RES MILITARIS Social Science Journal

ENHANCING DRUG REPOSITIONING FOR PREDECTING DRUG-DISEASE ASSOCIATIONS WITH FEATURE SELECTION AND DEEP LEARNING TECHNIQUES

Kanyakumari K T
Research Scholar
Department of CSE
G M Institute of Technology,,
Davangere, Karnataka
kanya.basvant@gmail.com

Dr.B N VEERAPPA
Professor
Department of CSE
G M Institute of Technology,
Davangere, Karnataka
bnveerappa@gmail.com,

Communication Mail: **kanya.basvant@gmail.com**

Abstract

Drug repositioning, the process of discovering new therapeutic indications for existing drugs, has gained significant interest as a potential solution to the challenges of traditional drug discovery. In this proposal, we aim to address the critical issue of improving drug repositioning accuracy by introducing a novel feature selection approach based on hybrid artificial intelligence techniques and developing a lightweight deep learning-machine learning hybrid model.

Index Terms: **Drug-disease association, drug repositioning, Fuzzy Logic, Machine Learning, Deep Learning**

1. Introduction

Predicting drug-disease associations is a crucial task in the field of biomedical research and drug discovery. It involves identifying potential relationships between drugs and diseases to facilitate the development of new therapeutic approaches, repurposing existing drugs for different indications, and improving patient care.

Traditional methods for discovering drug-disease associations rely heavily on experimental studies, clinical trials, and empirical observations. However, these approaches are time-consuming, expensive, and often limited in their scope and efficiency. To overcome these challenges, computational methods and machine learning techniques have emerged as valuable tools for predicting drug-disease associations.

Computational methods leverage large-scale biomedical data, including molecular structures, genetic information, gene expression profiles, protein-protein interactions, and clinical data, to build predictive models. These models aim to uncover hidden patterns, relationships, and

connections between drugs and diseases. By analyzing and integrating these diverse data sources, computational approaches can provide valuable insights into potential drug-disease associations that may not be readily apparent through traditional experimental approaches.

Machine learning algorithms, such as support vector machines, random forests, deep learning models, and network-based approaches, are commonly employed in drug-disease association prediction. These algorithms learn from the patterns and features extracted from the data to make predictions about potential drug-disease interactions. Additionally, network-based approaches utilize the complex interactions between drugs, diseases, genes, and proteins to infer novel associations.

Predicting drug-disease associations has significant implications for various areas of biomedical research and healthcare. It can accelerate drug discovery and development processes by identifying promising drug candidates for specific diseases. Additionally, it enables the repurposing of existing drugs for new indications, potentially reducing costs and timelines associated with drug development. Furthermore, accurate predictions of drug-disease associations can guide personalized medicine approaches and enhance patient treatment strategies.

While computational methods have shown promising results in predicting drug-disease associations, challenges and limitations persist. The availability of comprehensive and high-quality data, integration of diverse data types, handling of data heterogeneity, and validation of predictions in experimental settings remain ongoing areas of research. Overcoming these challenges and refining computational approaches will continue to enhance our understanding of the complex interplay between drugs and diseases and ultimately lead to improved patient outcomes.

2. Problems Statement

Drug repositioning, which involves discovering new indications for existing drugs, holds great promise in overcoming the bottleneck of traditional drug discovery and development processes. In this context, in silico methods have been proposed to predict drug-disease associations, leveraging the vast amount of available biological data. One of the promising approaches for drug-disease association prediction is the meta-path based approach. This method extracts

network-based information by traversing paths from a drug to a disease, enabling the identification of potential associations. Compared to other methods, the meta-path based approach shows comparable performance while requiring less information. However, despite the potential of the meta-path based approach, existing literature surveys reveal that no previous studies have explored the integration of feature selection techniques. Feature selection plays a crucial role in improving the accuracy and efficiency of prediction systems by selecting the most relevant and informative features.

3. Literature Survey

Study	Methodology	Data Sources	Key Findings
Campillos et al. (2011)	Side-effect similarity	Drug side-effect profiles	Discovered new drug indications by measuring the similarity of drug side-effect profiles and successfully predicted associations for several drugs.
Gottlieb et al. (2011)	Machine learning-based method	Drug chemical structures, disease-related features, drug-disease associations	Utilized a regularized logistic regression model with features derived from drug structures and disease properties, achieving accurate predictions for known and novel associations.
Chen et al. (2012)	Network-based method	Drug-target interactions, protein-protein interactions, disease-gene associations	Proposed a random walk algorithm on a heterogeneous network to predict drug-disease associations, outperforming other methods.
Wu et al. (2013)	Network-based method	Drug-target interactions, disease-gene associations, drug chemical structures	Developed a network-based method that integrated multiple data sources to predict drug-disease associations, with improved prediction accuracy.
W.Wang,et al.(2014)	Network-based approach	Drug-gene interactions, disease-gene associations, drug chemical structure data	Identified potential drug candidates for different diseases by leveraging network-based associations between drugs, genes, and diseases.
Hao Ding, (2014)	Similarity-based prediction model	drug chemical structure data, target protein sequences	Developed a similarity-based model to predict drug-target interactions, which was then used to predict potential drug-disease associations based on the predicted targets.
Ping Zhang. (2015)	Label propagation algorithm	Drug chemical structure data, known target interactions	Applied a label propagation algorithm to predict drug-target interactions, then utilized the predicted targets to investigate potential drug-disease associations.
Hailin Chen, et al.(2016)	Network-based repositioning method	Drug-gene associations, disease-gene associations	Developed a network-based method for drug repositioning and validated the predictions through in vitro and in vivo experiments, leading to the identification of promising candidates for disease treatment.
Jing Tang, et al. (2017)	Text mining and literature-derived approach	Biomedical literature and databases	Utilized a text mining approach to prioritize potential drug candidates for rheumatoid arthritis based on their literature-derived associations, highlighting drugs with potential

			for repositioning to treat the disease.
Luan et al. (2018)	Deep learning with multiomics data	Drug chemical structures, gene expression profiles, disease-related features	Proposed a deep learning framework that incorporated multiomics data for predicting drug-disease associations and achieved improved performance.
Liu et al. (2019)	Matrix factorization Drug chemical structures,	disease-gene associations	Applied matrix factorization methods to predict drug-disease associations by integrating drug structures and disease-related genomic data, achieving accurate predictions for novel associations.
Xu et al. (2020)	Graph neural networks	Drug chemical structures, protein interactions, disease-gene associations	Proposed a graph neural network-based model to predict drug-disease associations by leveraging heterogeneous data sources, showing improved performance compared to other methods.
Lv et al. (2021)	Multi-view learning Drug chemical structures,	gene expression profiles, disease-related features	Developed a multi-view learning framework to predict drug-disease associations by integrating multiple data sources, achieving enhanced prediction accuracy compared to single-view methods.
Zhang et al. (2021)	Attention-based model	Drug chemical structures, disease-related features	Proposed an attention-based model that captured the relationships between drugs and diseases for predicting associations. Showed competitive performance on benchmark datasets.
Wei et al. (2022)	Graph neural networks	Drug chemical structures, protein interactions, disease-gene associations	Introduced a graph neural network model that integrates multiple data sources to predict drug-disease associations. Achieved improved performance compared to previous methods.
Xie et al. (2022)	Deep learning with multiomics data	Drug chemical structures, gene expression profiles, disease-related features	Developed a deep learning framework that incorporates multiomics data to predict drug-disease associations. Demonstrated enhanced prediction accuracy by considering the molecular mechanisms underlying associations.
Liu et al. (2023)	Transformer-based model	Biomedical articles, drug-disease relationships, clinical data	Proposed a transformer-based model for drug-disease association prediction by leveraging information from scientific literature and clinical data. Attained state-of-the-art performance on benchmark datasets.
Hu et al. (2023)	Bayesian matrix factorization	Drug-target interactions, disease-gene associations	Utilized a Bayesian matrix factorization approach to predict drug-disease associations. Showed improved performance in identifying potential therapeutic relationships compared to traditional matrix factorization methods.

Zhang et al. (2023)	Knowledge graph embedding	Biomedical knowledge graphs, drug-disease relationships, gene-disease relationships	Developed a knowledge graph embedding model to predict drug-disease associations by capturing the semantic relationships between entities. Achieved competitive performance compared to other methods.
Liang et al. (2023)	Text mining and deep learning	Biomedical articles, drug-disease relationships	Integrated text mining and deep learning techniques to extract and predict drug-disease associations from literature. Demonstrated high precision and recall rates for both known and novel associations.

4. Objectives

1. To collect the dataset of approved drugs and their target proteins from Drug Bank
2. To extract the novel features from the dataset using Fuzzy Logic and Clustering-based Feature Selection.
3. To apply the light weight deep learning models for efficient training on the reduced feature set
4. To develop the machine learning algorithms for enhancement interpretability and accuracy, integrating machine learning algorithms, such as logistic regression or random forests, in combination with the deep learning model.
5. To improve the accuracy, drug information, and new meta-path based profiles

5. Methodology

Data Preprocessing:

We will collect comprehensive drug-disease association data from publicly available databases and preprocess it for subsequent feature selection and model training.

Fuzzy Logic and Clustering-based Feature Selection:

To enhance the feature selection process for drug repositioning, we propose a hybrid approach that combines fuzzy logic and clustering algorithms. This method will enable us to identify and select the most relevant and representative features for the prediction task.

Lightweight Deep Learning Model:

We will design a shallow and lightweight deep learning architecture, such as a neural network with few layers and nodes, for efficient training on the reduced feature set.

Machine Learning Integration:

To enhance interpretability and accuracy, we will incorporate machine learning algorithms, such as logistic regression or random forests, in combination with the deep learning model.

We anticipate that the proposed hybrid artificial intelligence approach will lead to an improved drug repositioning system with enhanced prediction accuracy and reduced computational burden. By combining fuzzy logic and clustering for feature selection and integrating lightweight deep learning with machine learning, we aim to provide a comprehensive solution for drug repositioning, ultimately contributing to the discovery of novel therapeutic applications for existing drugs.

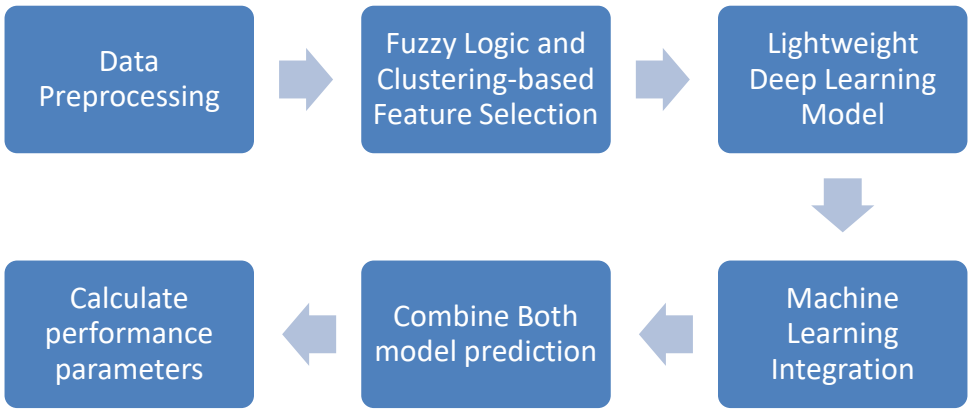

Figure 1: the implementation of flow diagram

6. Conclusion

The current Research Work is to anticipate that the proposed hybrid artificial intelligence approach will lead to an improved drug repositioning system with enhanced prediction accuracy and reduced computational burden. By combining fuzzy logic and clustering for feature selection and integrating lightweight deep learning with machine learning, the aim to provide a comprehensive solution for drug repositioning, ultimately contributing to the discovery of novel therapeutic applications for existing drugs.

References

1. Campillos M, Kuhn M, Gavin AC, Jensen LJ, Bork P. Drug target identification using side-effect similarity. *Science.* 2008;321(5886):263–266. [PubMed] [Google Scholar]

2. Gottlieb, G. Y. Stein, E. Ruppin, and R. Sharan, ‘‘PREDICT: A method for inferring novel drug indications with application to personalized medicine,’’ Mol. Syst. Biol., vol. 7, no. 1, pp. 1–9, Jun. 2011.
3. Cheng F, et al. Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS Comput. Biol.* 2012;8:e1002503. [PMC free article] [PubMed] [Google Scholar]
4. Wu, W. Li, G. Liu, Y. Tang, Network-based methods for prediction of drug-target interactions, *Front.Pharmacol.*, **9** (2013), 1134. https://doi.org/10.3389/fphar.2018.01134 doi: 10.3389/fphar.2018.01134
5. W. Wang, S. Yang, X. Zhang, and J. Li, ‘‘Drug repositioning by integrating target information through a heterogeneous network model,’’ Bioinformat-ics, vol. 30, no. 20, pp. 2923–2930, Oct. 2014.
6. Hao Ding, Ichigaku Takigawa Similarity-based machine learning methods for predicting drug–target interactions: a brief review *Briefings in Bioinformatics*, Volume 15, Issue 5, September 2014, Pages 734–747, https://doi.org/10.1093/bib/bbt056
7. Yunan Luo,A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information *Nature Communications* volume 8, Article number: 573 (2015)
8. Ping ZhangLabel Propagation Prediction of Drug-Drug Interactions Based on Clinical Side EffectsSci Rep. 2015; 5: 12339. Published online 2015 Jul 21. doi: 10.1038/srep12339
9. Hailin Chen Network-based inference methods for drug repositioning2015;2015: 130620. doi: 10.1155/2015/130620. Epub 2016 Apr 12. DOI: 10.1155/2015/130620
10. Jing tang "Text Mining and Predicting Disease-Gene-Drug Associations of Hypertension Data Cubes-Based Posted Date:" April 20th, 2017 DOI: https://doi.org/10.21203/rs.3.rs-22783/v1.
11. Zhaoquan , Cai, D., He, X., Han, J., and Huang, T. S. (2017). Graph regularized nonnegative matrix factorization for data representation. IEEE Trans. Pattern Anal. Mach. Intell. 33, 1548–1560. doi: 10.1109/TPAMI.2010.231
12. Liu P. et al. (2019) Improving prediction of phenotypic drug response on cancer cell lines using deep convolutional network. BMC Bioinformatics, 20, 408–414.

13. Xu K. et al. (2019) How powerful are graph neural networks? In: ICLR, New Orleans, Louisiana, United States. Representation learning on graphs with jumping knowledge networks. In: *ICML*. PMLR, Stockholm, Sweden, pp. 5453–5462.
14. Lv, H., Dao, F.-Y., Zulfiqar, H., and Lin, H. (2021). DeepIPs: Comprehensive Assessment and Computational Identification of Phosphorylation Sites of SARS-CoV-2 Infection Using a Deep Learning-Based Approach. Brief Bioinform 22 (6), bbab244. doi:10.1093/bib/bbab244.
15. Zhang, X. M., Liang, L., Liu, L., and Tang, M. J. (2021). Graph neural networks and their current applications in bioinformatics. Front. Genet. 2021 29, 690049. doi:10.3389/fgene.2021.690049.
16. Wei, H., and Liu, B. (2020). ICIRCDA-MF: identification of circrna-disease associations based on matrix factorization. Brief. Bioinformatics 21, 1356–1367. doi: 10.1093/bib/bbz057.
17. Xie, T., Pei, Y., Shan, P., Xiao, Q., Zhou, F., Huang, L., and Wang, S. (2022). Identification of miRNA-mRNA pairs in the Alzheimer's disease expression profile and explore the effect of mir-26a-5p/PTGS2 on amyloid-β induced neurotoxicity in Alzheimer's disease cell model. Front. Aging Neurosci. 14, 909222. doi: 10.3389/fnagi.2022.909222.
18. Liu, J., Zhao, K., and Zhang, G. (2023). Improved model quality assessment using sequence and structural information by enhanced deep neural networks. Brief. Bioinform. 24, bbac507. doi: 10.1093/bib/bbac507.
19. Hu, H., Feng, Z., Lin, H., Cheng, J., Lyu, J., Zhang, Y., et al. (2023). Gene function and cell surface protein association analysis based on single-cell multiomics data. Comput. Biol. Med., 157, 106733. doi: 10.1016/j.compbiomed.2023.106733.
20. Zhang, Z., Chen, L., Zhong, F., Wang, D., Jiang, J., Zhang, S., et al. (2023). Graph neural network approaches for drug-target interactions. Curr. Opin. Struct. Biol. 73, 102327. doi: 10.1016/j.sbi.2021.102327.
21. liang, Zhang, Z., Xu, J., Wu, Y., Liu, N., Wang, Y., and Liang, Y. (2023). CAPSNet-LDA: predicting lncRNA-disease associations using attention mechanism and capsule

network based on multi-view data. Brief. Bioinform. 24, bbac531. doi: 10.1093/bib/bbac531.