem, ClinicalTrials	9	Endocarditis, Bacterial	DrugBank, ClinicalTrials
	5	Gram-Negative Bacterial Infections	Inferred Candidate by 1 Literature Work	10	Pseudomonas Infections	PubChem
Doxorubicin	1	Urinary Tract Infections	CTD, PubChem	6	Leukemia, Lymphoid	CTD, DrugBank, ClinicalTrials
	2	Leukemia, Myeloid, Acute	CTD, DrugBank, ClinicalTrials	7	Bronchitis	CTD
	3	Escherichia coli Infections	CTD	8	Sarcoma	CTD, DrugBank, ClinicalTrials
	4	Neoplasms	ClinicalTrials, PubChem	9	Gonorrhea	Unconfirmed
	5	Staphylococcal Infections	CTD, PubChem	10	Precursor Cell Lymphoblastic Leukemia-Lymphoma	CTD
Erythromycin	1	Gonorrhea	DrugBank, PubChem	6	Gram-Positive Bacterial Infections	PubChem
	2	Gram-Negative Bacterial Infections	PubChem	7	Staphylococcal Infections	CTD, DrugBank, PubChem
	3	Chancroid	DrugBank, PubChem	8	Pneumonia, Mycoplasma	Unconfirmed
	4	Bacterial Infections	DrugBank, PubChem	9	Neurosyphilis	PubChem
	5	Neisseriaceae Infections	DrugBank	10	Chlamydiaceae Infections	DrugBank, ClinicalTrials
Itraconazole	1	Candidiasis, Cutaneous	DrugBank, PubChem, ClinicalTrials	6	Tinea Capitis	DrugBank, PubChem
	2	Tinea Versicolor	DrugBank, PubChem, ClinicalTrials	7	Fungemia	DrugBank, PubChem, ClinicalTrials
	3	Tinea Pedis	DrugBank, PubChem	8	Skin Diseases, Infectious	PubChem, ClinicalTrials
	4	Leishmaniasis	CTD, PubChem, ClinicalTrials	9	AIDS-Related Opportunistic Infections	ClinicalTrials
	5	Chromoblastomycosis	DrugBank, PubChem	10	Candidiasis	CTD, DrugBank, PubChem

2.4. Prediction of Novel Drug-Related Diseases

Finally, we applied the trained NAPred to 763 drugs to predict candidate diseases. The top-30 drug-related candidate diseases selected by our model are listed in Supplementary Table S3. They can be used by biologists to facilitate further wet experiments for validation.

3. Materials and Methods

Figure 3 shows our proposed predictive model for drug-related disease candidates; the model comprises two branches. Three drug–disease heterogeneity networks were first established to correlate the similarities between drugs and diseases from different perspectives. For the first branch, we obtained the sets of neighbor nodes for drugs and diseases based on meta-paths of different scales. Neighbor-scale-level and neighbor-topology-level attention mechanisms are proposed for capturing drug and disease neighbor information, followed by encoding pairwise neighbor topology representations using convolutional neural networks. In the second branch, CAE was utilized to learn a pair of drug–disease attribute representations from the three drug–disease heterogeneous networks. The scores predicted from the two branches were weighted and summed to obtain the scores for the corresponding associations. A higher score signifies the higher possibility of an association.

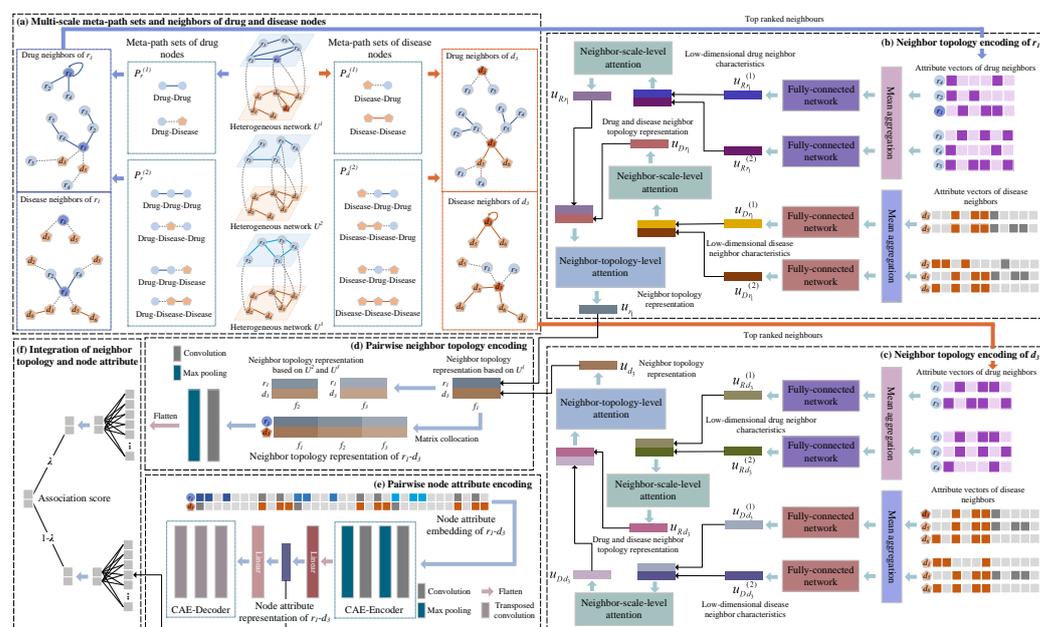


Figure 3. Framework of the proposed NAPred model. (a) Construct multi-scale meta-path sets and the sets composed of the same-type neighbor nodes. (b) Encode the attribute vectors of neighbor nodes of a drug. (c) Encode the attribute vectors of neighbor nodes of a disease. (d) Learn the neighbor topology of a drug–disease node pair. (e) Learn the attributes of the node pair. (f) Integrate multiple representations.

3.1. Dataset

Based on previous studies, we obtained drug–disease association data [15], chemical substructure data of drugs, protein structural domain data of target proteins, and gene ontology information of target proteins. Initially, data pertaining to drug–disease associations were obtained in the UMLS [38], which contains information regarding 763 drugs, 681 diseases, and 3051 known drug–disease associations. We extracted drug chemical substructure data from the PubChem database [39] and drug target protein structural domain data from the InterPro database [40]. The UniProt database was used to obtain gene ontology information regarding the target protein of the drug [41]. The numbers of drug chemical substructures, drug target protein structural domains, and drug target protein gene ontologies in our dataset were 623, 1426, and 4447, respectively.

3.2. Establishing Drug–Disease Heterogeneous Networks

3.2.1. Matrix of Drug Properties

Let the matrix T^c denote the case in which each drug contains a chemical substructure, and $T^c \in \mathbb{R}^{N_r \times N_c}$. N_r and N_c indicate the number of drugs and all relevant chemical substructures, respectively. A T_{ij}^c value of 1 implies that drug r_i contains the chemical substructure c_j , whereas a value of 0 implies otherwise. The vector of the chemical substructure attributes of r_i , which is obtained from the i -th row vector of T^c , is represented as T_i^c .

Let the matrix $T^p \in \mathbb{R}^{N_r \times N_p}$ denote the cases of protein structural domains discovered in the respective associated target proteins of N_r drugs; subsequently, N_p is the number of protein structural domains of all drug target proteins. T_{ij}^p is 1 for the target protein related to drug r_i containing the j -th protein structural, and 0 otherwise. The protein structural domain attribute vector of r_i is obtained from the i -th row of data in T^p .

The matrix $T^g \in \mathbb{R}^{N_r \times N_g}$ is used to indicate whether N_g gene ontology information is included in N_r drugs and their associated target proteins. A T_{ij}^g value of 1 implies that the target protein associated with drug r_i contains gene ontology g_j , whereas a value of 0 implies otherwise. The target protein gene ontology property vector of r_i is represented by the i -th row vector T_i^g .

3.2.2. Establishment of the Drug Network

For two drugs r_i and r_j , a higher number of identical chemical substructures between them signifies a higher level of similarity between them. The cosine similarity of their chemical substructures can be calculated using the strategy previously described by Liang et al. [15]; in fact, we used it as the first cosine similarity between r_i and r_j .

Similarly, based on the protein domains or protein-associated gene ontologies in the two drug-related target proteins, cosine similarity calculations can be applied to determine the second and third similarities of a drug.

We treated two drug nodes as having connected edges when the calculated drug similarity exceeded 0. The weights on the edges are expressed as the similarity between the two drugs (Figure 4). We used the matrices $R^c = [R_{ij}^c] \in \mathbb{R}^{N_r \times N_r}$, $R^p = [R_{ij}^p] \in \mathbb{R}^{N_r \times N_r}$, and $R^g = [R_{ij}^g] \in \mathbb{R}^{N_r \times N_r}$ to denote the drug networks obtained based on the similarity of the three drugs. For instance, based on the chemical substructure, R_{ij}^c represents the similarity between r_i and r_j .

3.3. Neighborhood Topology Encoding

3.3.1. Multi-Scale Meta-Path Sets

The meta-path [43] can be expressed as a path shaped as $G_1 \xrightarrow{R_1} G_2 \xrightarrow{R_2} \dots \xrightarrow{R_t} G_t$ (abbreviated as $G_1 G_2 \dots G_t$). The complex relationship of node types G_1 and G_t is described by $R = R_1 \circ R_2 \circ \dots \circ R_t$. Two nodes can be connected to each other via different meta-paths in a heterogeneous drug–disease network. Figure 1 shows the manner by which drugs r_1 and r_4 can be connected by meta-paths $r-r-r$ and $r-d-r$, with different meta-paths showing different semantics. For example, in $r_1-r_2-r_4(rrr)$, drugs r_1 and r_4 may be similar if both have functions similar to r_2 . In $r_1-d_5-r_4(rdr)$, an association is indicated between both drugs and d_5 , suggesting that r_1 may be similar to r_4 .

Based on the structural information from U^m , we can obtain the first-order meta-paths of drug nodes with $r-r$ and $r-d$ to form the set $P_r^{(1)} = \{rr, rd\}$ of the first-order meta-paths of the drug nodes. Similarly, the second-order meta-paths of the drug nodes include $r-r-r, r-d-r, r-r-d$, and $r-d-d$, which form set $P_r^{(2)} = \{rrr, rdr, rrd, rdd\}$ of the second-order meta-paths of the drug node. Finally, we obtain set $P_r^{(k)}(P_d^{(k)})$, $k = 1, 2, \dots, K$ of the multi-scale meta-paths of the drug (disease) nodes.

3.3.2. Neighbor Sets Based on Meta-Paths at Different Scales

For node $r_i(d_j)$ and the set of meta-paths $P_r^{(k)}(P_d^{(k)})$, we can capture the drug nodes or disease nodes connected to $r_i(d_j)$ based on meta-paths of different scales. This results in a set of drug neighbor nodes $N_{Rr_i}^{(k)}(N_{Rd_j}^{(k)})$ and the disease neighbor node set $N_{Dr_i}^{(k)}(N_{Dd_j}^{(k)})$ at different scales of $r_i(d_j)$, where the first-order neighbors of the node include itself.

For the drug (disease)-type neighbors of $r_i(d_j)$, we calculated the top- N_k neighbors that were the most similar to $r_i(d_j)$ based on their similarity to all other drugs (diseases). For the disease (drug)-type neighbors of $r_i(d_j)$, the disease (drug) nodes associated with $r_i(d_j)$ were ranked based on their occurrence frequency, and the top- N_k nodes of the ranking were retained as neighbors of $r_i(d_j)$.

As shown in Figure 3, for r_1 and the set of meta-paths $P_r^{(1)}$ and $P_r^{(2)}$, assuming $N_k = 3$, we can obtain the first-order drug neighbor nodes of r_1 based on $P_r^{(1)}$ via meta-paths $r-r$, retain the three top-ranked neighbors of r_1 , and obtain the set $N_{Rr_1}^{(1)} = \{r_1, r_2, r_4\}$. Similarly, r_1 captures and retains the top- N_k disease neighbors via meta-paths $r-r-d$ and $r-d-d$ in $P_r^{(2)}$, thereby forming its second-order disease neighbor set $N_{Dr_1}^{(2)} = \{d_2, d_5, d_6\}$.

3.3.3. Aggregation of Multi-Scale Neighbor Features

We propose a fully connected neural network with mean aggregation [44] to effectively combine the network topology in U^m with the characteristics of same-type nodes to learn the low-dimensional features of same-type neighbors at different scales. Because the learning frameworks of both drug and disease nodes are similar, we describe r_i and its drug (disease)-type neighbors as an example.

For the k th-order drug neighbor set $N_{Rr_i}^{(k)}$ of r_i , the attribute vector f_{r_n} of its neighbor node $r_n \in N_{Rr_i}^{(k)}$ can be obtained from the drug attribute matrix (T^c, T^p, T^g) corresponding to U^m . Because f_{r_n} is high-dimensional and sparse, we first performed the mean aggregation of the attribute vectors of the k th-order drug neighbors of r_i , and the aggregated vector $h_{Rr_i}^{(k)}$ is expressed as:

$$h_{Rr_i}^{(k)} = \text{mean}\left(\left\{f_{r_1}, \dots, f_{r_n}, \dots\right\}\right), r_n \in N_{Rr_i}^{(k)} \quad (3)$$

Subsequently, we project $h_{Rr_i}^{(k)}$ into the low-dimensional feature space through a fully connected network and obtain the low-dimensional k th-order drug neighbor feature vector $u_{Rr_i}^{(k)}$ as follows:

$$u_{Rr_i}^{(k)} = \sigma \left(W_R^{(k)} h_{Rr_i}^{(k)} + b_R^{(k)} \right) \quad (4)$$

where σ denotes the activation function *ReLU* [45], $W_R^{(k)}$ the weight matrix when the neighbor type is a drug, and $b_R^{(k)}$ the bias vector. K denotes the total number of orders, and $K = 2$ in our model.

3.3.4. Same-Type Neighbor Topology Encoding Based on Neighbor-Scale-Level Attention

Because the drug (disease)-type neighbor node information at different scales of r_i contributes differently to the learning of the drug (disease) neighbor topological representation of r_i , we established a neighbor-scale-level attention to learn the attention weights of order $1-k$ neighbor feature vectors of the same type. For the k th-order drug neighbor feature $u_{Rr_i}^{(k)}$ of r_i , with attention score s_k^{Scale} ,

$$s_k^{Scale} = h^{Scale} \tanh \left(W^{Scale} u_{Rr_i}^{(k)} + b^{Scale} \right), \quad (5)$$

where h^{Scale} is the weight vector at the neighbor scale level; W^{Scale} and b^{Scale} are the weight matrix and bias vector, respectively. The normalized attention coefficient is α_k^{Scale} , which can be obtained using the *softmax* function, as follows:

$$\alpha_k^{Scale} = \frac{\exp \left(s_k^{Scale} \right)}{\sum_{n \in K} \exp \left(s_n^{Scale} \right)} \quad (6)$$

The drug neighbor topology representation u_{Rr_i} of r_i obtained using the attention mechanism is:

$$u_{Rr_i} = \sum_k \alpha_k^{Scale} u_{Rr_i}^{(k)} \quad (7)$$

3.3.5. Neighbor Topology Encoding Based on Attention Enhancement at the Neighbor Topology Level

r_i contains two types of neighbor nodes, drug and disease, whose neighbor topologies are represented as u_{Rr_i} and u_{Dr_i} , respectively. However, the importance of different types of neighbor nodes for association prediction varies, and neighbor-topology-level attention is proposed to enhance the neighbor topology representation of r_i . The attention score for the same-type neighbor topology representation of r_i is:

$$s_t^{Topo} = h^{Topo} \tanh \left(W^{Topo} u_{tr_i} + b^{Topo} \right), \quad (8)$$

where $t \in \{R, D\}$, W^{Topo} and h^{Topo} are the neighbor-topology-level weight matrix and weight vector, respectively, and b^{Topo} is a bias vector. The normalized attention weights α_t^{Topo} are expressed as follows:

$$\alpha_t^{Topo} = \frac{\exp \left(s_t^{Topo} \right)}{\sum_{n \in 2} \exp \left(s_n^{Topo} \right)} \quad (9)$$

Finally, the augmented representation of the r_i neighbor topology obtained using the attention mechanism is u_{r_i} , expressed as follows:

$$u_{r_i} = \sum_t \alpha_t^{Topo} u_{tr_i} \quad (10)$$

Here, $u_{r_i}^{(m)}$ denotes the neighboring topological representation obtained by r_i in U^m , where $m \in \{1, 2, 3\}$.

Similarly, the neighbor topology representation $u_{d_j}^{(m)}$ of d_j in U^m can be obtained. These neighboring topological representations are used to form the feature matrices S of r_i - d_j node pairs, as follows:

$$S = \begin{bmatrix} u_{r_i}^{(1)} & u_{r_i}^{(2)} & u_{r_i}^{(3)} \\ u_{d_j}^{(1)} & u_{d_j}^{(2)} & u_{d_j}^{(3)} \end{bmatrix} \in \mathbb{R}^{2 \times (N_f + N_f + N_f)}, \quad (11)$$

where N_f denotes the dimension number of the neighbor topology representation.

3.3.6. CNN-Based Pairwise Neighbor Topology Encoding

The feature matrix of the first branch S is passed into the CNN, which learns the r_i - d_j neighbor topology representations. We filled the periphery of S with zeros to learn the edge features of S and then obtained the new matrix \hat{S} .

We established a CNN module using convolutional and pooling layers. The filter length and breadth relative to the convolution layer are denoted by w_l and w_h , respectively; a total of n_{conv} filters were used. After applying the convolution filter $W_{conv} \in \mathbb{R}^{w_l \times w_h \times n_{conv}}$ to \hat{S} , a feature map $Z \in \mathbb{R}^{n_{conv} \times (4 - w_l + 1) \times (2 + N_f + N_f + N_f - w_h + 1)}$ was generated. $\hat{S}_{k,i,j}$ represents the sliding of the k -th filter to position (i, j) of \hat{S} , and it is defined as:

$$\hat{S}_{k,i,j} = \hat{S}(i : i + w_l, j : j + w_h), \quad \hat{S}_{k,i,j} \in \mathbb{R}^{w_l \times w_h}, \quad (12)$$

where $i \in [1, 4 - w_l + 1]$, $j \in [1, 2 + (N_f + N_f + N_f) - w_h + 1]$, and $k \in [1, n_{conv}]$. The element value $Z_k(i, j)$ of the filter $W_{k,i,j}$ sliding on $\hat{S}_{k,i,j}$ to Z_k is:

$$Z_k(i, j) = \sigma(W_{k,i,j} * \hat{S}_{k,i,j} + b(k)), \quad (13)$$

where σ is the *ReLU* function and b the bias vector. The position (i, j) in the feature map Z_k is represented by $Z_k(i, j)$.

The more significant features of Z_k were extracted using the max-pooling layer. The filter length of the max-pooling layer is w_e , and the width is w_b . The k -th feature map of all feature maps P output by the pooling layer is P_k , and $P_k(i, j)$ can be calculated as:

$$P_k(i, j) = \text{Max}(Z_k(i : i + w_e, j : j + w_b)), \quad (14)$$

where $i \in [1, 5 - w_l - w_e + 1]$, $j \in [1, (N_f + N_f + N_f) + 3 - w_h - w_b + 1]$, and $k \in [1, n_{conv}]$.

In the CNN module, we set the number of filters in the convolutional layer to 16, the kernel size to 2×2 , and the stride size to 1. In the pooling layer, the kernel size was set to 2×2 , and the step size and zero-padding were set to 1 and 0, respectively. After performing processing in the convolution and max-pooling layers, the output vector z_{NT} was obtained. Subsequently, z_{NT} was input to the fully connected and *softmax* layer [46], which yielded the association probability distributed for the first branch, as follows:

$$\text{Score}_{NT} = \text{softmax}(W_{soft1} z_{NT} + b_{soft1}), \quad (15)$$

where W_{soft1} is the first branch of the fully connected layer's weight matrix and b_{soft1} is the corresponding bias vector. Score_{NT} indicates the association probability distribution for the $C(C = 2)$ classification, including the likelihood of a drug and disease being associated and otherwise.

3.4. Encoding Pairwise Node Attributes

3.4.1. Attribute Embedding Matrix for Drug–Disease Pairs

We introduced an embedding strategy to extract the nodal attributes of drug–disease pairs (Figure 5). If $r_i(d_j)$ is similar (related) to a more typical drug or related (similar) to a

disease, then r_i-d_j is likely to be related. Therefore, information regarding the properties of drugs and diseases must be learned from the pairwise node level.

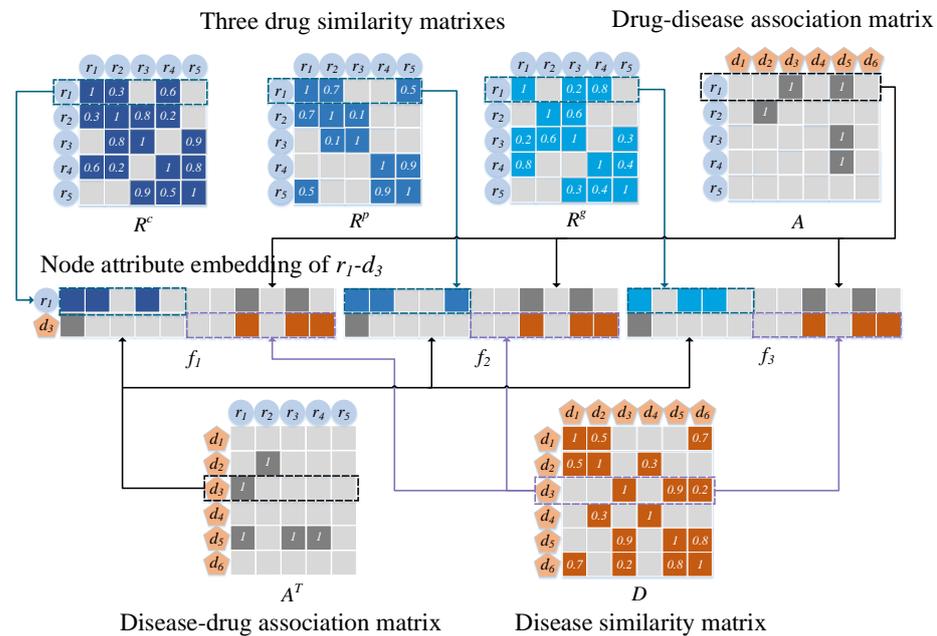


Figure 5. Illustration of constructing an attribute embedding matrix for a pair of drug and disease nodes.

For a heterogeneous drug–disease network U^m , U_i^m contains the m -th similarity of r_i with all drugs and the association with all diseases, and $U_{N_r+j}^m$ contains the association of d_j with all drugs and the similarity with all diseases. Therefore, we used the attribute vectors U_i^m and $U_{N_r+j}^m$ ($m = 1, 2, 3$) to perform splicing such that the attribute embedding matrix P of r_i and d_j can be obtained. P is expressed as follows:

$$P = \begin{bmatrix} U_i^1 & U_i^2 & U_i^3 \\ U_{N_r+j}^1 & U_{N_r+j}^2 & U_{N_r+j}^3 \end{bmatrix} \in \mathbb{R}^{2 \times ((N_r + N_d) \times 3)}, \quad (16)$$

where P has a dimension of $2 \times ((N_r + N_d) \times 3)$.

3.4.2. CAE-Based Pairwise Node Attribute Encoding

Because the node attribute matrix P obtained from the three heterogeneous networks is high-dimensional and sparse, meaningless and non-representative information may be present. Therefore, we performed encoding and decoding based on a CAE to comprehensively learn the attribute information of drug–disease pairs in the original data distribution, as shown in Figure 3.

Encoder: Two hidden layers, each comprising a convolutional layer and a max-pooling layer, constitute the encoder. The edge features of P should be preserved and learned via zero-padding. The first hidden layer uses the zero-padded P as input and yields the feature map $Z_{Encoder}^{(1)}$ encoded as:

$$Z_{Encoder}^{(1)} = \max \left(\sigma \left(W_{Encoder}^{(1)} * P + b_{Encoder}^{(1)} \right) \right) \quad (17)$$

Subsequently, the feature map of the t -th layer $Z_{Encoder}^{(t)}$ is generated as follows:

$$Z_{Encoder}^{(t)} = \max \left(\sigma \left(W_{Encoder}^{(t)} * Z_{Encoder}^{(t-1)} + b_{Encoder}^{(t)} \right) \right) \quad (18)$$

where σ is the *ReLU* function. $W_{Encoder}^{(t)}$ denotes the encoder's t -th hidden layer's weight matrix, and $b_{Encoder}^{(t)}$ is the corresponding bias vector. $t = 2, \dots, L_{En}$. L_{En} indicates the encoder's total number of layers, and the convolution computation is indicated by " \star "; *max* denotes the max-pooling processing for capturing the most critical features within every feature map by downsampling the potential representations acquired from the convolution layer.

Decoder: Using the decoder, we projected the $Z_{Encoder}^{(L_{En})}$ code such that it returns to its initial space and reassembled it to obtain the decoding matrix. The variance between the decoding matrix and the initial matrix P was evaluated, and an optimal coded feature map was obtained. Three hidden layers, each with a transposed convolutional layer, constitute the decoder. For $Z_{Encoder}^{(L_{En})}$ as the input of the first hidden layer of the decoder, the feature map $Z_{Decoder}^{(1)}$ is obtained as follows:

$$Z_{Decoder}^{(1)} = \sigma\left(W_{Decoder}^{(1)} \star Z_{Encoder}^{(L_{En})} + b_{Decoder}^{(1)}\right) \quad (19)$$

$$Z_{Decoder}^{(l)} = \sigma\left(W_{Decoder}^{(l)} \star Z_{Decoder}^{(l-1)} + b_{Decoder}^{(l)}\right) \quad (20)$$

where $W_{Decoder}^{(l)}$ is the weight matrix of the decoder and $b_{Decoder}^{(l)}$ is the decoder's bias vector. $l = 2, \dots, L_{De}$. A total of L_{De} decoder layers are involved. The operator " \star " indicates the transposed convolution computation. The reconstructed matrix \hat{P} is the output $Z_{Decoder}^{(L_{De})}$ of the last layer of the decoder.

Optimization: Our optimization objective was to render \hat{P} as consistent as possible with the input P . The loss function is expressed as:

$$loss_{auto} = \frac{1}{T_{train}} \sum_{n=1}^{T_{train}} (P_n - \hat{P}_n)^2, \quad (21)$$

where P is the input of the encoder, \hat{P} the output at the decoder, T_{train} the number of training samples, and P_n the embedding matrix of the n th drug–disease pair in the corresponding training sample. Adam's algorithm [47] was used to optimize $loss_{auto}$. The back propagation [48] approach was used to train the CAE and update $loss_{auto}$. Using the iterative algorithm, the pairwise property encoding was regarded as the output $Z_{Encoder}^{(L_{En})}$ of the last encoder layer, denoted by F_{PA} .

To acquire the association probability of the second branch of node pair r_i – d_j $Score_{PA}$, F_{PA} was processed in the fully connected and *softmax* layer. $Score_{PA}$ is expressed as:

$$Score_{PA} = softmax(W_{soft2}F_{PA} + b_{soft2}), \quad (22)$$

where W_{soft2} and b_{soft2} are the weight matrix and bias vector of the fully connected second branch, respectively. $Score_{PA}$ is the association probability distribution for the C ($C = 2$) classification.

3.5. Final Integration and Optimization

The loss function in the first branch can be expressed as the cross-entropy between the true label y_{NT} and the drug–disease association prediction result $Score_{NT}$, as follows:

$$loss_{NT} = - \sum_{i=1}^{N_{train}} \sum_{j=1}^C y_{NTj} \log(Score_{NTj}), \quad (23)$$

where N_{train} is the set of training samples and y_{NTj} represents the probability of a drug–disease association. If an r_i – d_j pair has an association, then y_{NTj} is 1; otherwise, it is 0. In the second branch, the cross-entropy loss function $loss_{PA}$ is defined as:

$$loss_{PA} = - \sum_{i=1}^{N_{train}} \sum_{j=1}^C y_{PAj} \log (Score_{PAj}) \quad (24)$$

We trained the loss functions $loss_{NT}$ and $loss_{PA}$ separately until their minimum values were attained. The final correlation prediction score is calculated as follows:

$$Score = \lambda \times Score_{NT} + (1 - \lambda) \times Score_{PA}, \quad (25)$$

where λ denotes a hyperparameter that ranges from 0 to 1 and was used to measure the contribution of neighboring topologies and pairwise node attributes to the association prediction score.

4. Conclusions

We proposed the NAPred method to determine the association between drug candidates and diseases. The three proposed heterogeneous networks facilitated neighbor topology extraction and pairwise node attribute embedding using multiscale meta-paths. A framework comprising a convolutional neural network with attention mechanisms and CAE was constructed to encode and integrate neighbor topological representations and pairwise attribute representations. Two attention mechanisms were proposed to assign greater weights to multi-scale features and topologies. NAPred's ability to discover potentially relevant diseases for drugs was validated through case studies and a cross-validation of five drugs. Numerous experimental results showed that NAPred's predictions outperformed existing methods. Our predictive model serves as a tool for screening to recognize potential drug–disease associations, thereby allowing biologists to conduct wet laboratory research for determining real drug–disease associations.

Supplementary Materials: The following are available at: <https://www.mdpi.com/article/10.3390/ijms23073870/s1>.

Author Contributions: P.X. designed the method and participated in manuscript writing; Z.L. developed the experiments and participated in manuscript writing; T.Z. participated in the method's design; Y.L. participated in the experimental design; T.N. participated in the experimental design. All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported by the Natural Science Foundation of China (62172143, 61972135); Heilongjiang Provincial Natural Science Foundation of China (LH2019F049, LH2019A029, LH2020F043); China Postdoctoral Science Foundation (2020M670939, 2019M650069); Heilongjiang Postdoctoral Scientific Research Staring Foundation (BHLQ18104).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Chen, H.; Cheng, F.; Li, J. iDrug: Integration of drug repositioning and drug-target prediction via cross-network embedding. *PLoS Comput. Biol.* **2020**, *16*, e1008040. [CrossRef] [PubMed]
- Ceddia, G.; Pinoli, P.; Ceri S.; Masseroli, M. Matrix Factorization-based Technique for Drug Repurposing Predictions. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 3162–3172. [CrossRef] [PubMed]
- Luo, H.; Li, M.; Yang, M.; Wu, F.X.; Li, Y.; Wang, J. Biomedical data and computational models for drug repositioning: A comprehensive review. *Briefings Bioinform.* **2021**, *22*, 1604–1619. [CrossRef]
- Pushpakom, S.; Iorio, F.; Eyers, P.A.; Escott, K.J.; Hopper, S.; Wells, A.; Andrew, D.; Tim, G.; Joanna, L.; Christine, M.; et al. Drug repurposing: Progress, challenges and recommendations. *Nat. Rev. Drug Discov.* **2019**, *18*, 41–58. [CrossRef] [PubMed]
- Turanli, B.; Altay, O.; Borén, J.; Hasan, T.; Jens, N.; Mathias, U.; Yalcin, A.K.; Adil, M. Systems biology based drug repositioning for development of cancer therapy. *Semin. Cancer Biol.* **2021**, *68*, 47–58. [CrossRef]
- Padhy, B.; Gupta, Y. Drug repositioning: Re-investigating existing drugs for new therapeutic indications. *J. Postgrad. Med.* **2011**, *57*, 153–160. [CrossRef]
- Pritchard, J.-L.E.; O'Mara, T.A.; Glubb, D.M. Enhancing the Promise of Drug Repositioning through Genetics. *Front. Pharmacol.* **2017**, *8*, 896. [CrossRef]
- Novac, N. Challenges and opportunities of drug repositioning. *Trends Pharmacol. Sci.* **2013**, *34*, 267–272. [CrossRef]

9. Alfedì, G.; Luffarelli, R.; Condò, I.; Pedini, G.; Mannucci, L.; Massaro, D.S. Drug repositioning screening identifies etravirine as a potential therapeutic for friedreich's ataxia. *Mov. Disord.* **2019**, *34*, 323–334. [[CrossRef](#)]
10. Karaman, B.; Sippl, W. Computational Drug Repurposing: Current Trends. *Curr. Med. Chem.* **2019**, *26*, 5389–5409. [[CrossRef](#)]
11. Shameer, K.; Readhead, B.; Dudley, J.T. Computational and experimental advances in drug repositioning for accelerated therapeutic stratification. *Curr. Top. Med. Chem.* **2015**, *15*, 5–20. [[CrossRef](#)] [[PubMed](#)]
12. Gottlieb, A.; Stein, G.Y.; Ruppín, E.; Sharan, R. PREDICT: A method for inferring novel drug indications with application to personalized medicine. *Mol. Syst. Biol.* **2011**, *7*, 496. [[CrossRef](#)]
13. Zhang, W.; Yue, X.; Lin, W.; Wu, W.; Liu, R.; Huang, F.; Liu, F. Predicting drug-disease associations by using similarity constrained matrix factorization. *BMC Bioinform.* **2018**, *19*, 233. [[CrossRef](#)] [[PubMed](#)]
14. Wang, Y.; Chen, S.; Deng, N.; Wang, Y. Drug Repositioning by Kernel-Based Integration of Molecular Structure, Molecular Activity, and Phenotype Data. *PLoS ONE* **2013**, *8*, e78518.
15. Liang, X.; Zhang, P.; Yan, L.; Fu, Y.; Peng, F.; Qu, L. LRSSL: predict and interpret drug-disease associations based on data integration using sparse subspace learning. *Bioinformatics* **2017**, *33*, 1187–1196. [[CrossRef](#)] [[PubMed](#)]
16. WWang, W.; Yang, S.; Zhang, X.; Li, J. Drug repositioning by integrating target information through a heterogeneous network model. *Bioinformatics* **2014**, *30*, 2923–2930. [[CrossRef](#)]
17. Liu, H.; Song, Y.; Guan, J.; Luo, L.; Zhuang, Z. Inferring new indications for approved drugs via random walk on drug-disease heterogeneous networks. *BMC Bioinform.* **2016**, *17*, 539. [[CrossRef](#)]
18. Luo, H.; Wang, J.; Li, M.; Luo, J.; Peng, X.; Wu, F.X.; Pan, Y. Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm. *Bioinformatics* **2016**, *32*, 2664–2671. [[CrossRef](#)]
19. Yu, L.; Su, R.; Wang, B.; Zhang, L.; Zou, Y.; Zhang, J.; Gao, L. Prediction of Novel Drugs for Hepatocellular Carcinoma Based on Multi-Source Random Walk. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2017**, *14*, 966–977. [[CrossRef](#)]
20. Huang, Y.-F.; Yeh, H.-Y.; Soo, V.-W. Inferring drug-disease associations from integration of chemical, genomic and phenotype data using network propagation. *BMC Med. Genom.* **2013**, *6*, S4. [[CrossRef](#)]
21. Chen, H.; Zhang, Z.; Peng, W. miRDDCR: A miRNA-based method to comprehensively infer drug-disease causal relationships. *Sci. Rep.* **2017**, *7*, 15921. [[CrossRef](#)] [[PubMed](#)]
22. Xuan, P.; Zhang, Y.; Zhang, T.; Li, L.; Zhao, L. Predicting MiRNA-Disease Associations by Incorporating Projections in Low-Dimensional Space and Local Topological Information. *Genes* **2019**, *10*, 685. [[CrossRef](#)] [[PubMed](#)]
23. Xuan, P.; Pan, S.; Zhang, T.; Liu, Y.; Sun, H. Graph Convolutional Network and Convolutional Neural Network Based Method for Predicting LncRNA-Disease Associations. *Cells* **2019**, *8*, 1012. [[CrossRef](#)]
24. Xuan, P.; Sheng, N.; Zhang, T.; Liu, Y.; Guo, Y. CNN-DLP: A Method Based on Convolutional Autoencoder and Convolutional Neural Network with Adjacent Edge Attention for Predicting LncRNA-Disease Associations. *Int. J. Mol. Sci.* **2019**, *20*, 4260. [[CrossRef](#)] [[PubMed](#)]
25. Xuan, P.; Gao, L.; Sheng, N.; Zhang, T.; Nakaguchi, T. Graph Convolutional Autoencoder and Fully-Connected Autoencoder with Attention Mechanism Based Method for Predicting Drug-Disease Associations. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 1793–1804. [[CrossRef](#)] [[PubMed](#)]
26. Xuan, P.; Ye, Y.; Zhang, T.; Zhao, L.; Sun, C. Convolutional Neural Network and Bidirectional Long Short-Term Memory-Based Method for Predicting Drug-Disease Associations. *Cells* **2019**, *8*, 705. [[CrossRef](#)]
27. Jiang, H.-J.; Huang, Y.-A.; You, Z.-H. Predicting Drug-Disease Associations via Using Gaussian Interaction Profile and Kernel-Based Autoencoder. *Biomed Res. Int.* **2019**, *2019*, 2426958. [[CrossRef](#)]
28. Hajian-Tilaki, K. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Casp. J. Intern. Med.* **2013**, *4*, 627–635.
29. Saito, T.; Rehmsmeier, M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS ONE* **2015**, *10*, e0118432. [[CrossRef](#)]
30. Ling, C.X.; Huang, J.; Zhang, H. AUC: A Better Measure than Accuracy in Comparing Learning Algorithms. In *Conference of the Canadian Society for Computational Studies of Intelligence*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 329–341.
31. Bolboacă, S.D.; Jäntschi, L. Predictivity Approach for Quantitative Structure-Property Models. Application for Blood-Brain Barrier Permeation of Diverse Drug-Like Compounds. *Int. J. Mol. Sci.* **2011**, *12*, 4348. [[CrossRef](#)]
32. Bolboacă, S.D.; Jäntschi, L. Sensitivity, Specificity, and Accuracy of Predictive Models on Phenols Toxicity. *J. Comput. Sci.* **2014**, *5*, 345–350. [[CrossRef](#)]
33. Pahikkala, T.; Airola, A.; Pietilä, S.; Shakyawar, S.; Szwajda, A.; Tang, J.; Aittokallio, T. Toward more realistic drug-target interaction predictions. *Briefings Bioinform.* **2015**, *16*, 325–337. [[CrossRef](#)] [[PubMed](#)]
34. Duran-Frigola, M.; Fernández-Torras, A.; Bertoni, M.; Aloy, P. Formatting biological big data for modern machine learning in drug discovery. *WIREs Comput. Mol. Sci.* **2019**, *9*, e1408. [[CrossRef](#)]
35. Davis, A.P.; Grondin, C.J.; Johnson, R.J.; Sciaky, D.; McMorran, R.; Wieggers, J. The Comparative Toxicogenomics Database: Update 2019. *Nucleic Acids Res.* **2019**, *47*, D948–D954. [[CrossRef](#)]
36. Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074–D1082. [[CrossRef](#)]
37. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A. PubChem Substance and Compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213. [[CrossRef](#)]

38. Wang, F.; Zhang, P.; Cao, N.; Hu, J.; Sorrentino, R. Exploring the associations between drug side-effects and therapeutic indications. *J. Biomed. Inform.* **2014**, *51*, 15–23. [[CrossRef](#)]
39. Wang, Y.; Xiao, J.; Suzek, T.O.; Zhang, J.; Wang, J.; Bryant, S.H. PubChem: A public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.* **2009**, *37*, W623–W633. [[CrossRef](#)]
40. Mitchell, A.; Chang, H.-Y.; Daugherty, L.; Fraser, M.; Hunter, S.; Lopez, R.; Craig, M.; Connor, M.; Gift, N.; Sebastien, P.; et al. The InterPro protein families database: The classification resource after 15 years. *Nucleic Acids Res.* **2015**, *43*, D213–D221. [[CrossRef](#)]
41. The UniProt Consortium. The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.* **2010**, *38*, D142–D148. [[CrossRef](#)]
42. Wang, D.; Wang, J.; Lu, M.; Song, F.; Cui, Q. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics* **2010**, *26*, 1644–1650. [[CrossRef](#)] [[PubMed](#)]
43. Wang, X.; Ji, H.; Shi, C.; Wang, B.; Ye, Y.; Cui, P.; Yu, P.S. Heterogeneous Graph Attention Network. *arXiv* **2019**, arXiv:1903.07293.
44. Cen, Y.; Zou, X.; Zhang, J.; Yang, H.; Zhou, J.; Tang, J. Representation Learning for Attributed Multiplex Heterogeneous Network. *arXiv* **2019**, arXiv:1905.01669.
45. Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th International Conference on International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010; pp. 807–814.
46. Bahdanau, D.; Cho, K.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv* **2014**, arXiv:1409.0473.
47. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
48. Petrini, M. Improvements to the Backpropagation Algorithm. *Ann. Univ. Petrosani Econ.* **2012**, *12*, 185–192.