DOI: 10.15849/IJASCA.251130.13

Int. J. Advance Soft Compu. Appl, Vol. 17, No. 3, November 2025
Print ISSN: 2710-1274, Online ISSN: 2074-8523

TSDR: A Two-Step Method for Drug Repurposing Based on Enhancing Drug Features

Saif Mohsin Najm[1*], Fatemeh Roohian[2], and Amir Lakizadeh[3]

[1] Department Computer Engineering and Information Technology, University of Qom, Qom 3716146611, Iran
e-mail: s.suhail@stu.qom.ac.ir; f.roohian@stu.ac.ir; lakizadeh@qom.ac.ir

Abstract

With the emergence of new diseases and the need for their treatment, the need to discover and develop new drugs has increased. In contrast, traditional drug development and discovery is a time-consuming, costly, and sometimes risky process. So, computational approaches have been considered widely for drug repurposing tasks. In this study, a new deep learning-based architecture for drug-disease association prediction is presented. The proposed model (TSDR) has two steps. In the first step, each subnetwork extracts a compact feature vector for the input drug-disease pair. For each input pair drug-disease, the initial drug features are constructed solely based on several primary drug datasets to assess separately the role of each initial drug dataset in the prediction of drug-disease association. In the second step, the individual drug feature vectors are combined and enhanced by NLP-based drug features to contract the final features of a drug. The enhanced drug vector, besides a disease feature vector, is used as input to train another convolutional neural network to predict the drug-disease association link. The final prediction is obtained at the output of the second step of the model. Experiment findings demonstrate the efficiency of the proposed model in extracting a more efficient feature vector and improving the accuracy of the drug repurposing task so that the proposed technique has achieved an accuracy of 96.34 in the AUC and 96.89 in the AUPR criterion.

Keywords: *Data fusion, deep learning, drug repurposing, drug-disease network, network embedding.*

1 Introduction

Because of the increasing spread of diseases, discovery and drug development for effective treatment will be a long process and, in some cases, impossible. Making medicine in traditional (laboratory) ways is a long and dangerous process and cannot be effective in finding the correct medication for diseases. Repurposing existing available drugs is an appropriate and relatively inexpensive way to find an effective treatment for diseases [1, 2]. Drug discovery and development is inherently a high-risk and resource-intensive process. Historically, traditional approaches have yielded low success rates, with only a small fraction of compounds progressing from early clinical trials to market approval. Recent empirical analyses indicate that the probability of a drug entering Phase I clinical trials and ultimately receiving FDA approval averages around 14.3%, though this rate varies widely across therapeutic areas and development strategies. The majority of

Received 4 August 2025; Accepted 22 September 2025

investigational drugs are discontinued due to safety concerns, adverse side effects, or insufficient efficacy in targeting the intended biological mechanism [3]. These attrition rates underscore the need for more predictive preclinical models and innovative trial designs to improve translational success. Recent analyses estimate that the average capitalized cost of developing a new molecular entity (NME) ranges from approximately $944 million to $2.8 billion, with development timelines spanning 10 to 17 years depending on therapeutic area and trial complexity [4, 5]. Drug repurposing has emerged as a promising alternative, often leveraging clinical observations and post-marketing data to identify new therapeutic indications for existing compounds [6]. While this strategy offers reduced cost and accelerated timelines, a major technical challenge in computational drug repurposing lies in the integration and harmonization of heterogeneous biomedical data sources including genomic, phenotypic, and pharmacological datasets which are often fragmented and context-dependent [7, 8].

The issue of repurposing approved drugs is a strategy for discovering and developing treatments for various diseases. Drug targeting was introduced by Ashburn and Thor [9] [10]. Repurposing is a new method of discovering effective drugs for a variety of diseases, especially for rare and unknown diseases with a low statistical population [11]. It is also a beneficial method for counteracting drug resistance. Drug resistance ranks high among the primary causes of reducing the effectiveness of drugs, for example, overcoming antimicrobial resistance with non-antibiotic medications [12, 13]. Using several drugs simultaneously can cause adverse effects on the patient's body. Drug repurposing can be used to find the right combination of drugs for treatment [14].

Much research has been done on drug repurposing, including the models according to deep learning; the following can be mentioned. The deepDR model [15] learns the high-level properties of drugs from heterogeneous networks by applying a deep automatic encoder network. Disease-drug interactions are encrypted and decrypted through a deep automated encryption network to infer a new option for approved drugs. The DeepDTA [16] model uses a convolution neural network to learn the sequence of proteins and drugs to predict drug-target communication. The LASSO-DNN model [17] is according to the properties extracted from LASSO regression models and utilizing proteins and explicit drug properties; it has been designed and shown that effective representations of drug properties and drug targets are essential for constructing predictive learning models of interactions between drug targets. The SCMFDD model [18] aims to predict new and unknown drug-disease relationships using a new method in matrix factorization that combines drug properties and disease semantic information and provides an optimal and efficient algorithm for predicting drug-disease relationships. The NeoDTI model [19] has been proposed for new drug repurposing and new drug delivery. NeoDTI introduces a nonlinear end-to-end model for forecasting. In this method, the properties of each node in the network that represents a drug are learned by maintaining the topology and using different drug networks. It is used to predict interactions between drug targets. The HNRD model is a model according to the NeoDTI model. In this method, neighborhood aggregation based on disease-disease and drug-drug networks is used, and the information about each node that represents a drug is integrated using the information about each node in the network. The drug-disease matrix that has been rebuilt is utilized to anticipate drug-disease links. The DDA-SKF model [20] proposes the prediction of associations between drugs and diseases according to fusion nuclei (similarity nuclei) to predict drug-disease associations, and the similarity nucleus combination is used to combine various similarity nuclei with Laplace regular least squares procedures. The NAPred model [21], a

modulation-based automatic convolution encoder, is proposed to encode drug-disease pair properties in three heterogeneous networks.

Despite significant progress in computational drug repurposing, a clear gap remains in effectively combining heterogeneous drug and disease data. Existing models often depend heavily on single data sources or perform early fusion strategies that overlook the complementary value of diverse feature types, while the potential of natural language processing (NLP)-based biomedical embeddings for drug representation has not been fully exploited. To address this limitation, the proposed TSDR framework introduces a two-step CNN-based architecture that first extracts compact feature vectors from multiple drug datasets independently, allowing a systematic evaluation of each source's predictive contribution, and then enriches these representations with BioWordVec-derived embeddings before fusing them with disease features. This staged design yields a semantically enhanced and more representative feature space, offering a structured and interpretable integration strategy that distinguishes TSDR from existing multi-source fusion approaches such as DeepDR, NeoDTI, and MGATRx.

A two-step deep learning-based architecture is proposed to address the drug–disease link prediction task. In the first step, subnetworks are designed to extract compact feature vectors from input drug–disease pairs. Initial drug features are constructed exclusively from multiple primary drug datasets, which are then concatenated with disease feature vectors and processed through convolutional layers for subnetwork training. Each subnetwork produces a compact feature vector of size 64 based solely on the corresponding input drug dataset. This stage enables a separate evaluation of the contribution of each drug dataset to drug–disease association prediction and facilitates the extraction of the most informative features. The resulting feature vectors from all subnetworks are integrated into a 256-dimensional representation, serving as input for the subsequent step. In the second step, the individual drug feature vectors are fused and further enriched with natural language processing (NLP)-derived drug features to construct final drug representations. Alongside disease feature vectors, these enhanced drug representations are used to train an additional convolutional neural network for predicting drug–disease associations. The final prediction is generated at the output layer of the second stage of the model.

In this study, various data sources have been used such as drug, chemical structure network, disease interaction network, drug-disease network, and protein network of drugs. The proposed model can obtain richer and more representative feature embeddings by combining all of these heterogeneous resources. By using several complementary datasets, it is guaranteed that structural and semantic attributes of drugs and diseases are well represented and thus allows making drug-disease associations more confident, contributing to the development of computational drug repurposing strategies.

2 The Proposed Model

This study uses a two-step deep learning-based architecture based on convolution neural networks to predict the link between drug use and disease. In the first step, several convolution neural networks are trained separately based on different drug datasets. After training the independent subnets, the individual obtained compact features are combined and enhanced by an NLP-based drug feature vector in the second step. Besides a disease feature vector, the enhanced drug vector is used as input to train a convolutional neural network to determine the drug- disease association link. The final prediction of the model is obtained from the second-level convolution network's output.

2.1 Individual feature extraction

In the first step of the model, four convolutional neural (sub) networks are trained separately using the datasets constructed for each drug-disease pair using disease feature vectors and drug-disease, drug- chemical structure, drug target protein domains and drug-gene ontology terms of drug target feature vectors. Each subnetwork extracts a compact feature vector for the input drug-disease pair in this step. In this architecture, for each input pair drug disease, the initial drug features are obtained only based on one of the primary drug datasets (as a binary matrix). Then, these drug features, in concatenation with the disease feature vector, are fed to the convolutional layers to train the subnetwork. Each subnetwork generates an output vector called a compact feature vector (with size 64) for input pair drug-disease only based on its input drug dataset. The main idea of this step is to assess separately the role of each initial drug dataset in drug-disease association prediction and extract the best features in this assessment. Integration of four output vectors of the subnetworks constructs a 256-dimension vector that is used in the next step. Fig. 1 shows a schema of the model in step 1.

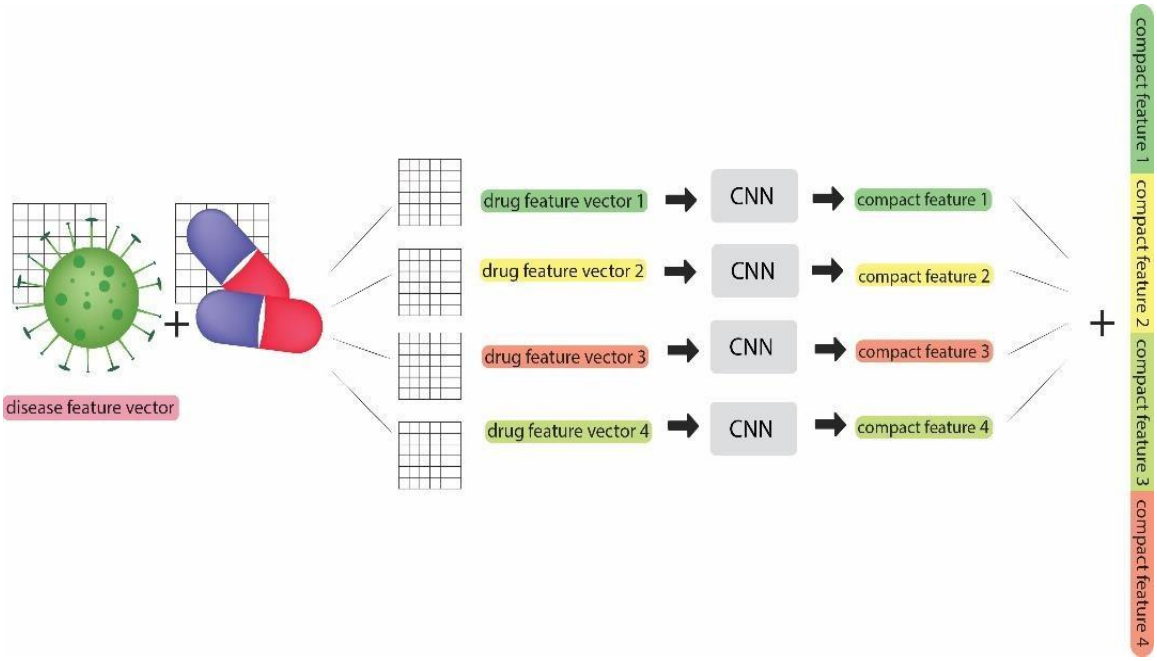

Figure 1. Step 1, extracting compact features.

2.2 Enhancing the features and prediction of the results

The second step predicts the disease-drug association link using a deep convolution neural network. In this step, the combined compact feature vector is enhanced by an NLP-based drug feature vector BioWordVec method [22]. Besides a disease feature vector, the enhanced drug vector is used as input to train a convolutional neural network to predict the link between drug use and disease. The final prediction of the model is obtained based on the outcome of the second-level convolution network. Fig 2 shows a schema of the model in step 2.

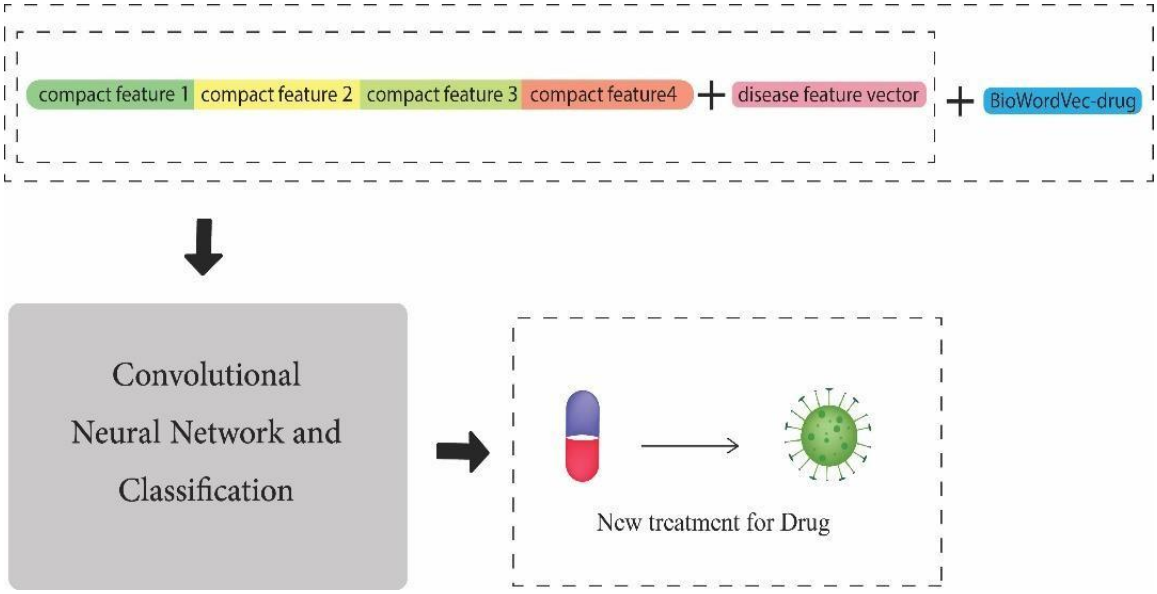

Figure 2. Step 2, drug-disease association prediction.

3 Data Collection

3.1 Drug dataset

The data set used in this study contains 681 diseases and 763 drugs and is taken from a study published by Liang et al. In 2017 [23]. This set includes three drug property matrices that examine drugs in terms of chemical structure, target protein, and gene ontology in the target protein. In addition, it includes a disease properties matrix and the drug-disease relationship matrix.

In addition to the data set used by Liang et al., A drug feature matrix was used using the BioWordVec method based on embedding words in medicine. This method, which is used to embed biomedical words, consists of 2 steps, Figs. 1 and 2:

- Creating a graph of the term MeSH according to its RDF data and sampling the sequences of the term MeSH.
- Learning distributed word embedding according to text and MeSH term sequences utilizing the subword embedding model. Graph relationships: The MeSH term is passed to sequential sequences of title nodes. This process leads to the main header sequences of the MeSH, which in turn combine with the PubMed sequences to learn word assignments.

3.2 Disease dataset

In this study, 681 diseases have been studied. For each disease, there is an aggregated similarity vector. Disease-disease similarity datasets are taken from a published study by Liang et al. [23]. Diseases are considered from the OMIM database [24] based on the MinMiner database's description, and the similarities are based on phenotype and downloaded from MinMiner. In this set, disease similarity data is extracted according to the number of occurrences of MeSH terms (medical terms) in the medical description of each disease pair. The range [0, 1] has already been used to normalize the similarities [25].

3.3 Preparation of neural network input dataset

The data set used in this study includes five separate datasets related to the drugs and one dataset for the disease. Table 1 shows details of the drug datasets utilized in this investigation. The number of drug-disease pairs (class 1) is randomly selected from the unrelated drug-disease pairs (class 0). It should be noted that the lack of a relationship between the drug-disease pair does not mean that it is a member of class 0, and it means that so far, according to the available information, no drug-disease relationship has been found. The number of related drug-disease pairs is 3051; all are included in the network input data set. In addition, 3051 drug- disease pairs unrelated to the available information are randomly selected from 516552 drug-disease pairs. Finally, the dataset used to train the model contains 6102 members connected to drug-disease combinations.

Table 1: The drug datasets

No.	Dataset	Size
1	Drug–disease	763,681
2	Drug-chemical structure	763,623
3	Drug–target protein domains	763,1425
4	Drug-gene ontology terms of drug target	763,4447
5	Drug–BioWordVec	763,763

Finally, five separate datasets are introduced into the model for training, and the data is distributed so that 90% of the data is left for the training process and the remaining 10% for the final testing process. 10% of the training data set is considered for validation.

4 Experiments and Results

The proposed model in this research has been implemented using the Python Keras library in the TensorFlow bed. The model is designed based on deep learning and implemented using functional API in Keras. This model consists of five different inputs and five convolution networks to extract high-level features from the input vectors to the network. Finally, the model has an output that uses a fully connected network, classifying between two classes of "related " or "unrelated" Drug-disease.

This research used a simple hold-out method to evaluate the model. This method's training and testing phases each take up 90% and 10% of the input data, respectively. Initially, 10% of the training data was discarded as validation data for model evaluation during training. Then, after adjusting the hyper-parameters, the model is trained once with all the training data, and with the test data, its final evaluation is performed.

4.1 Evaluating the proposed model

Model performance was quantified following systematic hyper-parameter optimization and full retraining on the complete training set. The final configuration achieved 97% accuracy on the evaluation data (macro-precision = 0.97, macro-recall = 0.97, macro-F1 = 0.97). The confusion matrix evidenced balanced classification, with 280 true negatives, 254 true positives, 8 false positives, and 9 false negatives, yielding per-class precision and recall of approximately 0.97/0.97 for both classes and corresponding. Training dynamics showed a monotonic decline in loss from ≈ 1.0 to ≈ 0.10 and a concomitant rise in accuracy from ≈ 0.65 to ≈ 0.98, reaching a plateau after roughly 20 epochs, indicative of stable convergence under the selected hyper-parameters (Figs. 3–4).

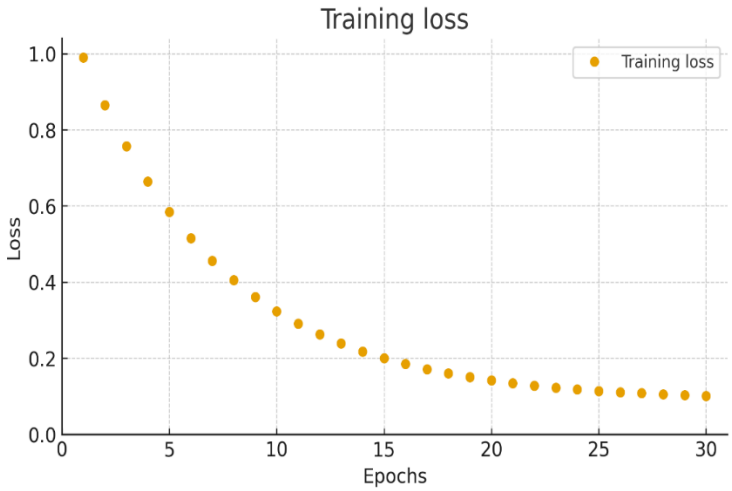

Figure 3. The process of loss changes when teaching the proposed model.

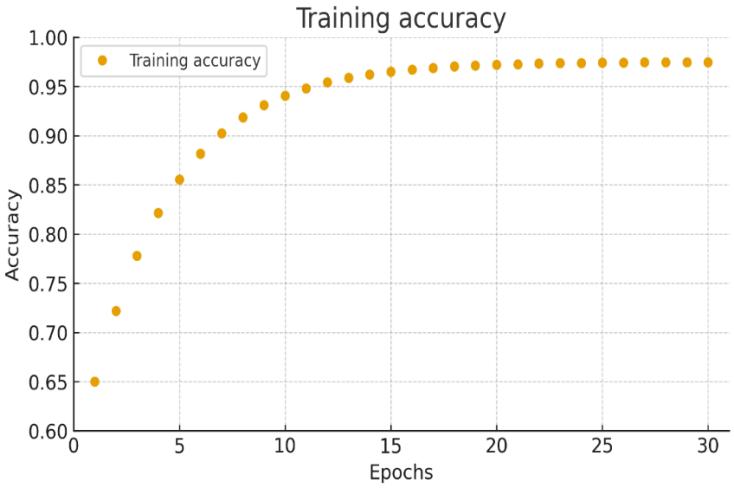

Figure 4. The process of accuracy changes when teaching the proposed model

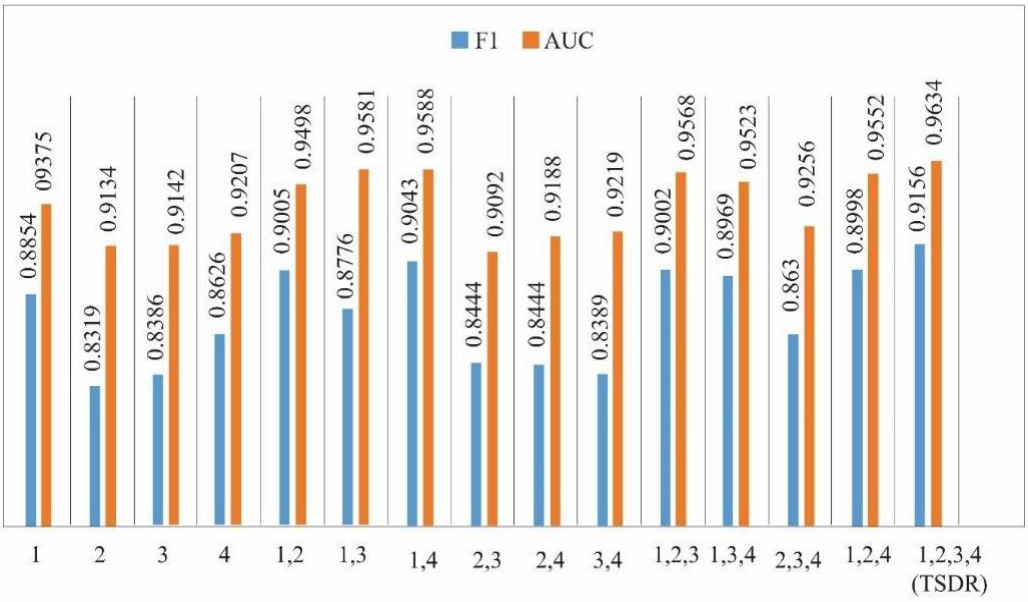

Figure 5. Comparison of model AUC and F1 in combining different drug sources.

4.2 The combination of data sources

To further investigate the suggested model's performance, we used different combinations of data sources, and the obtained results were analyzed. As can be seen in the results, using only one dataset (dataset 1 to 4 in Table 1) and using the drug-disease similarity matrix (dataset 1) gives better results than another dataset. When using the first two datasets, the combination of data sources 1 and 4 (drug-disease similarity matrix and drug-gene ontology) has the best results. In the case of using three datasets, the combination of datasets 1, 2, and 3 (combination of drug-chemical structure, drug-disease, and drug-target protein domains) has been able to achieve higher accuracy than the previous cases.

Finally, the best efficiency and accuracy of the model are found when using all four datasets in step 1 of the proposed model. Details of the outcomes of using the combination of the different datasets are shown in Table 2 and Fig. 5. The outcomes show that better results are obtained in the case of using all four available drug data sources.

Table 2: Classification report results

Class	Precision	Recall	F1-score
0.0	0.97	0.97	0.97
1.0	0.97	0.97	0.97
Accuracy		0.97	
Macro avg	0.97	0.97	0.97

Table 3: Comparison of different combinations of drug features

DATASET(STEP1)	AUC	F1
1	0.9375	0.8854
2	0.9134	0.8319
3	0.9142	0.8386
4	0.9207	0.8626
1, 2	0.9498	0.9005
1, 3	0.9581	0.8776
1, 4	0.9588	0.9043
2, 3	0.9092	0.8444
2, 4	0.9188	0.8444
3, 4	0.9219	0.8389
1, 2, 3	0.9568	0.9002
1, 3, 4	0.9523	0.8969
2, 3, 4	0.9256	0.8630
1, 2, 4	0.9552	0.8998
1, 2, 3, 4 (TSDR)	0.9634	0.9156

4.3 Comparison with the other methods

The suggested model has been compared in this study with several published advanced methods. The results of the advanced models on the published work dataset by Yang et al. Are compared with our proposed model (See Table 3 and Fig. 6). When it comes to drug-disease link prediction, the suggested model outperforms existing techniques and can, therefore, be used as a suitable computational tool in the application of drug repurposing.

The LRSSL method [23] provides a method for learning Laplace's regular scattered subspace. In this method, the drug's chemical information, the drug's target domain, and the target's annotation information are integrated.

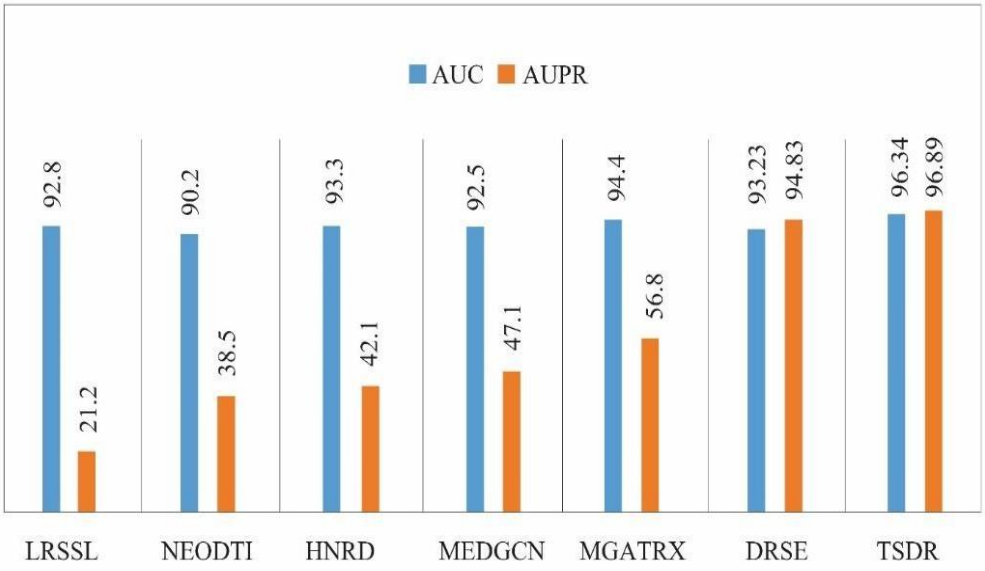

Figure 6. Comparing the performance of the proposed technique with other techniques.

In the NeoDTI method, diverse drug information is integrated into heterogeneous networks, and a nonlinear end-to-end framework is employed to predict drug–disease relationships [19]. Building on this approach, the HNRD method incorporates neighborhood aggregation over disease–disease and drug–drug networks to enhance prediction of drug–disease links.

The MedGCN method presents an intelligent clinical system that automates medication prescribing by formulating drug prediction as learning over a multivariate graph convolutional architecture [26]. The MGATRx method applies a graph attention network to heterogeneous drug and disease data to learn node representations and detect drug–disease relationships [27]. The DRSE method addresses drug repurposing via side-effect information, using stochastic patching to fuse heterogeneous drug–disease features, diffusion component analysis for feature extraction, and matrix factorization to generate final drug–disease predictions [26].

Table 4: Comparing the performance of the proposed technique with other techniques

Method	AUPR	AUC
LRSSL	21.2	92.8
NEODTI	38.5	90.2
HNRD	42.1	93.3
MEDGCN	47.1	92.5
MGATRX	56.8	94.4
DRSE	94.83	93.23
tsDR	96.89	0.9634

5 Case Study

To further investigate the function of the model presented in this study, its prediction for hypertension disease has been studied. Hypertension disease in the data set studied in this study includes 74 drugs, of which 82% (61 drugs) are predicted in the first 100 proposals (Fig. 7). All drugs related to this disease are suggested in the first 133 proposals as effective drugs for hypertension.

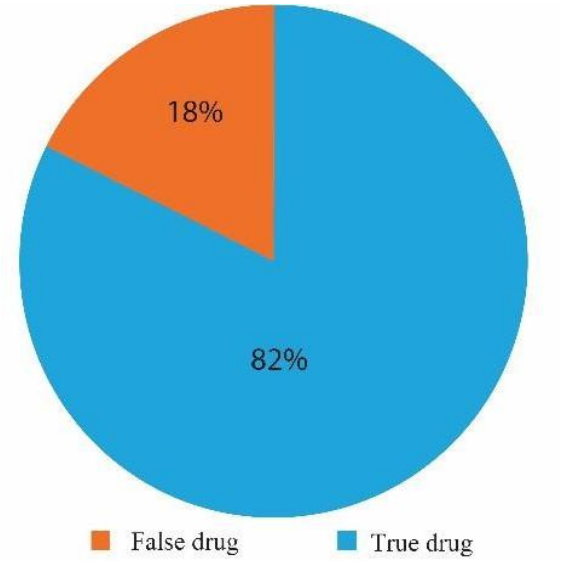

Figure 7. Predicting the proposed model in the first 100 drugs for hypertension.

6 Conclusion

This study introduced TSDR, a two-step deep learning framework for predicting drug–disease associations. The approach first learns compact, dataset-specific drug representations through independent subnetworks and subsequently enhances them by integrating NLP-based biomedical embeddings with disease vectors before final prediction. The methodological contribution lies in evaluating the predictive contribution of each drug feature set separately and then combining them in a structured late-fusion step, which provides richer and more representative embeddings for association tasks.

The observed improvements can be attributed to three design aspects. First, the subnetwork stage enforces compact feature representations that mitigate redundancy across correlated drug attributes and help control overfitting. Second, the use of independent subnetworks preserves complementary information from different data sources such as chemical, biological, and semantic features that early-fusion models may obscure. Third, the incorporation of NLP-derived embeddings supplements structured data by capturing semantic relations from biomedical text and ontologies, thereby enhancing generalization when explicit drug annotations are sparse or incomplete.

Despite these strengths, several limitations should be acknowledged. The current evaluation is restricted to a single benchmark and does not test robustness across different disease categories or external datasets. Negative samples are constructed under the assumption of “no known association,” which may introduce label noise and affect performance estimates. Furthermore, the model depends on static similarity and embedding resources, whose coverage and quality vary across therapeutic areas. The architecture also lacks interpretability beyond dataset-level ablations, which limits its utility for mechanistic insight or hypothesis generation.

Future research directions include:

- Conducting external and cross-dataset validations to assess robustness and generalizability.
- Designing evaluation protocols that incorporate time-aware and uncertainty-based analyses.
- Expanding data integration to include omics profiles, metabolic pathways, and clinical data.
- Exploring graph-based and hypergraph neural architectures to capture higher-order biomedical relationships.
- Enhancing model interpretability through attribution techniques that highlight the most influential drug and disease features.

In summary, TSDR demonstrates that a staged learning process combining compact feature extraction with semantically enriched fusion can improve the representation of heterogeneous biomedical data and strengthen drug–disease association prediction. Addressing the outlined limitations will be essential for developing more robust, interpretable, and clinically useful computational drug repurposing tools.

References

[1] Pinzi, L., Bisi, N., & Rastelli, G. (2024). How drug repurposing can advance drug discovery: challenges and opportunities. *Frontiers in Drug Discovery*, *4*, 1460100.

[2] Nossier, E. S., Anwar, M. M., & El-Zahabi, M. A. (2025). Recent advances in drug repositioning and rediscovery for different therapeutic activities utilizing updated technological approaches. *Molecular Diversity*, 1-28.

[3] Zhou, Y., Zhang, Y., Chen, Z., Huang, S., Li, Y., Fu, J., Zhang, H., Zhao, D., Lian, X., Zhou, Y., & Shen, X. (2024). Dynamic clinical success rates for drugs in the 21st century. *medRxiv*, 2024-02.

[4] Simoens, S., & Huys, I. (2021). R&D costs of new medicines: a landscape analysis. *Frontiers in Medicine*, *8*, 760762.

[5] Schlander, M., Hernandez-Villafuerte, K., Cheng, C. Y., Mestre-Ferrandiz, J., & Baumann, M. (2021). How much does it cost to research and develop a new drug? A systematic review and assessment. *Pharmacoeconomics*, *39(11)*, 1243-1269.

[6] Wiktorowska-Owczarek, A., Iacono, D., & Jasińska-Stroschein, M. (2025). Old drugs: confronting recent advancements and challenges. *Frontiers in Pharmacology*, *16*, 1565890.

[7] Pinzi, L., & Rastelli, G. (2023). Trends and applications in computationally driven drug repurposing. *International Journal of Molecular Sciences*, *24(22)*, 16511.

[8] Sosa, D. N., & Altman, R. B. (2022). Contexts and contradictions: a roadmap for computational drug repurposing with knowledge inference. *Briefings in Bioinformatics*, *23*(4), bbac268.

[9] Pinzi, L., Bisi, N., & Rastelli, G. (2024). How drug repurposing can advance drug discovery: challenges and opportunities. *Frontiers in Drug Discovery*, *4*, 1460100.

[10] Mishra, A. S., Vasanthan, M., & Malliappan, S. P. (2024). Drug repurposing: A leading strategy for new threats and targets. *ACS Pharmacology & Translational Science, 7*(4), 915-932.

[11] Nijim, S., Korsunska, A., Zinski, J., Bolden, S. E., Zuccato, M., Repasky, M., & Fajgenbaum, D. C. (2025). Rare disease drug repurposing. *JAMA Network Open, 8*(5), e258330.

[12] Xiao, G., Li, J., & Sun, Z. (2023). The combination of antibiotic and non-antibiotic compounds improves antibiotic efficacy against multidrug-resistant bacteria. *International Journal of Molecular Sciences, 24*(20), 15493.

[13] Muteeb, G., Rehman, M. T., Shahwan, M., & Aatif, M. (2023). Origin of antibiotics and antibiotic resistance, and their impacts on drug development: A narrative review. *Pharmaceuticals, 16*(11), 1615.

[14] Al Khzem, A. H., Gomaa, M. S., Alturki, M. S., Tawfeeq, N., Sarafroz, M., Alonaizi, S. M., ... & Alghamdi, A. A. (2024). Drug repurposing for cancer treatment: a comprehensive review. *International Journal of Molecular Sciences, 25*(22), 12441.

[15] Mohammadzadeh-Vardin, T., Ghareyazi, A., Gharizadeh, A., Abbasi, K., & Rabiee, H. R. (2024). DeepDRA: Drug repurposing using multi-omics data integration with autoencoders. *PLOS ONE, 19*(7), e0307649.

[16] Deng, J., Zhang, Y., Pan, Y., Li, X., & Lu, M. (2024). Multidta: drug-target binding affinity prediction via representation learning and graph convolutional neural networks. *International Journal of Machine Learning and Cybernetics, 15*(7), 2709-2718.

[17] Abdul Raheem, A. K., & Dhannoon, B. N. (2023). Automating drug discovery using machine learning. *Current Drug Discovery Technologies, 20*(6), 79-86.

[18] Tang, X., Hou, Y., Meng, Y., Wang, Z., Lu, C., Lv, J., Hu, X., Xu, J., & Yang, J. (2025). CDPMF-DDA: contrastive deep probabilistic matrix factorization for drug-disease association prediction. *BMC Bioinformatics, 26*(1), 5.

[19] Parini, P. (2025). The use of different types of networks, alone and in combination, for drug target identification. *British Journal of Pharmacology.*

[20] Gao, C.-Q., Zhou, Y.-K., Xin, X.-H., Min, H., & Du, P.-F. (2022). DDA-SKF: predicting drug–disease associations using similarity kernel fusion. *Frontiers in Pharmacology, 12*, 784171.

[21] Xuan, P., Lu, Z., Zhang, T., Liu, Y., & Nakaguchi, T. (2022). Integration of neighbor topologies based on meta-paths and node attributes for predicting drug-related diseases. *International Journal of Molecular Sciences, 23*(7), 3870.

[22] Kundu, S. (2023). Citation polarity identification from scientific articles using deep learning methods (Master's thesis, The University of Western Ontario).

[23] Liang, X., Zhang, P., Yan, L., Fu, Y., Peng, F., Qu, L., Shao, M., Chen, Y., & Chen, Z. (2017). LRSSL: predict and interpret drug–disease associations based on data integration using sparse subspace learning. *Bioinformatics, 33*(8), 1187-1196.

[24] Filatova, A., Reveguk, I., Piatkova, M., Bessonova, D., Kuziakova, O., Demakova, V., Romanishin, A., Fishman, V., Imanmalik, Y., Chekanov, N., & Skitchenko, R. (2023). Annotation of uORFs in the OMIM genes allows to reveal pathogenic variants in 5′ UTRs. *Nucleic Acids Research, 51*(3), 1229-1244.

[25] Xu, S., Zhang, Y., Chen, L., & An, X. (2024). Is metadata of articles about COVID-19 enough for multilabel topic classification task?. *Database, 2024*, baae106.

[26] Mao, C., Yao, L., & Luo, Y. (2022). MedGCN: Medication recommendation and lab test imputation via graph convolutional networks. *Journal of Biomedical Informatics, 127*, 104000.

[27] Tabatabaei, A., Mahdavi, M. E., Beiranvand, E., Gharaghani, S., & Adibi, P. (2025, February 19). Graph neural network-based approaches to drug repurposing: A comprehensive survey. *SSRN*. https://doi.org/10.2139/ssrn.5154387

Notes on contributors

Saif Mohsin Najm I hold a Bachelor's and a Master's degree from Qom State University, Iran, and I am currently pursuing my doctoral studies in Information Technology at the same university.

Fatemeh Roohian is a Master's student in Computer Engineering at the University of Qom, Iran.

Amir Lakizadeh As an Assistant Professor at the University of Qom, Amir Lakizadeh applies his expertise in machine and deep learning to fields such as multimedia, vision, and medical diagnostics. His research has produced frameworks for realistic virtual try-on, real-time drowsiness detection, and vision models for image classification. He leads an active research team and has an established publication history in solving complex vision tasks with advanced neural networks.