

Research Article

Identification of Drug-Disease Associations Using a Random Walk with Restart Method and Supervised Learning

Xiaoqing Liu,¹ Wenjing Yi,² Baohang Xi,² and Qi Dai ²

¹College of Sciences, Hangzhou Dianzi University, Hangzhou 310018, China

²College of Life Sciences, Zhejiang Sci-Tech University, Hangzhou 310018, China

Correspondence should be addressed to Qi Dai; daiailiu04@yahoo.com

Received 24 July 2022; Accepted 23 September 2022; Published 10 October 2022

Academic Editor: Lin Lu

Copyright © 2022 Xiaoqing Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Drug-disease correlations play an important role in revealing the mechanism of disease, finding new indications of available drugs, or drug repositioning. A variety of computational approaches were proposed to find drug-disease correlations and achieve good performances. However, these methods used a variety of network information, but integrated networks were rarely used. In addition, the role of known drug-disease association data has not been fully played. In this work, we designed a combination algorithm of random walk and supervised learning to find the drug-disease correlations. We used an integrated network to update the model and selected a gene set as the start of random walk based on the known drug-disease correlations data. The experimental results show that the proposed method can effectively find the correlation between drugs and diseases, and the prediction accuracy is 82.7%. We found that there are 8 pairs of drug-disease relationships that have not yet been reported, and 5 of them have pharmacodynamic effects on Parkinson's disease. We also found that a key linkage between Parkinson's disease and phenylhexol, a drug for the treatment of Parkinson's disease α -synuclein and tau protein, provides a useful exploration for the effectiveness of the treatment of Parkinson's disease.

1. Introduction

With the prevalence of complex diseases, the existing drugs are far from meeting the needs of human beings to fight against diseases. At the same time, due to the rising cost of drug research and development, long research and development cycle, large difference in research and development success rate, and high loss rate of new drugs, the research and development of innovative drugs has become a major challenge in the medical field.

At present, reusing compounds that have reduced risk to treat common or rare diseases has become a popular means of drug research and development. This strategy is called drug repositioning or drug reuse. This method not only reduces the overall development cost but also shortens a large amount of research and development time [1–3]. Through drug repositioning, pharmaceutical companies have achieved many successes, such as Pfizer's Viagra for erectile dysfunction [4] and Celgene's thalidomide for severe nodular leprosy erythema [5].

With the rapid expansion of large-scale genome, transcriptome and proteome data, computational drug repositioning study has emerged as one of the leading methods. Huang et al. developed a new drug repositioning pipeline to analyze four lung cancer microarray datasets, enrich biological processes, potential therapeutic drugs, and target genes for the treatment of non-small-cell lung cancer (NSCLC) [6]. They integrated two methods: machine learning algorithm and classification based on topological parameters. Zheng et al. designed a weighted ensemble similarity (WES) algorithm which provides a new perspective for drug repositioning and discovery [7]. Wang et al. integrated two drug transfer methods and proposed a new method for drug repositioning [8]. Cheng et al. [9] integrated the integration of chemical, gene, and disease networks, inferred the chemical hazard profile, studied the exposure data gap, and fully considered the gene and disease network in the chemical safety assessment [10]. A large number of genetic and molecular biology studies have shown that diseases reflect the interaction of multiple molecular components on a

certain level [11–14]. Therefore, drug repositioning study should consider the interaction between different disease-related genes [15–18]. Luo et al. found the potential indications of a given drug based on some comprehensive similarity measures and Bi-Random walk (BiRW) [19]. Yu et al. inferred the correlation between drugs and diseases by studying the characteristics of known protein complexes [20]. PREDICT (PREdicting Drug IndiCaTions) considers that similar drugs are suitable for similar diseases; the prediction task is achieved by designing similarity measures between multiple drugs and diseases [21].

The above method was successfully applied to drug-disease association study and achieved good performance. However, these methods have used a variety of network information, but the integrated network is still less used. With the increase of the related data of known drug diseases, a supervised learning method should be designed to further improve the drug-disease association research by using the related data of known drug diseases. In this paper, we used an integrated network consisting of HPRD, BioGRID, STRING, and other databases. Unlike previous network-based studies, which used the random walk method with restart on the network, we updated the model using the known data of the relationship between drugs and diseases and selected a gene set as the starting point of random walk, thus realizing the supervised learning of random walk with restart method. We also evaluated the performance of the proposed methods in various diseases and analyzed their GO and KEGG function enrichment.

2. Datasets and Methods

2.1. Protein–Protein Interaction (PPI) Network. Human protein–protein interaction (PPI) network is selected, which has been compiled by Jörg et al. that contained experimentally documented human physical interactions from TRANSFAC, IntAct39, MINT40, BioGRID41, HPRD42, KEGG43, BIGG44, CORUM45, PhosphoSitePlus46, and a large scale signaling network47. We used the largest connected component of the interaction in our analysis, consisting of 141,150 interactions between 13,329 proteins. Entrez Gene IDs were used to map disease-associated genes to the corresponding proteins in the interaction. The interaction and disease-gene association data is provided as a supplementary data set in Menche et al. [22]

2.2. Disease and Disease-Genes Data. Medical Subject Headings (MeSH) is an authoritative thesaurus compiled by the National Medical Library of the United States [23]. The disease subject words in the vocabulary provided by MeSH have perfect vocabulary classification. Our disease data and drug data are derived from Menche et al. [22], which integrate some genetic disease-related genes from the human Mendelian inheritance in man (OMIM: Online Mendelian Inheritance in Man) and trait gene association data from GWAS central. Through the medical topic title Ontology (MeSH) [24], the disease names of different disease nomenclature are combined into one name.

We screened diseases containing at least 20 disease-related genes from 1489 diseases in MeSH. In this paper, we considered at least 20 disease-related genes in order to understand the role of related genes in the interaction network, rather than the occurrence of disease due to the mutation of a gene. Finally, 299 diseases and their 3173 genes were obtained. In the process of disease screening, we required at least one drug for each disease. By searching the DrugBank database, the drug information that can treat 79 diseases corresponding to FDA approval is obtained, and Metab2Mesh is used for text mining [25]. If the text mining results indicate that there is a strong correlation between disease and drug, we added the relationship between the drug and disease into the known data set.

2.3. Drug and Drug-Target Data. DrugBank is a comprehensive drug information database, which not only includes the information of drug structure, drug target, and drug action mechanism but also integrates the information of drug experiment and clinical research. DrugBank has strong retrievable ability, coupled with its convenient web visualization function, which provides researchers with powerful convenience in drug research and development, drug mechanism exploration, and so on. DrugBank 5.0 contains information about 10971 drugs and 4900 protein targets, including 2391 FDA approved small molecule drugs, 934 approved biotechnology drugs, 109 nutritional drugs, and more than 5090 experimental drugs. We collected the drug and drug-target information certified by the FDA from DrugBank, and then searched for the strong literature evidence of drug-early-warning-disease association through Metab2Mesh, and finally obtained 238 drugs that can treat corresponding diseases.

2.4. Random Walk with Restart Method. PPI network can be expressed as $G = (V, E)$, where V denotes protein and E stands for protein–protein interaction. The $n * n$ adjacency matrix A is used to represent the PPI network, where n is the total number of the proteins. If there is interaction between protein i and protein j , $A_{[i,j]}$ is 1, otherwise it is 0. We then normalized the adjacency matrix A :

$$A'_{[i,j]} = \frac{A_{[i,j]}}{\sum_{k=1}^n A_{[k,j]}}. \quad (1)$$

Random walk is used to find potential gene association data of diseases or drugs. When the random walk converges, the probability of a disease or drug at each point of the PPI network can be obtained. The relationship between drugs and diseases can be calculated based on the correlation between the probability distribution of diseases and drugs.

Random walk starts with a set of seed genes. The initial vector of seed genes is defined as follows:

$$P_0 = [\psi_1, \psi_2, \dots, \psi_n]^T. \quad (2)$$

For a disease, we listed all the drugs that can treat it, incorporate all the genes of these drugs into the relevant genes of the disease, and took the combined gene set as the

seed gene of the disease. Among them, the genes directly related to the disease are defined as

$$P_{\text{dis}_{\text{dir}}} = \left[\psi_{\text{dis}_{\text{dir}_1}}, \psi_{\text{dis}_{\text{dir}_2}}, \dots, \psi_{\text{dis}_{\text{dir}_n}} \right]^T, \quad (3)$$

where the disease-related genes $\psi_{\text{dis}_{\text{dir}_i}}$ will be set to 1, otherwise it will be set to 0. Then $P_{\text{dis}_{\text{dir}}}$ is normalized as

$$P'_{\text{dis}_{\text{dir}_k}} = \frac{P_{\text{dis}_{\text{dir}_k}}}{\sum_{k=1}^n P_{\text{dis}_{\text{dir}_k}}}. \quad (4)$$

Suppose there are m drugs that can treat the same disease, they are represented as $P_{\text{dis}_{\text{drug}_1}}, P_{\text{dis}_{\text{drug}_2}}, \dots$, and $P_{\text{dis}_{\text{drug}_m}}$

$$P_{\text{dis}_{\text{drug}_m}} = \left[\psi_{\text{dis}_{\text{drug}_{m_1}}}, \psi_{\text{dis}_{\text{drug}_{m_2}}}, \dots, \psi_{\text{dis}_{\text{drug}_{m_n}}} \right]^T. \quad (5)$$

Sum all drugs for a disease:

$$P_{\text{dis}_{\text{drug}}} = \sum_{k=1}^m P_{\text{dis}_{\text{drug}_k}}. \quad (6)$$

Then we normalized $P_{\text{dis}_{\text{drug}}}$ as

$$P'_{\text{dis}_{\text{drug}_k}} = \frac{P_{\text{dis}_{\text{drug}_k}}}{\sum_{k=1}^m P_{\text{dis}_{\text{drug}_k}}}. \quad (7)$$

Finally we got its seed gene for a given disease,

$$P_{\text{disease}} = tP'_{\text{dis}_{\text{dir}}} + (1-t)P'_{\text{dis}_{\text{drug}}}, \quad (8)$$

where t is 0.5.

We also got the seed gene P_{drug} of a given drug following the same method. Start random walk and randomly access adjacent genes in each time scale ($t \rightarrow t+1$), the state probability P_{t+1} at time $t+1$ is

$$P_{t+1} = (1-r)A'P_t + rP_0, \quad (9)$$

where P_0 is the initial vector, P_t is the probabilities at time t , and r is the restart probability. If the difference between P_t and P_{t+1} is less than $1e-6$, it is considered that the process will reach a stable state. After reaching the stable state, the correlation between drugs and diseases, drugs and drugs, and diseases and diseases is calculated according to the probability of drugs and diseases accessing each node on the network.

2.5. Supervised Learning. Cross-validation is a frequently used model validation technology. It divides the known data into two subsets, adds the data of one subset to the model training, and verifies the model with the remaining subset to evaluate the performance of the model in unknown data. For example, when using k -fold cross validation, the known

data set needs to be randomly divided into k parts. In each cross-validation, $k-1$ data is selected to be added to the model training, and the remaining data is used for validation. Repeat for k times and select one piece of data for verification each time until each piece of data is tested.

The goal of cross-validation is to test the prediction ability of the model in new data, and it can also reflect the problem of overfitting or selection bias. In this paper, the idea of this method is used for supervised learning of random walk. For a certain disease, all drugs that can treat the disease in the data set are listed, and the genes associated with these drugs are incorporated into the relevant genes of the disease, and the combined gene set is used as the start of random walk. Needles are treated in the same way as drugs. In this paper, 403 known drug-disease associations between 78 diseases and 238 drugs were randomly divided into 10 parts. Nine of the disease and drug association data were selected to update the model, and the updated model was used to process the other data, so as to achieve the effect of supervised learning.

2.6. Evaluation Method. Receiver operating characteristic (ROC) curve is a curve based on the true positive rate (TPR) and false positive ratio (FPR) under various threshold settings. Area under the curve of ROC, also known as AUC value, can well reflect the performance of the classifier. The value of AUC varies between 0 and 1. When the AUC value is equal to 0.5, it means that the classifier cannot work. The larger the AUC value, the better the performance of the classifier. When the AUC value is 1, the classifier can produce perfect results.

3. Results

3.1. Performance Evaluation of the Random Walk with Restart Method Based on Supervised Learning. In order to evaluate the effectiveness of the proposed method, we first took the known drug-disease association as an independent validation data set. According to the relevant genes of 78 diseases and the drug targets of 238 drugs, the correlation information between diseases and drugs was obtained through restart random walk on PPI network. According to the ranking of drug-disease information pair by correlation, the AUC value was calculated. Three PPI networks BioGrid, HPRD, and STRING were independently verified, and their AUC results were 0.64, 0.52, and 0.66.

In order to further explore the efficiency of methods in different diseases, MeSH was used to classify all diseases. There are also some diseases in the classification that belong to a variety of disease types, such as colorectal tumors, which belong to C04 tumor diseases and C06 digestive system diseases. For the above case, we only calculated the average AUC value according to one of them. The AUC value was calculated on the basis of PPI network and optimal parameters. The classification results of various diseases are shown in Table 1.

From Table 1, it is easy to note that the performance of the random walk with the restart method is different among various diseases. It achieves good performance in the diseases of blood and lymphatic system C15, endocrine system diseases C19, eye diseases C11, and Male genitourinary

TABLE 1: Average AUC value of various diseases based on the random walk with restart method and three PPI networks BioGrid, HPRD, and STRING.

Disease classification based on MeSH	Number of diseases	Average AUC
Viral diseases C02	2	0.70467
Tumor C04	16	0.764632
Musculoskeletal diseases C05	5	0.748488
Digestive system diseases C06	9	0.760698
Respiratory diseases C08	2	0.68288
Nervous system diseases C10	9	0.62232
Eye diseases C11	2	0.83772
Male genitourinary system C12	1	0.81407
Cardiovascular disease C14	11	0.66685
Blood and lymphatic system C15	4	0.876637
Skin and connective tissue diseases C17	5	0.675556
Nutritional and metabolic diseases C18	6	0.734407
Endocrine system diseases C19	2	0.87446
Immune system diseases C20	4	0.62869

system C12, with AUC values above 0.8. The highest is blood and lymphatic system C15, with an AUC value of 0.877. The AUC value of nervous system diseases is low, only 0.62.

In order to further verify the efficiency of the random walk with the restart method and supervised learning, we randomly divided all known drug-disease relationships into ten parts, nine pieces of data are used as the training set and the other is used to calculate AUC. For a certain disease, we listed all the drugs that can treat the disease in the known training set, and then integrated all the related genes of these drugs into the related genes of the disease, and took the combined gene set as the start of random walk [26]. For drugs, the same method is used; that is, the relevant genes of diseases that can be treated by a drug in the training set were combined into the target information of the drug. Ten AUC values were obtained for each experiment. In order to reduce random interference, the above experiment was repeated 10 times, and a total of 100 AUC values were obtained, as shown in Figure 1.

The results show that the average value of 100 AUC values is 0.827, indicating that the proposed method found the relationship between drugs and diseases. With the help of the training data of the known network relationship between drugs and diseases, the prediction sensitivity of drugs and diseases was further improved. Adding the target information of drugs that can treat a disease to the disease-gene information will indirectly add some potential disease-gene information, making the disease-gene information more abundant. Similarly, adding the genes of all diseases that can be treated by a drug to the target information of drugs can also enrich the information of drug-action targets and make the relationship between drugs and dis-

eases more discovered, thus improving the prediction of drug sensitivity.

3.2. Analysis of the Relationship between Drugs and Diseases.

In this work, disease-related genes were taken as the starting point of random walk on one side, and the target genes of drugs were taken as the starting point of random walk on the other side. Through the restart random walk on the whole PPI network, the relationship between each disease and each drug on the PPI network was obtained, and their correlation coefficient was further calculated. We got 18564 group correlations of 78 diseases and 238 drugs. According to their correlation coefficients, 61 pairs of disease drugs with a correlation degree of more than 0.8 are found, of which 53 diseases and drugs have been confirmed by research, and 8 pairs belong to unknown drug-disease relationship. The relevant information of 8 pairs of diseases and drugs is shown in Table 2.

Methylprednisolone (DB00959) can treat autoimmune diseases, but we found that methylprednisolone is strongly associated with hematological diseases. According to the definition of MeSH, blood diseases include blood tumors, bone marrow diseases, and other diseases. Methylprednisolone is a biological macromolecular drug, a steroid derivative, and also a glucocorticoid. It can affect the expression of some genes through the cell membrane, thus interfering with the inflammatory response, inhibiting humoral immune response, and has a strong anti-inflammatory effect. Bowen et al. found that high-dose methylprednisolone has a certain effect on patients with recurrent chronic lymphoblastic leukemia [27]. Yao et al. found that methylprednisolone inhibited Wnt signaling pathway by downregulating the expression of LEF-1 protein, and Wnt signaling pathway is highly related to recurrent chronic lymphoblastic leukemia [28].

Mitoxantrone (DB01204) is associated with non-Hodgkin's lymphoma (NHL) and multiple sclerosis (MS). We found that it is also strongly correlated with lymphoid leukemia [29, 30]. Mitoxantrone has significant benefits for tumor control and overall survival in patients with recurrent acute lymphoblastic leukemia.

Prednisolone (DB00860) is a typical steroid drug, which can treat a variety of diseases, including rheumatoid arthritis, asthma, allergies, psoriasis, and multiple sclerosis [31]. However, these diseases are all autoimmune diseases. Therefore, we also found that prednisolone has a strong connection with autoimmune diseases.

We also found that apomorphine (DB00714), cabergoline (DB00248), bromocriptine (DB01200), and rotigotine (DB05271) are related to Parkinson's disease. After querying DrugBank, we knew that these four drugs have therapeutic effects on Parkinson's disease, but they are not included in the known data set.

3.3. GO Function Enrichment Analysis.

Eight reusable drugs were found in this work, five of which have pharmacodynamic effects on Parkinson's diseases. We further performed GO function enrichment analysis on disease-related genes of the disease before drug action. The results are shown in Figure 2(a). It is easy to note that genes are mainly enriched

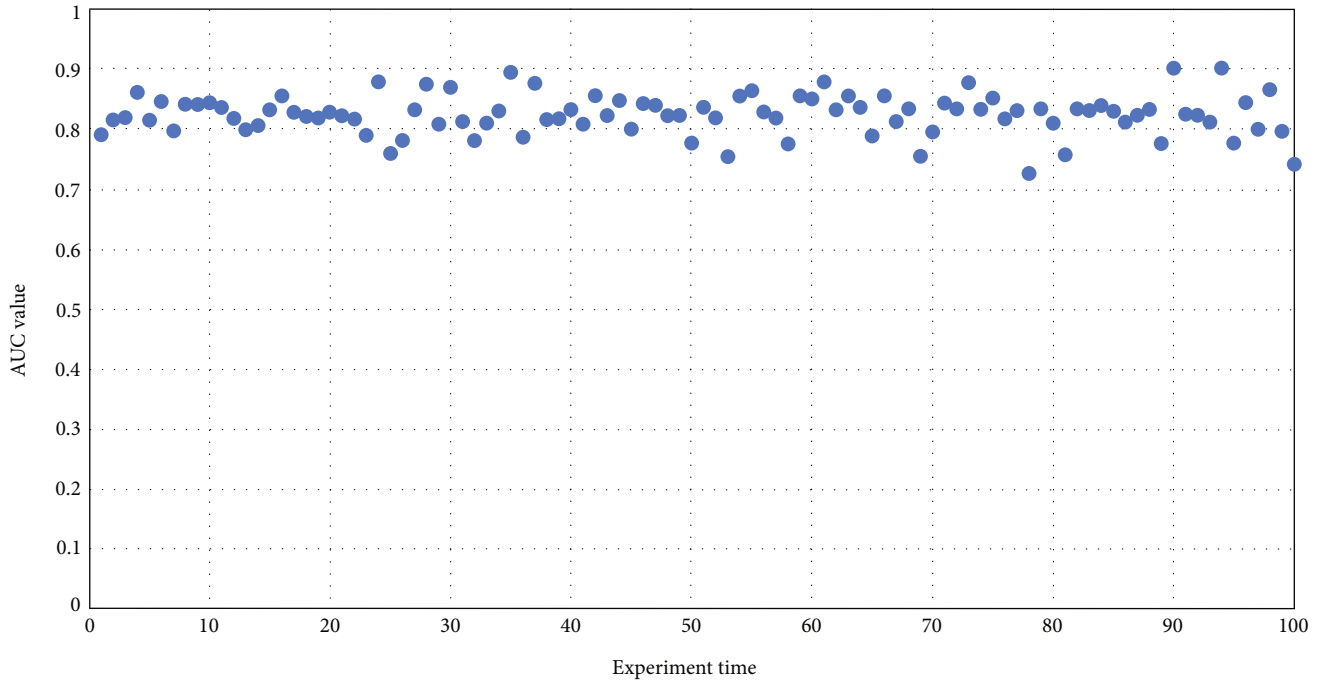


FIGURE 1: The AUC distribution of the random walk with restart method and supervised learning.

TABLE 2: The relevant information of eight diseases and eight drugs.

Disease	Drug	Pearson
Parkinsonian disorders	Apomorphine	0.876
Parkinsonian disorders	Cabergoline	0.876
Bone diseases metabolic	Calcitriol	0.841
Parkinsonian disorders	Bromocriptine	0.840
Leukemia lymphoid	Mitoxantrone	0.834
Hematologic diseases	Methylprednisolone	0.811
Parkinsonian disorders	Rotigotine	0.806
Autoimmune diseases	Prednisolone	0.806

in functional modules such as chromosome breakage (GO:0031052), upregulated cell migration (GO:0030335), and chain replacement (GO:0000732).

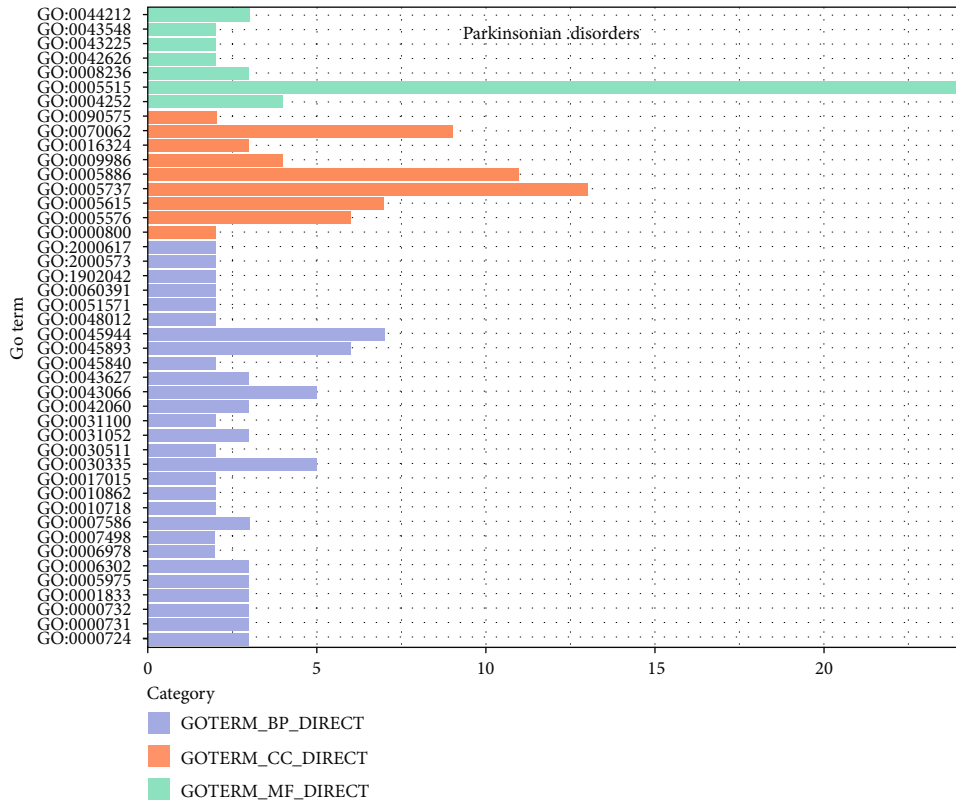
We then analyzed the related genes of Parkinson's diseases after drug action. The results are shown in Figure 2(b). The results show that the gene is enriched in the following functional modules, such as the regulation of exercise (GO:0040012), dopamine binding (GO:0035240), and serotonin binding (GO:0051378).

Before and after random walk, the GO enrichment module of Parkinson's disease has changed significantly. Before random walk, the main enrichment module of Parkinson's syndrome is related to gene expression and cell movement in cells, which may be related to the pathogenesis of Parkinson's disease. After random walk, the relevant genes of Parkinson's syndrome are mainly enriched in some neural transmission modules, which are closely related to the treatment of Parkinson's syndrome.

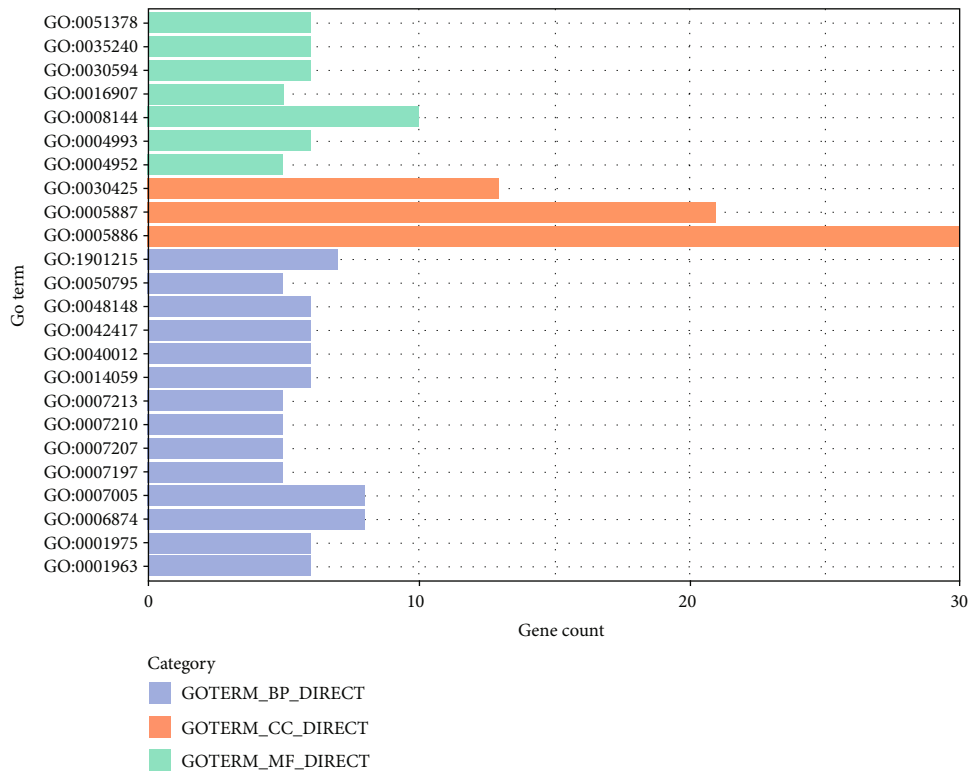
3.4. KEGG Pathway Analysis. We further analyzed the genes related to Parkinson's disease by KEGG pathway. The results are shown in Figure 3. Figure 3(a) shows that genes are mainly enriched in pancreatic secretion (hsa04972), PI3K Akt signaling pathway (hsa04151), and other pathways. After adding drug information and random walk, we conducted KEGG pathway analysis on relevant genes. The results show that the genes are mainly enriched in neural active ligand receptor interaction (hsa04080), calcium signaling pathway (hsa04020), serotonin receptor synapse (hsa04726), and dopamine receptor synapse (hsa04728).

Before and after random walk, the KEGG pathway enrichment module of Parkinson's disease has changed significantly. The approximate change is similar to the result of GO enrichment analysis. Before random walk, the main enrichment pathways of Parkinson's syndrome are related to intracellular signaling pathways. After random walk, the relevant genes of Parkinson's syndrome are mainly enriched in some neural transmission pathways, which are closely related to the treatment of Parkinson's syndrome.

3.5. Key Gene Analysis. In order to further study key genes of Parkinson's disease, we studied the local relationship between Parkinson's disease and trihexyphenidyl, a drug that can treat Parkinson's disease and their related genes on the network (Figure 4). It can be seen from Figure 4 that the key genes of Parkinson's disease are α -synuclein (Gene ID: 6622) and tau protein (Gene ID:4137). α -Synuclein mainly exists at the synapse of the nerve cells and plays a key role in the transmission of neurotransmitters. Tau protein is a microtubule-associated protein that mainly exists in nerve cells. These two proteins are closely related to the pathogenesis of Parkinson's disease.



(a)



(b)

FIGURE 2: The distribution of the related genes of Parkinson's diseases. (a) GO function enrichment on the disease-related genes of Parkinson's disease before drug action; (b) GO function enrichment on disease-related genes of Parkinson's disease after drug action.

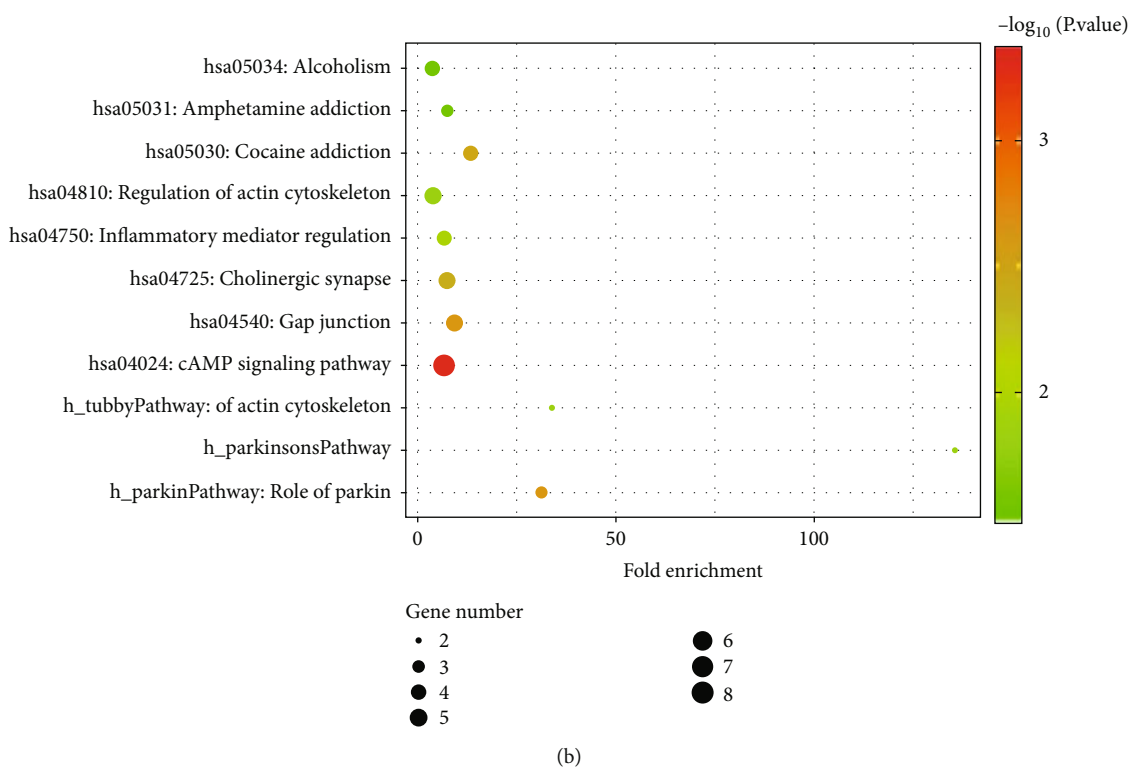
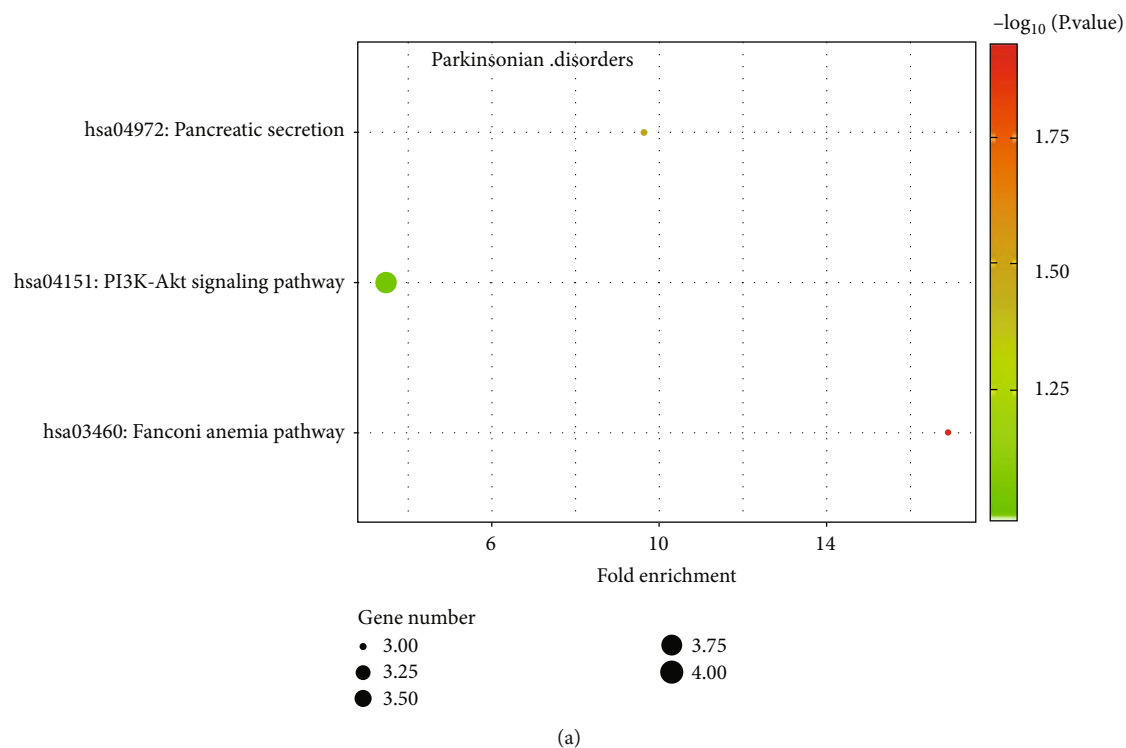


FIGURE 3: KEGG pathway of the related genes of Parkinson’s diseases. (a) KEGG pathway on the disease-related genes of Parkinson’s disease before drug action; (b) KEGG pathway on disease-related genes of Parkinson’s disease after drug action.

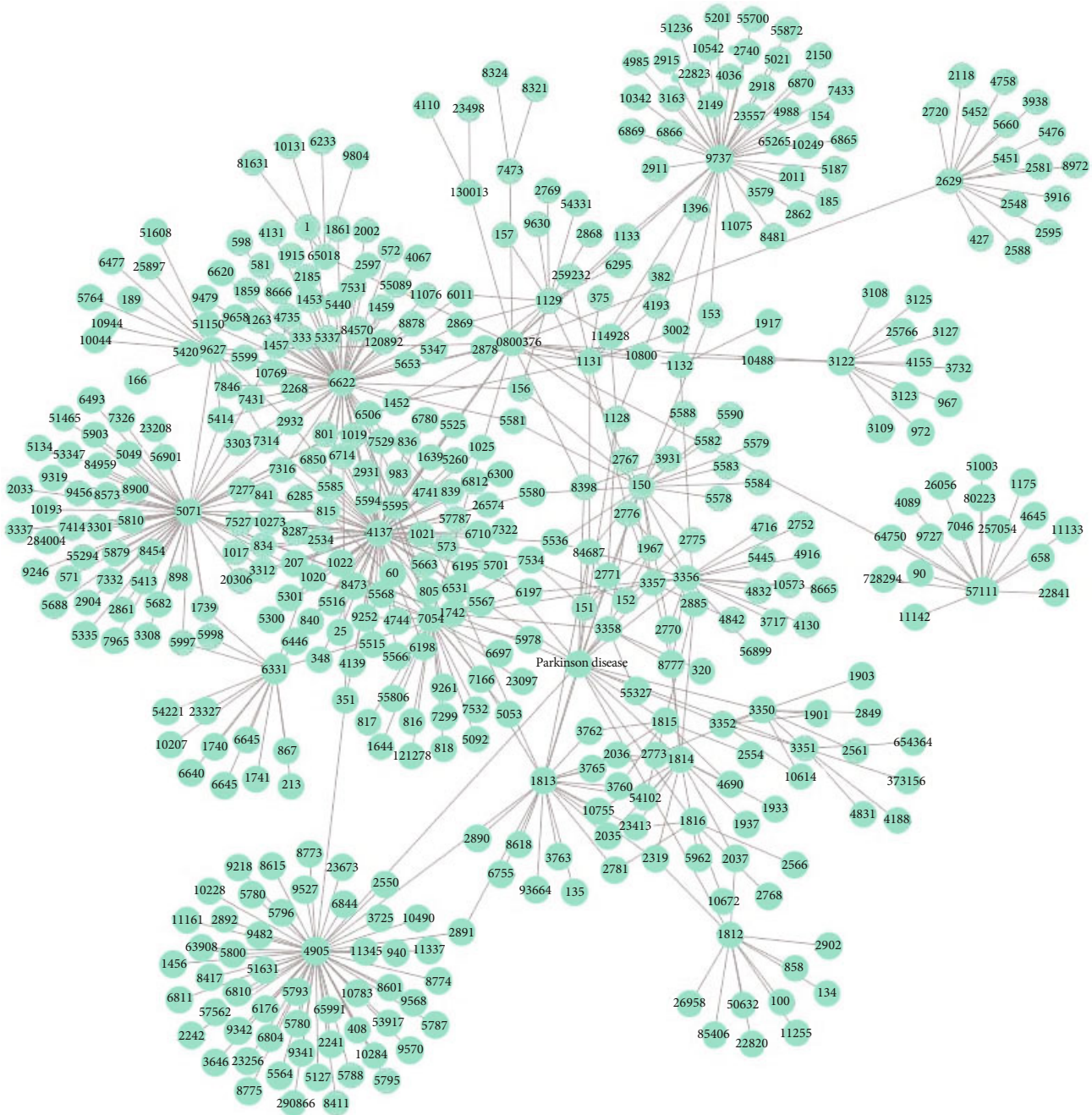


FIGURE 4: The gene network between Parkinson’s disease and trihexyphenidyl.

4. Conclusion

With the prevalence of complex diseases, the existing drugs are far from meeting the needs of human beings to fight diseases. At the same time, due to the rising cost of drug research and development and the long research and development cycle, the research and development means of innovative drugs have become a major challenge in the medical field. In recent years, with the continuous enrichment of disease and drug databases, researchers have realized drug reuse through the correlation analysis of disease-related genes, drugs, and drug-target data. This is a new research

and development idea in the field of pharmaceutical research and development, which reduces the research and development cost of innovative drugs and saves resources. Because most diseases are not single gene defects, they often involve the destruction of the coordination function between genes [32]. Therefore, we explored the relationship between drugs and diseases based on the biological function network. Using HPRD, BioGRID, STRING, and other databases; the protein-protein interaction (PPI) network was constructed. We designed a combination algorithm of random walk and supervised learning to predict the sensitivity of drugs. The accuracy of sensitivity prediction is 82.7%.

With the help of the proposed method, we found that 8 pairs of drug-disease relationships have not been reported, and 5 of them have pharmacodynamic effects on Parkinson's diseases. For Parkinson's disease, we found the changes of its functional modules by adding drug information and comparing before and after random walk, combined with the results of GO and KEGG function enrichment analysis. Using the network diagram of disease and drug-related genes after random walk, we found the key linkage between Parkinson's disease and phenylhexol, a drug for the treatment of Parkinson's disease α -synuclein and tau protein, which provide a useful exploration for the effectiveness of the treatment of Parkinson's disease.

Data Availability

The data used to support the findings of this study are available from the Protein-Protein Interaction (PPI) Network: <https://ppi-net.org>

Conflicts of Interest

The authors declare no conflict of interest, financial or otherwise.

Acknowledgments

The authors thank all the anonymous referees for their valuable suggestions and support. This work is supported by the National Natural Science Foundation of China (62172369); the Key Research and Development Plan of Zhejiang Province (2021C02039); and the Natural Science Foundation of Zhejiang Province (LY20F020016).

References

- [1] A. Joseph, "New drug development in the United States from 1963 to 1999," *Clinical Pharmacology & Therapeutics*, vol. 69, no. 5, pp. 286–296, 2001.
- [2] C. P. Adams and V. V. Brantner, "Estimating the cost of new drug development: is it really \$802 million?," *Health Affairs*, vol. 25, no. 2, pp. 420–428, 2006.
- [3] S. H. Sleight and C. L. Barton, "Repurposing strategies for therapeutics," *Pharmaceutical Medicine*, vol. 24, no. 3, pp. 151–159, 2010.
- [4] N. Novac, "Challenges and opportunities of drug repositioning," *Trends in Pharmacological Sciences*, vol. 34, no. 5, pp. 267–272, 2013.
- [5] S. L. Walker, M. F. R. Waters, and D. N. J. Lockwood, "The role of thalidomide in the management of erythema nodosum leprosum," *Leprosy Review*, vol. 78, no. 3, pp. 197–215, 2007.
- [6] C. H. Huang, P. M. Chang, C. W. Hsu, C. Y. Huang, and K. L. Ng, "Drug repositioning for non-small cell lung cancer by using machine learning algorithms and topological graph theory," *BMC Bioinformatics*, vol. 32, no. 17, pp. 2664–2671, 2016.
- [7] C. Zheng, Z. Guo, C. Huang et al., "Large-scale direct targeting for drug repositioning and discovery," *Scientific Reports*, vol. 5, article 11970, 2015.
- [8] H. Wang, Q. Gu, J. Wei, Z. Cao, and Q. Liu, "Mining drug-disease relationships as a complement to medical genetics-based drug repositioning: where a recommendation system meets genome-wide association studies," *Clinical Pharmacology and Therapeutics*, vol. 97, no. 5, pp. 451–454, 2015.
- [9] F. Cheng, W. Li, Y. Zhou et al., "Prediction of human genes and diseases targeted by xenobiotics using predictive toxicogenomic-derived models (PTDMs)," *Molecular BioSystems*, vol. 9, no. 6, pp. 1316–1325, 2013.
- [10] J. P. Zhou, L. Chen, and Z. H. Guo, "iATC-NRAKEL: an efficient multi-label classifier for recognizing anatomical therapeutic chemical classes of drugs," *Bioinformatics*, vol. 36, no. 5, pp. 1391–1396, 2020.
- [11] E. E. Schadt, "Molecular networks as sensors and drivers of common human diseases," *Nature*, vol. 461, no. 7261, pp. 218–223, 2018.
- [12] A. Califano, A. J. Butte, S. Friend, T. Ideker, and E. Schadt, "Leveraging models of cell regulation and GWAS data in integrative network-based association studies," *Nature Genetics*, vol. 44, no. 8, pp. 841–847, 2012.
- [13] A. Zanzoni, M. Soler-López, and P. Aloy, "A network medicine approach to human disease," *22nd IUBMB Congress/37th FEBS Congress*, vol. 583, no. 11, pp. 1759–1765, 2012.
- [14] S. D. Ghiassian, "Network medicine: a network-based approach to human diseases," *Dissertations & Theses-Gradworks*, vol. 12, no. 1, pp. 56–68, 2015.
- [15] K. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A. L. Barabási, "The human disease network," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 21, pp. 8685–8690, 2007.
- [16] K. Lage, K. Møllgård, S. Greenway et al., "Dissecting spatio-temporal protein networks driving human heart development and related disorders," *Molecular Systems Biology*, vol. 6, no. 1, pp. 381–381, 2010.
- [17] H. Y. Chuang, E. Lee, Y. T. Liu, D. Lee, and T. Ideker, "Network-based classification of breast cancer metastasis," *Molecular Systems Biology*, vol. 3, p. 140, 2007.
- [18] T. Rolland, M. Taşan, B. Charlotiaux et al., "A proteome-scale map of the human interactome network," *Cell Cambridge Ma*, vol. 159, no. 5, pp. 1212–1226, 2014.
- [19] H. Luo, J. Wang, M. Li et al., "Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm," *Bioinformatics*, vol. 32, no. 17, pp. 2664–2671, 2016.
- [20] L. Yu, J. Huang, Z. Ma, J. Zhang, Y. Zou, and L. Gao, "Inferring drug-disease associations based on known protein complexes," *BMC Medical Genomics*, vol. 8, no. S2, p. 2, 2015.
- [21] A. Gottlieb, G. Y. Stein, E. Ruppim, and R. Sharan, "PREDICT: a method for inferring novel drug indications with application to personalized medicine," *Molecular Systems Biology*, vol. 7, no. 1, p. 496, 2011.
- [22] J. Menche, A. Sharma, M. Kitsak et al., "Disease networks. Uncovering disease-disease relationships through the incomplete interactome," *Science*, vol. 347, no. 6224, article 1257601, 2015.
- [23] S. Peng, R. You, H. Wang, C. Zhai, H. Mamitsuka, and S. Zhu, "DeepMeSH: deep semantic representation for improving large-scale MeSH indexing," *Bioinformatics*, vol. 32, no. 12, pp. i70–i79, 2016.
- [24] S. D. Jani, G. L. Argraves, J. L. Barth, and W. S. Argraves, "GeneMesh: a web-based microarray analysis tool for relating differentially expressed genes to MeSH terms," *BMC Bioinformatics*, vol. 11, no. 1, pp. 166–166, 2010.

- [25] M. A. Sartor, A. Ade, Z. Wright et al., “Metab2MeSH: annotating compounds with medical subject headings,” *Bioinformatics*, vol. 28, no. 10, pp. 1408–1410, 2012.
- [26] X. Pan, L. Chen, Liu, Z. Niu, T. Huang, and Y. D. Cai, “Identifying protein subcellular locations with embeddings-based node2loc,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 1, pp. 1–675, 2021.
- [27] D. A. Bowen, T. G. Call, G. D. Jenkins et al., “Methylprednisolone-rituximab is an effective salvage therapy for patients with relapsed chronic lymphocytic leukemia including those with unfavorable cytogenetic features,” *Leukemia & Lymphoma*, vol. 48, no. 12, pp. 2412–2417, 2007.
- [28] Q. M. Yao, P. P. Li, S. M. Liang et al., “Methylprednisolone suppresses the Wnt signaling pathway in chronic lymphocytic leukemia cell line MEC-1 regulated by LEF-1 expression,” *International Journal of Clinical and Experimental Pathology*, vol. 8, no. 7, pp. 7921–7928, 2015.
- [29] P. J. Rosen, C. Rankin, D. R. Head et al., “A phase II study of high dose ARA-C and mitoxantrone for treatment of relapsed or refractory adult acute lymphoblastic leukemia,” *Leukemia Research*, vol. 24, no. 3, pp. 183–187, 2000.
- [30] C. Parker, R. Waters, C. Leighton et al., “Effect of mitoxantrone on outcome of children with first relapse of acute lymphoblastic leukaemia (ALL R3): an open-label randomised trial,” *The Lancet*, vol. 376, no. 9757, pp. 2009–2017, 2010.
- [31] Y. Yang and L. Chen, “Identification of drug-disease associations by using multiple drug and disease networks,” *Current Bioinformatics*, vol. 17, no. 1, pp. 48–59, 2022.
- [32] X. Li, L. Lu, and L. Chen, “Lei Chen, identification of protein functions in mouse with a label space partition method,” *Mathematical Biosciences and Engineering*, vol. 19, no. 4, pp. 3820–3824, 2021.