Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 11, No. 3, 2023, 417-427.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

DcDiRNeSa, Drug Combination Prediction by Integrating Dimension Reduction and Negative Sampling Techniques

Mina Tabatabaei, Hossein Rahmani* and Motahareh Nasiri

School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran.

Article Info

Article History:
Received 26 March 2023
Revised 06 June 2023
Accepted 18 July 2023

DOI:10.22044/jadm.2023.12883.2437

Keywords:
Drug Combination Prediction, Synergistic Effect, Computational Techniques, Deep Learning.

**Corresponding author: h_rahmani@iust.ac.ir (H. Rahmani).*

Abstract

The search for effective treatments for complex diseases, while minimizing toxicity and side effects has become crucial. However, identifying synergistic combinations of drugs is often a time-consuming and expensive process, relying on trial and error due to the vast search space involved. Addressing this issue, we present a deep learning framework in this study. Our framework utilizes a diverse set of features including chemical structure, biomedical literature embedding, and biological network interaction data to predict potential synergistic combinations. Additionally, we employ autoencoders and principal component analysis (PCA) for dimension reduction in sparse data. Through 10-fold cross-validation, we achieved an impressive 98 percent area under the curve (AUC), surpassing the performance of seven previous state-of-the-art approaches by an average of 8%.

1. Introduction

Theoretically, combination therapy uses multiple drugs to treat a specific disease and shows better therapeutic efficacy than the sum of individual effects of the two drugs [1]. The use of combination therapy in the treatment of a variety of complex disorders such as hypertension, cancer, and infections [2]. The benefits of combination therapy over monotherapy such as increasing the therapeutic effect, decreasing the required dose per drug to avoid toxicity and the risk of adverse effects, and slowing the development of drug resistance have made this method a vital strategy [3]. As a result, determining the best drug combinations is a critical task with clinical and economic implications. It is unfeasible to test the complete combinatorial space due to the enormous search space. For example, the number of pairwise combinations that need to be examined across about 3000 human diseases and different dosage configurations is 499,500 for a sample of 1000 Food and Drug Administration (FDA)- approved drugs [4]. Also as the number of drugs rises, the search space extends quadratically [5].

Search algorithms were the most basic methods for predicting new drug combinations. In this category, Zinner *et al*. [6] used search algorithms with high convergence speed and no need for positive and negative samples of training data. However, it may reach a local optimum rather than a global optimum. Our proposed deep learning framework offers a more robust and comprehensive approach by incorporating a varied range of features.

Under the network analysis category, Chen *et al*. used the relationship between drug targets and disease proteins to evaluate drug combination efficiency for a disease. They concluded that drug-target module closeness in the protein-protein interaction network correlates with chemical and functional similarity. While the focus of their method was on the relationship between drug targets and disease proteins, our approach considers a broader range of features and captures a more holistic understanding of the intricate relationships between drugs which enables us to make more accurate predictions of synergistic effects, offering valuable insights for optimizing drug combination therapy.

Machine learning is the category with the most proposed approaches. Shi *et al*. [7] used a support vector machine (SVM) classifier on a two-layer structure. Bai *et al*. [8] used an upgraded Naïve Bayesian (NB) classifier on five distinct types of features, including drug targets, pathways, side-effects, metabolic enzymes, and drug transporters. Li *et al*. [9] compared the ensemble model with classic machine learning methods and determined that it outperformed them. Numerous further strategies have been proposed in this category.

By utilizing deep learning, our method can effectively capture complex relationships in drug data and provide more accurate and reliable predictions compared to ensemble models and traditional machine learning methods.

In this paper, we have considered interactions, side-effects, and indications of drugs and even their usage in biomedical texts available in PubMed [10]. To overcome the challenge of the imbalanced dataset, we have applied negative sampling and considered 2000 least similar drugs as negative samples. Additionally, various regularization methods including activity regularization, dropout, and early stopping have been used to eliminate overfitting caused by small datasets.

In this paper, we have made the following contributions:

- We have introduced a new dataset that gathered the chemical (molecular structure), biological (interactions and targets), phenotype (side effects and indications), and biomedical literature (PubMed) information of drugs together. To overcome the challenge of imbalanced dataset, we have applied negative sampling and considered 2000 least similar drugs based on the sum of 6 different feature similarities as negative samples alongside augmented positive samples, which have improved the performance compared to a random selection of negative samples.
- To reduce the effect of overfitting due to the small dataset in the field of synergistic drug combinations, we used a combination of dimensional reduction and negative sampling methods.

The rest of the paper is organized as what follows. Section 2 provides a literature assessment in the field of drug combinations prediction, Section 3 explains the methodology in further depth, Section 4 contains the empirical results of the DcDiRNeSa model and biological interpretation, and Section 5 discusses the conclusions and policy implications of the study.The source codes, datasets, and additional files used in this work are all available at:https://github.com/minta76/DcDiRNeSa.

2. Related Works

In drug combination therapy, more than one drug is used to treat complex diseases like HIV [11] and cancer [12]. As a result, computational methods have become increasingly important in predicting drug synergy. The Computational methods that are used to detect drug combinations can be divided into several categories, including search algorithms [6], network-based [4,13,14], machine learning [7-9,15-26], and deep learning methods [27-31].

The first computational method for dealing with drug combinations was a stochastic search algorithm known as the genetic algorithm [32]. The genetic algorithm, although a pioneering computational method for drug combination prediction, may suffer from converging to suboptimal solutions due to its stochastic nature and limited ability to capture complex relationships. Another technique for identifying drug combinations is network analysis, in which the interaction between components may uncover more information about the mechanisms of action of drugs that are either synergistic or antagonistic [33]. Guney *et al*. [13] proposed a drug-disease proximity metric that quantifies the connection between drug targets and diseases in order to discover the therapeutic effects of drugs and they concluded that drugs with close targets to disease can have a better effect than drugs with distant targets. Network analysis approaches often focus on capturing the interactions between components, such as drugs and target proteins, to uncover synergistic or antagonistic mechanisms. However, they may not fully capture the complex and multifaceted nature of drug combinations, which involve diverse factors beyond direct interactions. Chen *et al*. [14] constructed a two-layer heterogeneous network by integrating multiple data sources describing drugs, target proteins, and diseases in which nodes are drugs combination and diseases, and the edges are relations between the two layers. Using this network, they converted the drug combinations predictions problem to the link prediction problem solved by Regularized Least Squares (KRLS) [34] algorithm due to its easy implementation. While this technique offer simplicity in implementation, it may not effectively capture the intricate relationships and patterns that contribute to drug synergy.

Another type of computational method is one that employs machine learning to develop a prediction model and to learn underlying patterns in labeled input data, allowing for complex integration of

various data types. Three main ML approaches, including supervised [7-9,16,21-23,25,35], unsupervised [17,18,36], and semi-supervised learning [19,20] can be considered in the task of drug combination prediction.

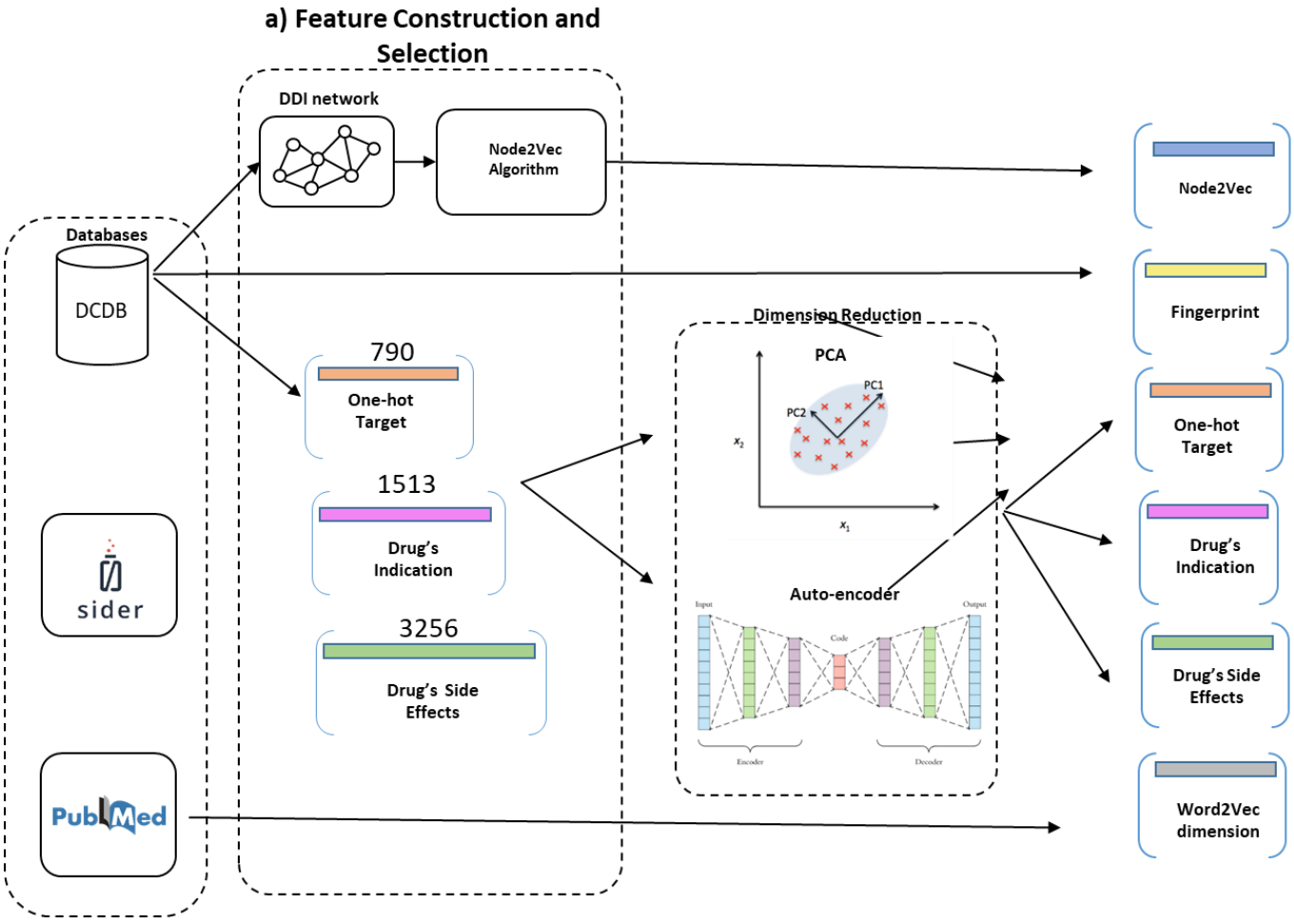
Machine learning-based approaches may focus on a subset of features, such as drug targets or pathways, neglecting the integration of other relevant information. This limited feature integration can lead to incomplete representations of drug combinations and potentially impact the accuracy of predictions. While these models can learn underlying patterns in labeled data, they may struggle with capturing complex relationships and interactions present in drug combinations.

Another category of drug combination prediction is the deep learning method which necessitates more training data, hyperparameters, computational resources, and memory because of its several processing layers [37]. The performance of deep learning models improves significantly as the amount of input data increases. Preuer *et al.* [27] outperformed four classic machine learning methods including SVM, RF, GBM (Gradient Boosting Machine), and EN (Elastic Nets), using a two-layer feedforward neural network that the input is a concatenated vector of chemical descriptors of both drugs and gene expression values of corresponding cancer cell lines.

To the best of our knowledge, none of the previous research has used a combination of features such as Node2vec (extracted from DDI interactions), word2vec (extracted from PubMed), drug-protein targets, indications, side effects, and drug fingerprints to investigate the relationship between the two drugs in predicting the synergistic effect of drug pairs. Also, the impact of employing dimensional reduction in sparse data in addition to the different negative sampling approaches to provide balanced training data on the prediction outcomes has not been explored yet.

3. Method

The main steps of the DcDiRNeSa model are illustrated in Figure 1; these steps are as follows:



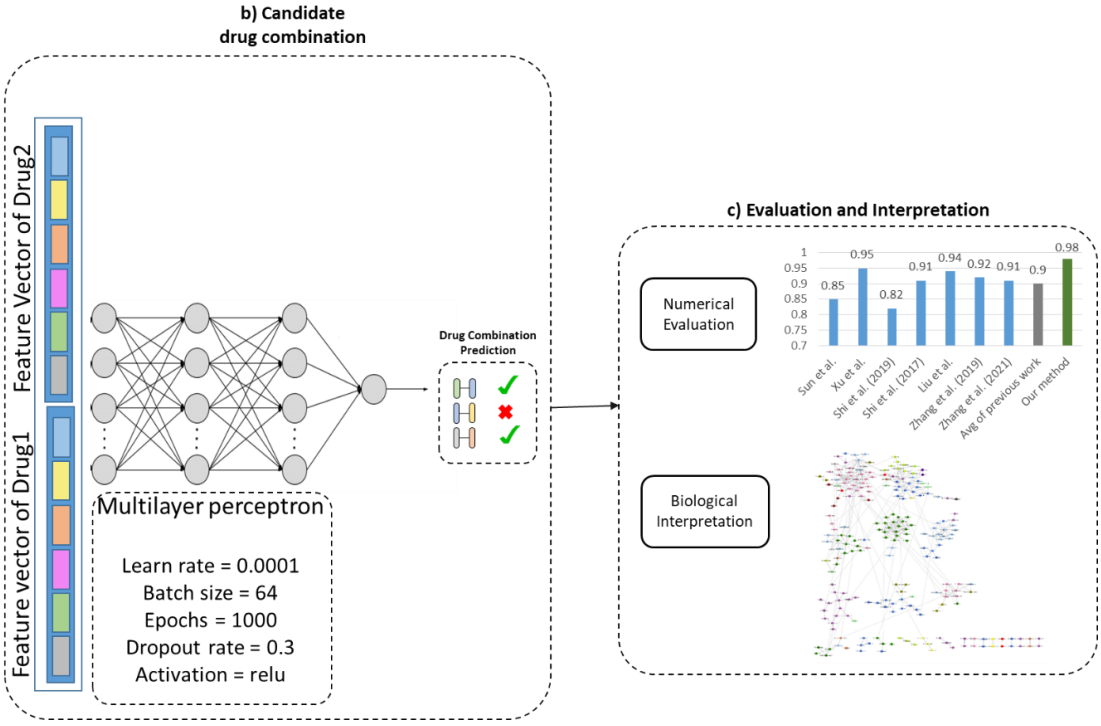


Figure 1. Overview of the DcDiRNeSa method with three main steps. Step a) Feature Construction and Selection: Feature extraction and dimension reduction techniques, followed by sampling methods to obtain a representative set of features (Section 3.1, 3.2, and 3.3). Step b) Candidate Drug Combination: Model is constructed using the candidate drug combinations generated from the previous step (Section 3.4). Step c) Evaluation and Interpretation: The numerical evaluation of the model's performance and the subsequent biological interpretation of the results (Section 4).

3.1. Databases and features construction

Drug combinations used in this article are taken from DCDB (Version 2.0) [38]. DCDB is collected from about 140,000 clinical studies and the Food and Drug Administration (FDA) Orange Electronic Book, which has a total of 1363 combinations.

Other drug features such as side-effects and indications are extracted from an external dataset known as SIDER [39]. Moreover, Word2Vec embedding of the drugs is extracted from PubMed. Chemical structure of drugs can be represented by SMILES (simplifying the molecular linear input specification) string [40]. A drug has a list of target proteins which is represented by a 790-dimension binary sparse vector showing the presence or absence of each protein. We have used a model constructed by Allahgholi *et al*. [41], which was based on the Word2vec word embedding technique [42]. We assigned a 3256-dimension binary sparse vector showing the presence or absence of side effects for each drug. Drug indications are represented in the same way, using a 1513-dimension binary sparse vector for each drug.

To create a positive sample, we extracted 295 drugs that have all the 6 mentioned features and participated in 329 positive combination drugs.

In creating negative samples, we used the similarity of all the drug pairs in the positive samples. The similarity of binary features, including chemical structure, drug targets, side effects, and indications are calculated based on the Tanimoto coefficient (also known as the Jaccard coefficient). While the similarity of Word2Vec and Node2Vec embedding of drugs which are numerical features are calculated using the cosine similarity measure. In the next step, we sorted the drug pairs based on the sum of the six obtained similarity scores, and after removing the positive samples from them, we considered the 2000 least similar drug pairs as negative samples.

The Tanimoto coefficient score is calculated using equation 1 and ranges from 0 to 1:

$$T = \frac{|F A \cap F B|}{|F A| + |F B| - |F A \cap F B|} \quad (1)$$

$F A$ is the number of chemical structures, targets, side effects, or indications related to drug A and $F B$ is the number of chemical structures, targets, side effects, or indications related to drug B, and $F A \cap F B$ is the number of common chemical structure, targets, side effects and indications for drug A and drug B.

3.2. Dimension reduction

Dimensionality reduction is used in this method to avoid overfitting and redundancy, which can be caused because of the curse of dimensionality phenomena. In addition, simplifying models improves the quality of human interpretations and reduces computational costs. We have applied two

different dimension reduction methods including PCA and autoencoder in order to identify the salient aspects of features with high dimensions and to make them easier to employ in a subsequent task. Using these two methods, the three features, including drugs target, side effects, and indications are reduced from 790, 3256, and 1513 to 200 dimensions, respectively.

3.2.1. Autoencoder

Three autoencoders (AE) [43] are used to reduce the dimension of the drug's targets, indications, and side effects. We have considered $x \epsilon Rd$ as the input with dimensions d and x' the reconstructed output. $h \epsilon Rt$ is the encoded feature representation with dimension t, which is smaller than d because the encoder is designed to produce low-dimensional and abstract features that can be used in different applications. In order to gain the input's hidden representation h, an encoder performs transformations f on the original input equation 2. A decoder, on the other hand, uses different transformations g to extract the reconstruction vector from the hidden representation h Equation 3. Back-propagation is used to update the parameters of f and g and reduce the reconstruction loss $L(x, x')$ in Equation 4 [44].

$$h = f\left(x\right) \tag{2}$$

$$x' = g\left(h\right) = g\left(f\left(x\right)\right) \tag{3}$$

$$L(x, x') = L(x, g\left(f\left(x\right)\right) \tag{4}$$

3.2.2. PCA

PCA is a method for converting multidimensional data into lower dimensions with minimal loss of information. PCA is based on a decomposition of the data matrix Y (nxm) into two matrices T and P Equation 5, where both matrices are orthogonal, plus a matrix of residual E. The matrix P (mxf) is usually called the loadings matrix, and the matrix T (nxf) is called the scores matrix, where f is the number of factors $f < m$.

$$Y = TP^T + E \tag{5}$$

As a condition of factorization optimality, the Euclidean norm of the residual matrix, $\|E\|$, must be reduced for the specified number of factors. PCA can be considered as a linear data mapping from Rm to Rf Equation 6, where y represents a row of Y as a single data, t represents the corresponding row of T, or the coordinates of y in the feature space, P are the coefficients for the linear transformation.

$$t = yP \tag{6}$$

To examine the information loss in this mapping, the measurement vector $y' = y - E$ can be reconstructed by reversing the projection back to Rm Equation 7 and measured by $|E|$ for individual measurement vectors or $\|E\|$ for the overall dataset.

$$y' = tP^T \tag{7}$$

3.3. Sampling

We concatenated the vector of the five specified features to produce a feature vector for each drug, and by using dimension reduction methods, the dimension of this vector was reduced from 6054 to 1095.

To represent each drug combination sample, we have concatenated vectors related to the participating drugs. Given that the model should be able to predict the synergistic combination regardless of the order of the drugs, each sample is used twice in the training set. The sequence of drug features is used once in (X, Y) and once in (Y , X) order. To deal with an imbalanced dataset that includes 329 positive samples and 43,365 possible negative samples, the undersampling as well as oversampling method is used. Therefore, we have selected a number of negative samples with two different methods. In one case, negative samples are picked up randomly from all the pairwise combinations of drugs. In the other case, a negative sampling method is used, and among all possible drug combinations, drugs with the least similarity to each other are considered negative samples.

3.4. Model

A deep learning approach is presented in this paper which uses a feedforward neural network composed of multiple processing layers for predicting the synergistic effect of drug combinations and comparing the results to other state-of-the-art machine learning methods. This neural network converts drug sample vectors into a single output value that indicates whether a drug combination is synergistic or not.

The forward propagation procedure starts with the inputs, and each neuron gets the linear combination of neural unit outputs from the previous layer as input and applies the activation function on it to get the output. With x as the input to layer i, W as the weight matrix, b as the bias vector, and f as the activation function, the i-th layer's output can be shown as Equation 8:

$$z_i = f\left(W_{i^x} + b_i\right) \tag{8}$$

After forward propagating an input example to the output layer, utilizing the specified loss function, the error is determined to minimize the difference between predicted and real values. Then the gradient of the loss function is calculated with respect to each of the weights of the network, and the backpropagation algorithm is applied to

propagate the error from the output to the input layer. As a result, each weight can be independently updated, allowing the loss function to be gradually reduced across numerous training iterations. Learning is accomplished by iteratively adjusting the weights using gradient algorithms, such as stochastic gradient descent (SGD) or modifications such as the Adam algorithm.

4. Empirical Results

In the following sub-sections, we first explain the hyper-parameters of our model in greater detail. Next, we'll go over the performance metrics that were utilized to evaluate the model, as well as the findings of the evaluation. The performance of our model is then compared to that of earlier state-of-the-art models. Finally, we evaluate some of the new predicted drug combinations with recent medical work and perform a network analysis based on Anatomical Therapeutic Chemical (ATC) codes of predicted synergistic drug combinations.

4.1. Method configuration

The feedforward neural network and autoencoder implementation details, as well as the values of the parameters, are discussed in this section.

4.1.1. Feedforward neural network

The feedforward neural network model is implemented using Keras, and to choose the best parameters grid search is used.The loss function is binary cross entropy and Adam is chosen as the optimization algorithm. To choose the best learning rate, $10-3$, $10-4$ and $10-5$ values are considered. To overcome overfitting, early-stopping, dropout, and activity regularizer are used as regularization techniques. For the input and the hidden layer's dropout rates of 0.3 and 0.5 were tested, and an early stopping strategy [45] was implemented with a patience of 20, which stops the training after 20 epochs if no improvement is seen. This method could help to avoid overfitting while also speeding up the training process.

The activity regularizer applies a penalty on the output of layers, which are summed into the loss function that the network optimizes. We have used L1 regularization with a factor of 0.01 in all layers. Besides the input and output layers, the remaining hidden layers have 128, 64, and 16 neural units, and all these three layers share the same activation function of 'relu' or 'selu'. In addition, we looked into different learning rates and regularization techniques. To achieve the optimum result, the selected hyperparameter space based on the best practices is summarized in Table 1.

Table 1. **Neural network parameters chose based on the best practices.**

Grid search hyper-parameter	Best hyper-parameters
Learn rate = [0.0001, 0.001, 0.01]	learn rate = 0.0001
batch size = [16, 32, 64]	batch size = 64
epochs = [10, 100, 1000]	epochs = 1000
dropout rate = [0.3, 0.5, 0.7]	dropout rate = 0.3
activation = ['selu', 'relu']	activation = 'relu'

4.1.2. Autoencoder

The encoder includes four hidden layers in addition to the input layer. For dimension reduction of side effects and indications features, the first layer contains 1,000 neural units, while target features have 600. The second and third layer contains 500 and 300, respectively. The encoded layer, which corresponds to the retrieved features with a length of 200, is the fourth layer. The decoder includes three hidden levels with unit numbers 300, 500, and 1000, respectively, as well as an output layer with the same unit number as the encoder's input layer. The autoencoder's loss function is Mean absolute, and the optimization technique is Adam, with a learning rate of 0.001.

4.2. Performance metrics

We used a 10-fold cross-validation method to evaluate performance. In our method, we employed accuracy as equation 9, recall (REC) as equation 10, precision (PRE) as Equation 11, F-measure as Equation 12, and the area under the receiver operating characteristic curve (AUC) as performance measurements

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

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$F_1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{12}$$

4.3. Model comparison

In this section, we have compared the influence of different methods on performance. In the first step, we considered 2000 negative samples, which are chosen in two different ways. In the first method, the negative drug combinations are random pairwise drugs picked up from positive sample drugs. The second method chooses the dug pairs which have the least similarity to each other based on the sum of 6 different feature similarities calculated using the Tanimoto coefficient score or cosine similarity. Figure 2 shows that selecting negative samples from less similar drugs could improve the AUC measure from 0.92 to 0.95.

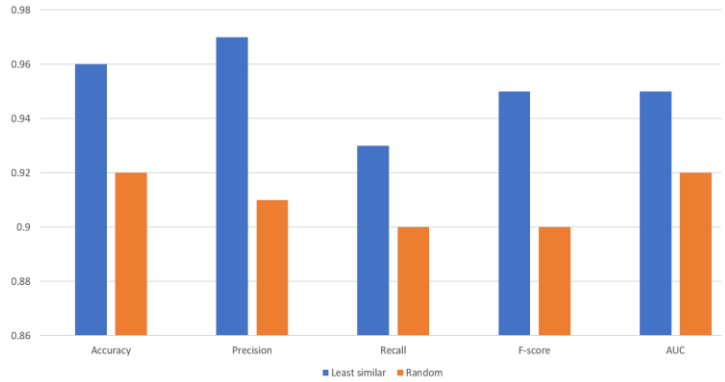


Figure 2. Negative sampling methods comparison. In the first method, the negative drug combinations are random pairwise drugs picked up from positive sample drugs. The second method chooses the dug pairs which have the least similarity to each other based on the sum of 6 different feature similarities.

Table 2 shows the influence of different negative sample sizes on performance for the second method. According to the table with 1316 positive data, the best result is obtained with 4000 negative data, although the improvement is not significant and in order to maintain the high speed of execution, the same 2000 negative samples have been used in later stages. In the next step, we decided to utilize autoencoder and PCA to minimize the dimension of three features, including side effects, indications, and targets. Utilizing these two methods, the dimension of the features was reduced to 200. However, we can see that using the autoencoder (green in Figure 3) produces better results than PCA.

Table 2. Performance with different negative sample sizes.

Negative sample size	ACC	AUC	F1	Rec	Pre
2000	0.96	0.95	0.95	0.93	0.95
4000	0.98	0.97	0.96	0.94	0.98
8000	0.99	0.96	0.95	0.92	0.99

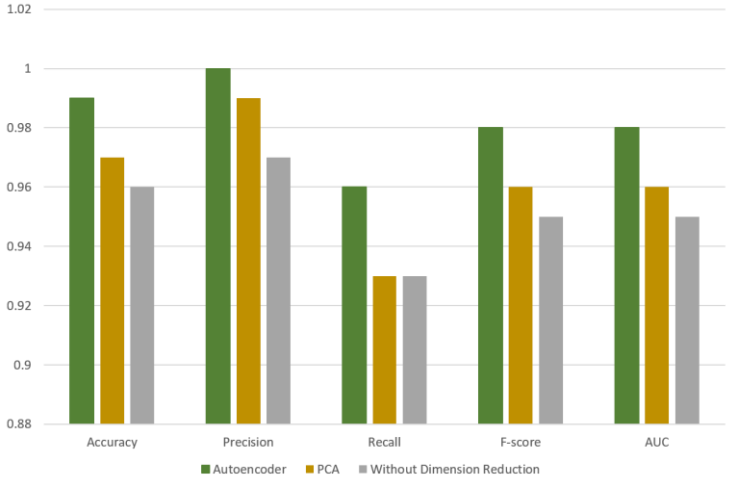


Figure 3. Dimensional reduction approaches significantly improved the results when compared to the unused conditions.

4.4. Comparison with related work

In this section, we compare the DcDiRNeSa method with 7 other models used to predict synergistic drug combinations in the DCDB dataset.

The comparison of validation results is shown in Figure 4. Compared to past techniques, the AUC score of our approach was an impressive 8 percent higher than the average. In addition, the method in this article is 7 percent better than the deep learning method provided by Zhang *et al*. (2021) [29]. Also compared to Xu *et al*. [46], it performs 3 percent better than the top-performing machine learning-based method.

Based on the paired t-test analysis, the obtained p-value of 0.0043 is smaller than the commonly used significance level of 0.05, indicating a statistically significant difference between our method's AUC score and the AUC scores of the other models. This result further strengthens the confidence in the superior performance of our approach.

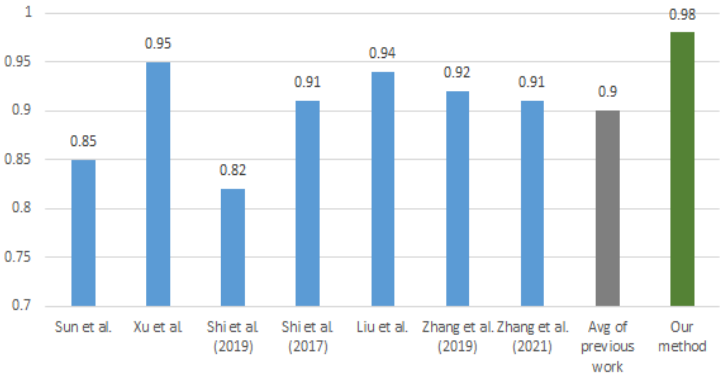


Figure 4. Methods result comparison. The AUC score of our approach is 8 percent higher than the average.

4.5. Biological interpretation

In the preceding section, we intend to examine the suggested model based on biological interpretation, in addition to the numerical evaluation completed. In the following, the concept called Anatomical Therapeutic Chemical (ATC) has been used for further biological interpretation. ATC is a system for classifying medications based on their therapeutic intent and chemical characteristics, as well as the organ or system on which they work. In this classification system maintained by the World Health Organization (WHO), one drug may have more than one code, and each code classified drugs into groups at five different levels. Only the first level of coding is used in this article, which specifies the anatomical main group and consists of one letter, with 14 main groups. Each of these groups has its own color, which can be seen in the top right corner of Figure 5.

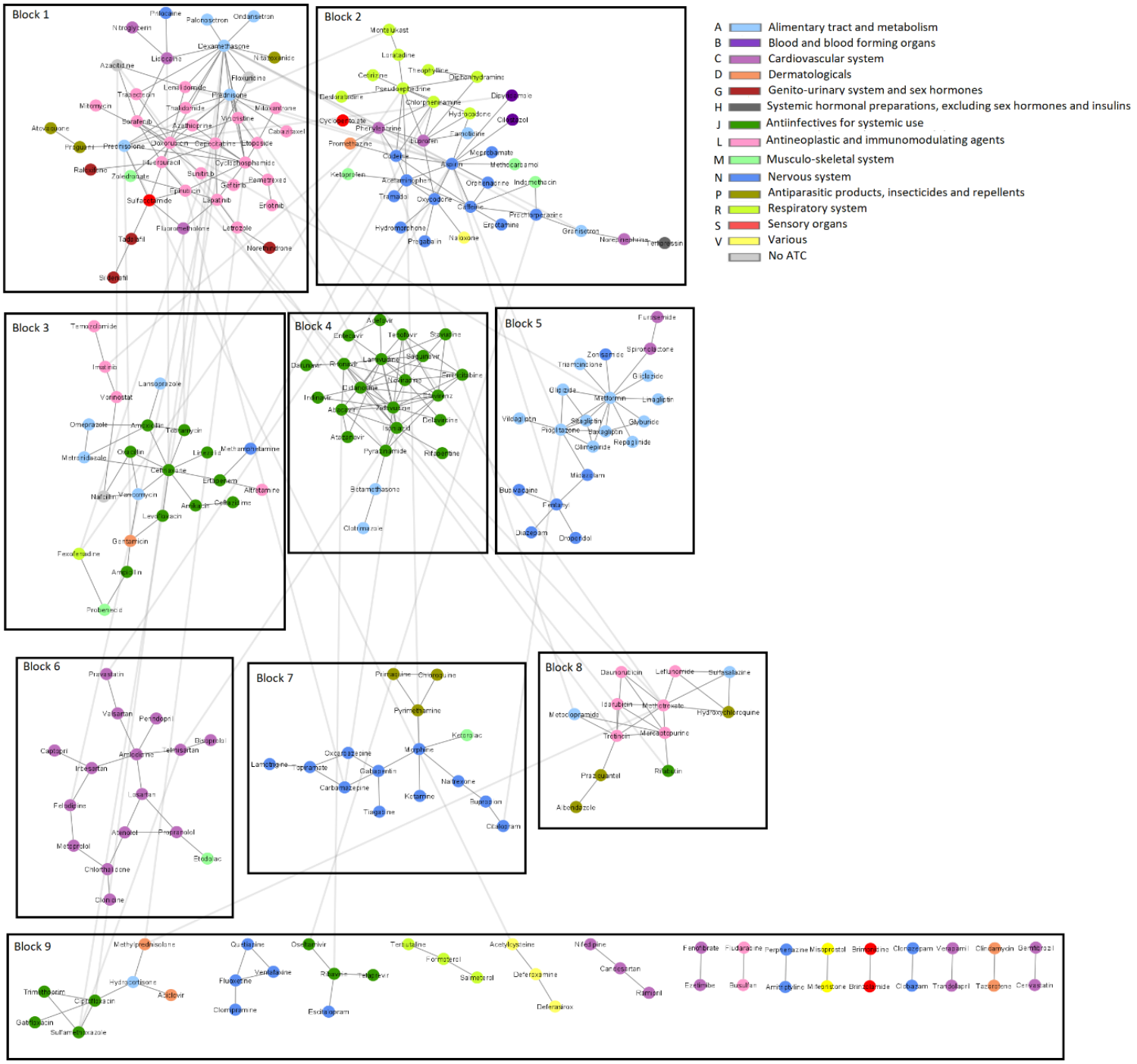


Figure 5. Predicted Drug Combinations Graph. The graph is constructed of new predicted drug combinations by the DcDiRNeSa method. A significant degree of homophily is shown when coloring the detected communities of the network based on Anatomical Therapeutic Chemical (ATC) codes. Further discussion in subsection 4.5.

In this graph, the nodes are drugs, and the edges are drug pairs that DcDiRNeSa predicts to be synergistic. In the next step, we applied the GLay algorithm [47] as a community structure analysis of biological networks to identify communities that are likely to share similar goals. In Figure 5, each block shows an identified community, and it can be seen that drugs with the same ATC codes are closer to each other than drugs with different ones. In block 1 most of the drugs including Mitoxantrone, Doxorubicin, Lenalidomide, Cyclophosphamide, and Methotrexate are used to treat certain types of cancer and their first level ATC code is "Antineoplastic and immunomodulating agents".

In block 2, there are two major types of medications: One for respiratory pharmaceuticals such as pseudoephedrine, chlorpheniramine, and hydrocodone, and another for nervous system

drugs such as aspirin, acetaminophen, caffeine, codeine, and oxycodone. Among some of the formed blocks, all drugs have almost the same first-order ATC code such as blocks 4 and 6, which correspond to antiinfectives for systemic use and the Cardiovascular system, respectively. Block 5 includes a group of drugs used to treat type 2 diabetes, and the first level ATC code is "alimentary tract and metabolism" such as Sitagliptin and Pioglitazone.

5. Conclusions and Future Work

The clinical and economic benefits of combination therapy, which employs numerous medications to treat a complex disease rather than monotherapy, have made it a critical strategy. Due to the exponential growth in the number of medications, it is necessary to create computational algorithms to minimize the search space for drug combinations. We introduced a synergistic drug combination prediction model named DcDiRNeSa in this research, which considers each drug combination as a set of features of two drugs. The features used include chemical structure, biological, biomedical literature embedding, and biological network interactions aspects of drugs extracted from various datasets. Then we applied a feedforward neural network with different overfitting handling methods on candidates and showed that dimension reduction on features and negative sampling on created imbalanced datasets could improve results considerably.

In terms of predicting drug combinations, our results show that DcDiRNeSa outperforms state-of-the-art techniques. In future work, using larger and more up-to-date datasets can achieve better results. Additionally, experimental validation can be used to improve machine Learning methods [48] by adjusting the computation model in order to gain better prediction results.

References

[1] D. Chen, X. Liu, Y. Yang, H. Yang, P. Lu, "Systematic synergy modeling: understanding drug synergy from a systems biology perspective," *BMC systems biology, 2015*, val. 9, no. 1, pp. 1-10.

[2] B. He, C. Lu, G. Zheng, X. He, M. Wang, G. Chen, G. Zhang, A. Lu, "Combination therapeutics in complex diseases," *Journal of Cellular and Molecular Medicine, 2016*, val. 20, no. 12, pp. 2231-2240.

[3] T. C. Chou, "Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies," *Pharmacological reviews, 2006*, val. 58, no. 3, pp. 621-681.

[4] F. Cheng, I. A. Kovács, A. L. Barabási, "Network-based prediction of drug combinations," *Nature communications, 2019*, val. 10, no. 1, p. 1197.

[5] L. Wu, Y. Wen, D. Leng, Q. Zhang, C. Dai, Z. Wang, Z. Liu, B. Yan, Y. Zhang, J. Wang, S. He, "Machine learning methods, databases and tools for drug combination prediction," *Briefings in bioinformatics, 2022*, val. 23, no. 1, p. bbab355.

[6] R. G. Zinner, B. L. Barrett, E. Popova, P. Damien, A. Y. Volgin, J. G. Gelovani, R. Lotan, H. T. Tran, C. Pisano, G. B. Mills, L. Mao, "Algorithmic guided screening of drug combinations of arbitrary size for activity against cancer cells," *Molecular cancer therapeutics, 2009*, val. 8, no. 3, pp .521-532.

[7] J. Y. Shi, J. X. Li, K. T. Mao, J. B. Cao, P. Lei, H. M. Lu, S. M. Yiu, "Predicting combinative drug pairs via multiple classifier system with positive samples only," *Computer methods and programs in biomedicine, 2019*, pp. 1-10.

[8] L. Y. Bai, H. Dai, Q. Xu, M. Junaid, S. L. Peng, X. Zhu, Y. Xiong, D. Q. Wei, "Prediction of effective drug combinations by an improved naïve bayesian algorithm," *International journal of molecular sciences, 2018*, val. 19, no. 2, p. 467.

[9] J. Li, X. Y. Tong, L. D. Zhu, H. Y. Zhang, "A machine learning method for drug combination prediction," *Frontiers in Genetics, 2020*, p. 1000.

[10] "PubMed," [Online]. Available: https://ftp.ncbi.nlm.nih.gov/pubmed/baseline.

[11] K. Traynor, FDA approves four-drug anti-HIV combination tablet, Oxford University Press, 2012.

[12] B. Al-Lazikani, U. Banerji, P. Workman, "Combinatorial drug therapy for cancer in the post-genomic era," *Nature biotechnology, 2012*, val. 30, no. 7, pp. 679-692.

[13] E. Guney, J. Menche, M. Vidal, A. L. Barábasi, "Network-based in silico drug efficacy screening," *Nature communications, 2016*, val. 7, no. 1, p. 10331.

[14] H. Chen, S. K. Iyengar, J. Li, "Large-scale analysis of drug combinations by integrating multiple heterogeneous information networks," in *InProceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, 2019*, pp. 67-76.

[15] T. Hastie, R. Tibshirani, J. H. Friedman, J. H. Friedman, The elements of statistical learning: data mining, inference, and prediction, 2009, springer.

[16] V. Y. Tan, C. Févotte, "Tan VY, Févotte C. Automatic relevance determination in nonnegative matrix factorization with the/spl beta/-divergence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012.

[17] L. Huang, F. Li, J. Sheng, X. Xia, J. Ma, M. Zhan, S. T. Wong, "DrugComboRanker: drug combination

discovery based on target network analysis," *Bioinformatics, 2014*, val. 30, no. 12, pp. i228-236.

[18] D. Chen, H. Zhang, P. Lu, X. Liu, H. Cao, "Synergy evaluation by a pathway--pathway interaction network: a new way to predict drug combination," *Molecular BioSystems, 2016*, val. 12, no. 2, pp. 614-623.

[19] Y. Sun, Z. Sheng, C. Ma, K. Tang, R. Zhu, Z. Wu, R. Shen, J. Feng, D. Wu, D. Huang, D. Huang, "Combining genomic and network characteristics for extended capability in predicting synergistic drugs for cancer," *Nature communications, 2015*, val. 6, no. 1, p. 8481.

[20] X. Chen, B. Ren, M. Chen, Q. Wang, L. Zhang, G. Yan, "NLLSS: predicting synergistic drug combinations based on semi-supervised learning," *PLoS computational biology, 2016*, val. 12, no. 7, p. e1004975.

[21] H. Liu, W. Zhang, L. Nie, X. Ding, J. Luo, L. Zou, "Predicting effective drug combinations using gradient tree boosting based on features extracted from drug-protein heterogeneous network," *BMC bioinformatics, 2019*, val. 20, no. 1, pp. 1-2.

[22] P. Sidorov, S. Naulaerts, J. Ariey-Bonnet, E. Pasquier, P. J. Ballester, "Predicting synergism of cancer drug combinations using NCI-ALMANAC data," *Frontiers in chemistry, 2019*, p. 509.

[23] C. Zhang, G. Yan, "Synergistic drug combinations prediction by integrating pharmacological data," *Synthetic and systems biotechnology, 2019*, val. 4, no. 1, pp. 67-72.

[24] Y. Juan, Y. Zhuang, W. S. Chin, C. J. Lin, "Field-aware factorization machines for CTR prediction," in *InProceedings of the 10th ACM conference on recommender systems, 2016*, pp.43-50.

[25] H. Julkunen, A. Cichonska, P. Gautam, S. Szedmak, J. Douat, T. Pahikkala, T. Aittokallio, J. Rousu, "comboFM: leveraging multi-way interactions for systematic prediction of drug combination effects," *bioRxiv, 2020*, pp. 2020-2029.

[26] A. Ianevski, A. K. Giri, P. Gautam, A. Kononov, S. Potdar, J. Saarela, K. Wennerberg, T. Aittokallio, "Prediction of drug combination effects with a minimal set of experiments," *Nature machine intelligence, 2019*, val. 1, no. 12, pp. 568-577.

[27] K. Preuer, R. P. Lewis, S. Hochreiter, A. Bender, K. C. Bulusu, G. Klambauer, "DeepSynergy: predicting anti-cancer drug synergy with Deep Learning," *Bioinformatics, 2018*, val. 34, no. 9, pp. 1538-1546.

[28] F. Xia, M. Shukla, T. Brettin, C. Garcia-Cardona, J. Cohn, .J. E. Allen, S. Maslov, S. L. Holbeck, J. H. Doroshow, Y. A. Evrard, E. A. Stahlberg, "Predicting tumor cell line response to drug pairs with deep learning," *BMC bioinformatics, 2018*, pp.71-79.

[29] T. Zhang, L. Zhang, .P. R. Payne, F. Li, "Synergistic drug combination prediction by integrating multiomics data in deep learning models," in *InTranslational bioinformatics for therapeutic development, 2021*, pp. 223-238.

[30] H. I. Kuru, O. Tastan, A. E. Cicek, "MatchMaker: a deep learning framework for drug synergy prediction," *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2021*, val. 19, no. 4, pp. 2334-2344.

[31] P. Jiang, S. Huang, Z. Fu, Z. Sun, T. M. Lakowski, P. Hu, "Deep graph embedding for prioritizing synergistic anticancer drug combinations," *Computational and structural biotechnology journal, 2020*, pp. 427-438.

[32] S. Forrest, "Genetic algorithms: principles of natural selection applied to computation," *Science, 1993*, val. 261, no. 5123, pp. 872-878.

[33] N. Yin, W. Ma, J. Pei, Q. Ouyang, C. Tang, L. Lai, "Synergistic and antagonistic drug combinations depend on network topology," *PloS one, 2014*, val. 9, no. 4, p. e93960.

[34] A. C. Nascimento, R. B. Prudêncio, I. G. Costa, "A multiple kernel learning algorithm for drug-target interaction prediction," *BMC bioinformatics, 2016*, pp. 1-6.

[35] T. Chen, C. Guestrin, "Xgboost: A scalable tree boosting system," in *InProceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016*, pp. 785-794.

[36] Z. Wu, X. M. Zhao, L. Chen, "A systems biology approach to identify effective cocktail drugs," in *InBMC systems biology, 2010*, Sep (Vol. 4, pp. 1-14). BioMed Central.

[37] I. F. Tsigelny, "Artificial intelligence in drug combination therapy," *Briefings in bioinformatics, 2019*, val. 20, no. 4, pp. 1434-1448.

[38] Y. Liu, Q. Wei, G. Yu, W. Gai, Y. Li, X. Chen, "DCDB 2.0: a major update of the drug combination database," *Database, 2014*, Jan 1;2014.

[39] M. Kuhn, I. Letunic, L. J. Jensen, P. Bork, "The SIDER database of drugs and side effects," *Nucleic acids research, 2016*, val. 44, no. D1, p. D1075-9.

[40] D. Weininger, "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules," *Journal of chemical information and computer sciences, 1988*, val. 28, no. 1, pp. 31-36.

[41] M. Allahgholi, H. Rahmani, D. Javdani , G. Weiss, D. Módos, "ADDI: Recommending alternatives for drug--drug interactions with negative health effects," *Computers in Biology and Medicine, 2020*, p. 103969.

[42] T. Mikolov, K. Chen, G. Corrado, L. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781, 2013*, Jan 16.

[43] M. A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks," *AIChE Journal, 1991*, val. 37, no. 2, pp. 233-243.

[44] M. Li, Y. Wang, R. Zheng, X. Shi, Y. Li, F. X. Wu, J. Wang, "DeepDSC: a deep learning method to predict drug sensitivity of cancer cell lines," *IEEE/ACM transactions on computational biology and bioinformatics, 2019,* val. 18, no. 2, pp. 575-582.

[45] L. Prechelt, "Automatic early stopping using cross validation: quantifying the criteria," *Neural Networks, 1998,* val. 11, no. 4, pp. 761-767.

[46] Q. Xu, Y. Xiong, H. Dai, K. M. Kumari, Q. Xu, H. Y. Ou, D. Q. Wei, "PDC-SGB: Prediction of effective drug combinations using a stochastic gradient boosting algorithm," *Journal of theoretical biology, 2017,* pp. 1-7.

[47] G. Su, A. Kuchinsky, J. H. Morris, D. J. States, F. Meng, "GLay: community structure analysis of biological networks," *Bioinformatics, 2010,* val. 26, no. 24, pp. 3135-3137.

[48] O. Oladimeji, O. Oladimeji, "Detecting Breast Cancer through Blood Analysis Data using Classification Algorithms," *Journal of AI and Data Mining, 2021,* val. 9, no. 3, pp. 351-359.

DcDiRNeSa: پیش‌بینی ترکیب دارو با ادغام روش‌های کاهش ابعاد و نمونه‌گیری منفی

مینا طباطبایی، حسین رحمانی* و مطهره نصیری

دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران.

ارسال ۲۰۲۳/۰۳/۲۶؛ بازنگری ۲۰۲۳/۰۶/۰۶؛ پذیرش ۲۰۲۳/۰۷/۱۸

چکیده:

جستجو برای درمان‌های مؤثر برای بیماری‌های پیچیده، درحالی‌که سمیت و عوارض جانبی را به حداقل می‌رساند، بسیار مهم شده است. بااین‌حال، شناسایی ترکیب‌های هم‌افزایی داروها اغلب فرآیندی زمان‌بر و پرهزینه است و به دلیل فضای جستجوی وسیعی که درگیر است، بر آزمون‌وخطا تکیه می‌کند. با پرداختن به این موضوع، ما یک چارچوب یادگیری عمیق را در این مطالعه ارائه می‌کنیم. چارچوب ما از مجموعه متنوعی از ویژگی‌ها، ازجمله ساختار شیمیایی، تعبیه ادبیات زیست پزشکی، و داده‌های تعامل شبکه بیولوژیکی برای پیش‌بینی ترکیب‌های هم‌افزایی بالقوه استفاده می‌کند. علاوه بر این، ما از رمزگذارهای خودکار و تجزیه‌وتحلیل اجزای اصلی (PCA) برای کاهش ابعاد در داده‌های پراکنده استفاده می‌کنیم. از طریق اعتبارسنجی متقاطع ۱۰ برابری، ما به سطح چشمگیر ۹۸ درصدی زیر منحنی (AUC) دست یافتیم که به‌طور متوسط ۸ درصد از عملکرد هفت رویکرد پیشرفته قبلی پیشی گرفت.

کلمات کلیدی: پیش‌بینی ترکیبی از داروها، اثر هم‌افزایی، تکنیک‌های محاسباتی، یادگیری عمیق.