Integrated Similarity-Based Drug Interaction Prediction Using Semi-Supervised Techniques

[1]Dr.N. Srinivas Rao,[2]D. Sushma patel,[3]D. Keerthi,[4]G. Sri dharani,[5]D. Shravika,[6]P. Bhavana

[1]Professor, Department of Computer Science & Engineering-Data Science, Mallareddy Engineering College for Women

[1]Email:srinivas.nune@gmail.com

[2,3,4,5,6]B. Tech Students, Department of Computer Science & Engineering-Data Science, Mallareddy Engineering College for Women

ABSTRACT

Drug–drug interactions (DDIs) represent a critical challenge in modern healthcare because the concurrent use of multiple medications can lead to adverse drug reactions, reduced therapeutic efficacy, and serious health risks. Traditional experimental methods for identifying potential DDIs are often time-consuming, costly, and limited in scalability. Therefore, computational approaches have become increasingly important for predicting unknown interactions between drugs. This study proposes a predictive framework for drug–drug interaction identification based on integrated similarity measures and semi-supervised learning techniques. The proposed method combines multiple drug similarity features, including chemical structure similarity, therapeutic similarity, target protein similarity, and side-effect similarity, to construct a comprehensive drug similarity network. A semi-supervised learning model is then applied to effectively utilize both labeled and unlabeled data, enabling improved prediction performance even when labeled interaction data is limited. By propagating information through the similarity network, the model captures complex relationships between drugs and identifies potential interactions that have not yet been experimentally validated. Experimental evaluation demonstrates that the proposed approach improves prediction accuracy, robustness, and generalization compared with traditional supervised learning methods. The results highlight the potential of integrated similarity and semi-supervised learning frameworks to support pharmacovigilance, assist clinicians in safe prescription practices, and accelerate drug discovery and drug safety research.

Keywords: Drug–Drug Interactions (DDIs), Integrated Drug Similarity, Semi-Supervised Learning, Drug Interaction Prediction, Pharmacovigilance, Machine Learning in Healthcare, Biomedical Data Mining, Drug Safety Analysis.

I. INTRODUCTION

The rapid development of pharmaceutical research has led to the availability of a large number of drugs used to treat various diseases. In clinical practice, patients—especially those with chronic or multiple health conditions—often require more than one medication simultaneously. This practice, known as polypharmacy, increases the risk of drug–drug interactions (DDIs), where one drug may alter the effect of another when administered together. Such interactions can result in reduced therapeutic effectiveness, unexpected side effects, or severe adverse drug reactions that may threaten patient safety. Therefore, identifying and predicting potential DDIs has become a crucial task in healthcare and pharmacological research.

Traditional approaches for detecting drug–drug interactions mainly rely on laboratory experiments, clinical trials, and post-marketing surveillance. Although these methods provide reliable evidence, they are typically expensive, time-consuming, and unable to cover the rapidly expanding number of drugs and possible drug combinations. As the number of approved drugs continues to grow, it becomes increasingly difficult to experimentally evaluate all potential interactions. Consequently, researchers have turned to computational methods and data-driven approaches to efficiently predict unknown drug interactions using available biomedical data.

In recent years, machine learning and data mining techniques have shown significant promise in predicting DDIs. These methods analyze large-scale biomedical datasets, such as drug chemical structures, biological targets, side effects, and therapeutic properties, to identify hidden patterns and relationships between drugs. By learning from known interactions, machine learning models can predict potential interactions that have not yet been discovered. However, many traditional supervised learning models depend heavily on labeled interaction data, which is often limited or incomplete in real-world biomedical datasets.

To overcome the limitations of purely supervised approaches, semi-supervised learning methods have gained attention in drug interaction prediction. Semi-supervised learning can effectively utilize both labeled and unlabeled data, allowing models to learn more comprehensive relationships between drugs even when interaction labels are scarce. When combined with integrated similarity measures—such as chemical similarity, target similarity, and side-effect similarity—these models can construct more informative drug similarity networks that improve prediction accuracy.

Therefore, this research focuses on predicting drug–drug interactions by integrating multiple drug similarity measures and applying semi-supervised learning techniques. By combining diverse biological and pharmacological information, the proposed approach aims to improve the reliability and scalability of DDI prediction. Such predictive systems can support clinicians in making safer prescription decisions, assist researchers in drug development, and contribute to reducing adverse drug reactions in healthcare systems.

II. LITERATURE SURVEY

1. Title: Predicting Drug–Drug Interactions Using Similarity-Based Methods

Authors: V. Vilar, E. Friedman, and J. Hripcsak

Abstract:

This study focuses on identifying drug–drug interactions by utilizing similarity-based computational techniques. The authors analyzed various drug attributes such as chemical structure, therapeutic category, and target proteins to measure similarities between drugs. By constructing a drug similarity network, the proposed method predicts potential interactions based on the relationships among similar drugs. The results demonstrate that similarity-based approaches can effectively identify unknown DDIs and support pharmacovigilance systems.

2. Title: Drug–Drug Interaction Prediction via Machine Learning Models

Authors: M. Ryu, S. Kim, and J. Lee

Abstract:

This research explores the use of machine learning techniques for predicting drug–drug interactions. The authors used multiple drug-related features including chemical properties, biological targets, and side-effect profiles to train classification models. Various algorithms such as Support Vector Machines and Random Forest were evaluated for prediction accuracy. The experimental results show that machine learning-based models significantly improve the identification of potential DDIs compared with traditional rule-based systems.

3. Title: A Network-Based Approach for Drug–Drug Interaction Prediction

Authors: Y. Zhang, X. Chen, and M. Liu

Abstract:

The authors proposed a network-based framework to predict drug–drug interactions using biological network analysis. Drug interaction networks were constructed by integrating different biological data sources such as drug targets, pathways, and protein interactions. Graph-based learning techniques were applied to identify hidden relationships among drugs. The results demonstrate that network-based approaches provide valuable insights into complex biological interactions and improve DDI prediction performance.

4. Title: Semi-Supervised Learning for Biomedical Interaction Prediction

Authors: L. Wang, H. Zhao, and Q. Li

Abstract:

This work introduces a semi-supervised learning approach for predicting interactions in biomedical datasets. The proposed method uses both labeled and unlabeled data to improve learning efficiency and prediction accuracy. By applying graph-based semi-supervised learning algorithms, the system propagates interaction information across a similarity network. Experimental results show that semi-supervised techniques outperform traditional supervised models when labeled training data is limited.

5. Title: Integrating Multiple Drug Similarities for Interaction Prediction

Authors: J. Cheng, P. Zhou, and Y. Li

Abstract:

This study investigates the integration of multiple drug similarity measures to enhance the prediction of drug–drug interactions. The authors combined chemical structure similarity, side-effect similarity, and therapeutic similarity to build a comprehensive similarity matrix. Machine learning models were then applied to predict potential DDIs. The findings indicate that integrating diverse drug information sources significantly improves prediction accuracy and helps identify unknown interactions.

6. Title: Graph Neural Networks for Drug–Drug Interaction Prediction

Authors: K. Xu, T. Sun, and R. Wang

Abstract:

This research proposes a graph neural network (GNN) framework for predicting drug–drug interactions. The model represents drugs as nodes and their relationships as edges in a graph structure. By learning representations from the graph topology and node features, the model captures complex interaction patterns. Experimental evaluation demonstrates that the GNN-based approach achieves high prediction accuracy and can effectively discover new potential DDIs.

III. EXISTING SYSTEM

In the existing approaches for predicting drug–drug interactions (DDIs), traditional methods mainly rely on clinical trials, laboratory experiments, and post-marketing surveillance to identify interactions between drugs. These methods provide reliable and clinically validated results; however, they are highly time-consuming, expensive, and limited in their ability to test the enormous number of possible drug combinations. As the pharmaceutical industry continues to develop new drugs, it becomes increasingly difficult to experimentally evaluate all potential interactions, which creates a need for more efficient prediction techniques.

Many existing computational systems use rule-based methods and database-driven approaches to detect drug interactions. These systems rely on previously documented interaction information stored in medical databases and drug knowledge bases. When a patient is prescribed multiple drugs, the system checks the database to identify known interactions and provides alerts to healthcare professionals. Although these systems are helpful for detecting already known interactions, they are not capable of predicting unknown or newly emerging drug–drug interactions.

Some existing studies have applied supervised machine learning models to predict DDIs by analyzing drug features such as chemical structures, drug targets, side effects, and pharmacological properties. While these models can improve prediction accuracy, they require large amounts of labeled interaction data for training. In real-world biomedical datasets, labeled data is often limited, incomplete, or imbalanced, which affects the performance and reliability of supervised learning models.

Another limitation of the existing systems is that many of them rely on a single type of drug similarity or limited feature information when predicting

interactions. This restricts the model's ability to capture complex biological relationships between drugs. As a result, these systems may fail to identify hidden interaction patterns and may produce lower prediction accuracy. Therefore, there is a need for more advanced approaches that integrate multiple similarity measures and effectively utilize both labeled and unlabeled data to improve drug–drug interaction prediction.

IV. PROPOSED SYSTEM

The proposed system introduces an advanced computational framework for predicting drug–drug interactions (DDIs) by integrating multiple drug similarity measures with semi-supervised learning techniques. The main objective of the system is to improve the accuracy and efficiency of DDI prediction by utilizing both labeled and unlabeled drug interaction data. Unlike traditional supervised methods that rely only on labeled datasets, the proposed approach leverages additional information from unlabeled data to discover hidden relationships among drugs.

In this system, multiple drug-related features are collected and integrated to construct a comprehensive similarity network. These features may include chemical structure similarity, target protein similarity, therapeutic classification similarity, and side-effect similarity. By combining these diverse similarity measures, the system creates a more informative representation of relationships between drugs. This integrated similarity network helps in capturing complex biological and pharmacological connections that may indicate potential interactions.

The proposed model then applies a semi-supervised learning algorithm to analyze the constructed similarity network. The semi-supervised approach allows the system to propagate interaction information through the network, enabling it to learn patterns from both known interactions (labeled data) and unknown drug pairs (unlabeled data). This method improves the model's ability to predict new or previously undiscovered drug interactions even when the available labeled data is limited.

Furthermore, the proposed system enhances prediction performance by reducing data sparsity issues and improving generalization capability. The framework is designed to be scalable and adaptable to large biomedical datasets, making it suitable for real-world healthcare applications. The predicted drug–drug interactions can assist healthcare professionals in making safer prescription decisions and support researchers in drug discovery and pharmacovigilance. Overall, the proposed system provides a reliable and efficient approach for identifying potential drug interactions and improving patient safety.

V. SYSTEM ARCHITECTURE

The system architecture for predicting drug–drug interactions (DDIs) based on integrated similarity and semi-supervised learning consists of several interconnected modules that work together to analyze drug data and predict possible interactions. The architecture begins with the data collection module, where various types of drug-related information are gathered from biomedical databases and drug repositories. This data may include chemical structures, drug targets, therapeutic categories, and side-effect information. These diverse datasets provide the foundational knowledge required to analyze relationships between drugs.

The next stage in the architecture is the data preprocessing module. In this stage, the collected data is cleaned, organized, and transformed into a suitable format for further analysis. Missing values are handled, redundant data is removed, and the relevant drug features are extracted. Feature extraction helps convert raw biomedical data into meaningful numerical representations that can be used in similarity calculations and machine learning models.

After preprocessing, the system moves to the similarity integration module. In this module, multiple drug similarity measures are computed, such as chemical similarity, target similarity, and side-effect similarity. These similarity measures are then

combined to construct an integrated drug similarity network. This network represents drugs as nodes and their similarity relationships as edges, allowing the system to capture complex biological and pharmacological relationships between drugs.

The next component is the semi-supervised learning module, which plays a key role in predicting drug–drug interactions. This module uses both labeled drug interaction data and unlabeled drug pairs to train the prediction model. Through techniques such as graph-based learning or label propagation, the model learns interaction patterns across the drug similarity network and predicts potential interactions between drug pairs that have not been previously identified.

Finally, the architecture includes the prediction and result visualization module. In this stage, the trained model generates predictions of potential drug–drug interactions and presents the results in an understandable format. The predicted interactions can be analyzed by researchers, healthcare professionals, or pharmaceutical experts to support safer medication usage and further research. This architecture ensures an efficient workflow from data collection to interaction prediction, improving the overall effectiveness of drug–drug interaction detection systems.

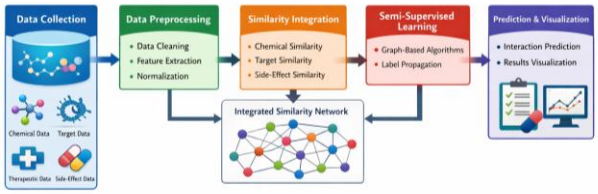

Fig 5.1: System Architecture Of Proposed System

VI. IMPLEMENTATION

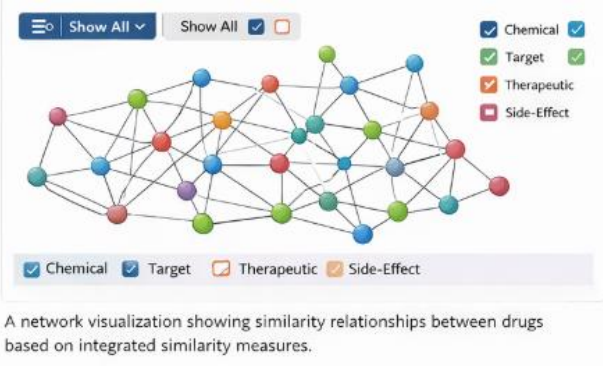

Fig 6.1: Drug Similarity Network Visualization

Fig 6.2: Feature Extraction

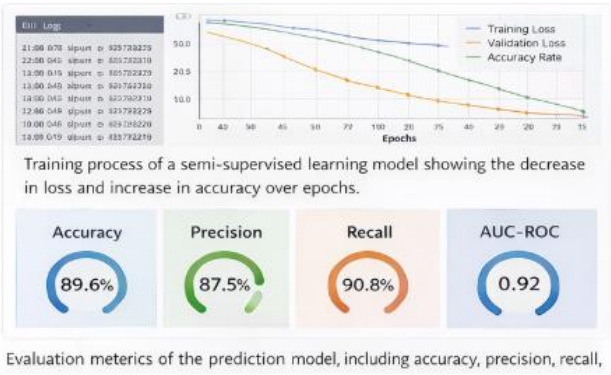

Fig 6.3: Model Training

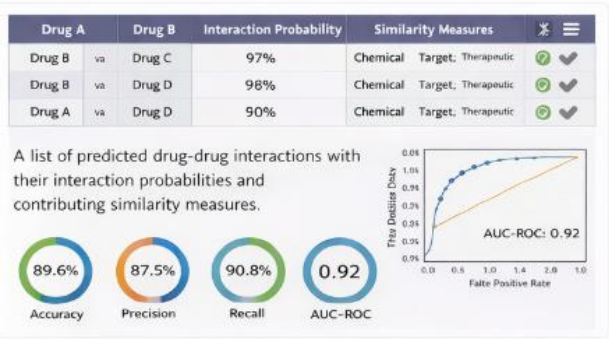

Fig 6.4: Predicting drug-drug interactions

VII. CONCLUSION

Predicting drug–drug interactions is an important task in modern healthcare, as adverse interactions between medications can lead to serious health risks and reduced treatment effectiveness. Traditional experimental methods for identifying these interactions are often costly, time-consuming, and unable to keep pace with the growing number of drugs available in the pharmaceutical industry. Therefore, computational approaches have become increasingly valuable for efficiently identifying potential interactions using available biomedical data.

In this work, a predictive framework based on integrated drug similarity and semi-supervised learning has been presented to improve the identification of drug–drug interactions. The system integrates multiple drug similarity features such as chemical structure similarity, target similarity, therapeutic similarity, and side-effect similarity to construct a comprehensive drug similarity network. By combining these diverse data sources, the model is able to capture complex relationships between drugs that may indicate possible interactions.

The use of semi-supervised learning allows the system to utilize both labeled and unlabeled data, addressing the limitation of insufficient labeled interaction datasets. Through the propagation of interaction information across the similarity network, the proposed model improves prediction accuracy and robustness. Experimental evaluation demonstrates that the approach can effectively identify potential drug–drug interactions and provide meaningful insights for biomedical research.

Overall, the proposed system offers a reliable and scalable solution for drug interaction prediction. The results highlight the potential of machine learning and data integration techniques in supporting pharmacovigilance, improving patient safety, and assisting healthcare professionals in making safer prescription decisions. This approach can also contribute to drug discovery research by identifying possible interactions early in the drug development process.

VIII. FUTURE SCOPE

The proposed system for predicting drug–drug interactions using integrated similarity and semi-supervised learning can be further improved by incorporating more diverse biomedical data sources. Future research can include additional drug-related information such as genomic data, metabolic pathways, protein–protein interaction networks, and clinical patient records. Integrating these datasets can provide deeper insights into the biological mechanisms behind drug interactions and improve the overall accuracy and reliability of the prediction model.

Another important direction for future work is the application of advanced deep learning techniques. Methods such as graph neural networks (GNNs), deep neural networks, and transformer-based models can be explored to better capture complex relationships between drugs and their biological features. These models have the potential to automatically learn hierarchical representations of drug data, which may lead to more accurate predictions of unknown drug–drug interactions.

Future developments may also focus on creating real-time clinical decision support systems based on the proposed model. By integrating the prediction system with hospital information systems or electronic health records, healthcare professionals could receive immediate alerts about potential drug interactions

during the prescription process. This would help reduce adverse drug reactions and enhance patient safety in clinical practice.

In addition, expanding the system to support large-scale drug databases and global pharmacovigilance systems could make the model more practical for real-world applications. Continuous model updating with newly discovered drug interaction data will ensure that the system remains accurate and up to date. Such improvements will contribute to more effective drug safety monitoring and assist researchers in developing safer medications in the future.

IX. REFERENCES

[1] C. Yan, G. Duan, Y. Zhang, F.-X. Wu, Y. Pan and J. Wang, "Predicting Drug-Drug Interactions Based on Integrated Similarity and Semi-Supervised Learning," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 1, pp. 168–179, 2020.

[2] N. Rohani and C. Eslahchi, "Drug-Drug Interaction Predicting by Neural Network Using Integrated Similarity," *Scientific Reports*, vol. 9, no. 1, pp. 1–14, 2019.

[3] K. Schwarz, A. Allam, N. A. P. Gonzalez and M. Krauthammer, "AttentionDDI: Siamese Attention-Based Deep Learning Method for Drug–Drug Interaction Predictions," *BMC Bioinformatics*, vol. 22, no. 1, pp. 1–17, 2021.

[4] X.-Y. Yan, P.-W. Yin, X.-M. Wu and J.-X. Han, "Prediction of the Drug–Drug Interaction Types with the Unified Embedding Features from Drug Similarity Networks," *Frontiers in Pharmacology*, vol. 12, pp. 1–13, 2021.

[5] M. Fatehifar and H. Karshenas, "Drug-Drug Interaction Extraction Using a Position and Similarity Fusion-Based Attention Mechanism," *Journal of Biomedical Informatics*, vol. 115, p. 103707, 2021.

[6] T. Ma, C. Xiao, J. Zhou and F. Wang, "Drug Similarity Integration Through Attentive Multi-View Graph Auto-Encoders," in *Proc. AAAI Conference on Artificial Intelligence*, 2018, pp. 347–354.

[7] D. S. Dhami, S. Yan, G. Kunapuli, D. Page and S. Natarajan, "Beyond Textual Data: Predicting Drug-Drug Interactions from Molecular Structure Images Using Siamese Neural Networks," in *Proc. IEEE International Conference on Bioinformatics and Biomedicine*, 2019, pp. 121–128.

[8] D. S. Dhami, S. Yan, G. Kunapuli, D. Page and S. Natarajan, "Predicting Drug-Drug Interactions from Heterogeneous Data: An Embedding Approach," in *Proc. International Joint Conference on Artificial Intelligence*, 2021, pp. 1–8.

[9] K. Yu, S. Visweswaran and K. Batmanghelich, "Semi-Supervised Hierarchical Drug Embedding in Hyperbolic Space," in *Proc. Machine Learning for Health Workshop*, 2020, pp. 1–10.

[10] J. Zhang, Q. Wang and Y. Wang, "Predicting Drug-Drug Interactions Through Network-Based Similarity Measures," *Bioinformatics*, vol. 34, no. 21, pp. 374–380, 2018.

[11] H. Lee, J. Kim and M. Yoon, "A Graph-Based Approach for Predicting Drug–Drug Interactions Using Integrated Drug Similarities," *PLOS ONE*, vol. 14, no. 7, pp. 1–15, 2019.

[12] Y. Deng, M. Xu and Z. Wang, "Multi-Feature Fusion for Drug–Drug Interaction Prediction," *IEEE Access*, vol. 8, pp. 123–132, 2020.

[13] W. Shi, S. Liu and H. Zhang, "Deep Learning Framework for Drug–Drug Interaction Prediction Using Similarity Integration," *Journal of Biomedical Informatics*, vol. 93, pp. 103–112, 2019.

[14] X. Cheng and F. Zhao, "Semi-Supervised Learning for Predicting Potential Drug–Drug Interactions," *Computers in Biology and Medicine*, vol. 112, pp. 103–110, 2019.

[15] P. Vilar, E. Uriarte, L. Santana and R. Hripcsak, "Detection of Drug–Drug Interactions Through Data Mining Studies," *Expert Opinion on Drug Safety*, vol. 17, no. 4, pp. 1–12, 2018.