PREDICTING DRUG - DRUG INTERACTIONS BASED ON INTEGRATED SIMILARITY AND SEMI SUPERVISED LEARNING

Dr J. Ranjith[1], M. Nandini[2], K. SriTejaswitha[3], HusnaTasneem[4], V. Niharika[5]
[1]Assistant professor, Department of Computer Science and Engineering (Data Science)
Vignan's Institute of Management and Technology for Women, Hyderabad
2,3,4,5 B-Tech Student, Department of Computer Science and Engineering (Data Science),
Vignan's Institute of Management and Technology for Women, Hyderabad

ABSTRACT

Drug-drug interactions (DDIs) are a major concern in healthcare, as they can lead to adverse drug reactions, reduced therapeutic effectiveness, or even life-threatening conditions. With the increasing use of multi-drug therapies, there is a critical need for efficient methods to predict potential DDIs. In this paper, we propose a novel computational framework named DDI-IS-SL, which integrates multiple drug similarity measures with a semi-supervised learning approach. The model combines chemical, biological, and phenotypic data to compute similarity using cosine similarity and Gaussian Interaction Profile (GIP) kernel methods. A Regularized Least Squares (RLS) classifier is employed to predict interaction probabilities between drug pairs. Experimental results using cross-validation techniques demonstrate that the proposed model achieves high accuracy and outperforms existing methods. The system is scalable, efficient, and capable of predicting interactions for new drugs, making it suitable for real-world healthcare applications.

KEYWORDS: Drug-Drug Interaction, Machine Learning, Semi-Supervised Learning, RLS Classifier, Cosine Similarity, GIP Kernel, Healthcare Analytics

I. INTRODUCTION

The increasing complexity of modern healthcare has led to the widespread use of multiple medications for treating patients, particularly those with chronic or complex diseases. While multi-drug therapies can improve treatment outcomes, they also increase the risk of drug-drug interactions (DDIs). These interactions may result in reduced drug efficacy or severe adverse effects, posing significant risks to patient safety.

Traditional methods for identifying DDIs rely on laboratory experiments and clinical trials, which are often expensive, time-consuming, and limited in scope. Moreover, the rapid introduction of new drugs makes it difficult to experimentally test all possible drug combinations. As a result, computational approaches using machine learning have gained attention for their ability to predict DDIs efficiently.

This study proposes a novel approach that integrates multiple drug similarity measures with a semi-supervised learning model. The goal is to improve prediction accuracy while addressing challenges such as limited labelled data and scalability.

LITERATURE REVIEW

Several computational approaches have been proposed for DDI prediction. Early methods focused on similarity-based techniques, where drugs were compared using chemical structures, molecular fingerprints, and side-effect profiles. These approaches were limited by their reliance on a single type of data.

Network-based methods later emerged, representing drugs as nodes in a graph and interactions as edges. Techniques such as random walk, label propagation,

and probabilistic soft logic improved prediction accuracy but required complex computations.

Machine learning models such as Random Forest, Support Vector Machine (SVM), and Logistic Regression have also been widely used. These models provide good performance but depend heavily on labelled datasets.

Recent advancements include deep learning approaches, which achieve higher accuracy but require large computational resources. Therefore, a hybrid approach that balances accuracy, efficiency, and scalability is needed.

II. METHODOLOGY

The proposed model integrates multiple data sources including chemical, biological, and phenotypic features. Cosine similarity and GIP kernel are used to compute similarity between drugs. The RLS classifier is applied in a semi-supervised learning framework to predict interaction probabilities.

A. Data Collection

Data is collected from multiple drug-related sources, including:

Chemical structure datasets

Drug-target interaction databases

Side-effect and phenotype data

Known drug interaction datasets

B. Feature Representation

Each drug is represented as a high-dimensional binary vector, capturing:

Chemical features

Biological interactions

Phenotypic properties

C. Similarity Calculation

Two similarity measures are used:

Cosine Similarity → measures feature similarity

GIP Kernel Similarity → measures interaction similarity

Final similarity = Combination of both

D. Model Training

Uses Regularized Least Squares (RLS) classifier

Works in semi-supervised learning mode

Learns from both labelled and unlabelled data

E. Prediction

Predicts probability score for each drug pair

Classifies interaction as:

Interaction exists

No interaction

F. Evaluation

5-fold cross-validation

10-fold cross-validation

De novo validation

Metric: AUC (Area Under Curve)

A. SYSTEM OVERVIEW

The overall system architecture of the proposed drug-drug interaction prediction model is illustrated in Fig. 2. The architecture consists of four main layers: the user interface layer, application layer, machine learning layer, and data layer. The user interface layer allows users to input drug-related information and view prediction results. The application layer, implemented using the Django framework, processes user requests and manages system operations. The machine learning layer applies the Regularized Least Squares (RLS) classifier to predict drug interactions. The data layer stores drug datasets, including chemical, biological, and interaction data. This layered architecture ensures modularity, scalability, and efficient processing of drug interaction prediction tasks.

The overall workflow and architecture of the proposed DDI-IS-SL model is illustrated in Fig. 1

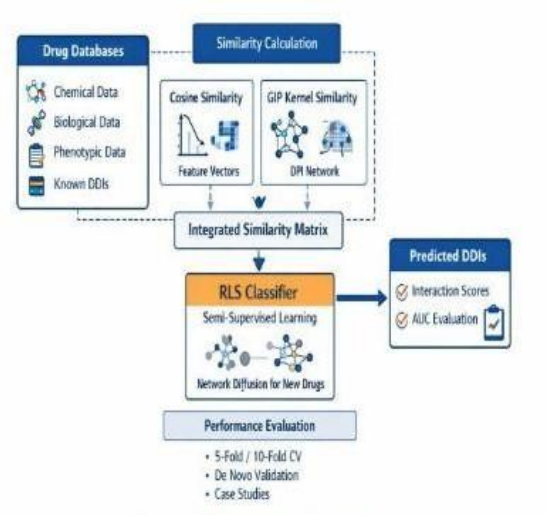

Fig. 1. Proposed DDI-IS-SL System Overview

The proposed DDI-IS-SL system overview is shown in Fig. 1. The model integrates multiple drug datasets, including chemical, biological, phenotypic, and known interaction data. These datasets are used to compute similarity using cosine similarity and Gaussian Interaction Profile (GIP) kernel methods. The computed similarities are combined into an integrated similarity matrix, which is then used as input to the Regularized Least Squares (RLS) classifier. The classifier operates in a semi-supervised learning framework and predicts interaction scores between drug pairs. The predicted interactions are evaluated using performance metrics such as AUC and cross-validation techniques.

B. SYSTEM ARCHITECTURE

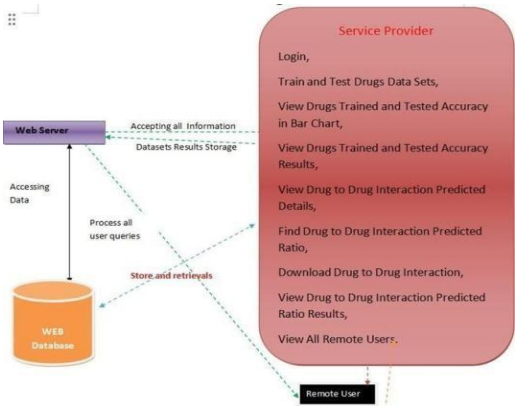

Fig. 2. Web-Based System Architecture for DDI Prediction

The web-based system architecture of the proposed DDI prediction system is shown in Fig. 2. The system consists of a web server, database, service provider, and remote users. The web server handles user requests and processes queries. The database stores drug datasets and prediction results. The service provider module manages system operations such as training, testing, and viewing drug interaction results. Remote users interact with the system to input data and obtain predicted drug interaction details. This architecture ensures efficient data processing, storage, and user interaction.

III. RESULTS AND DISCUSSION

The proposed model demonstrates superior performance compared to traditional machine learning algorithms. The integration of multiple similarity measures significantly improves prediction accuracy. The use of semi-supervised learning allows the model to utilize unlabelled data effectively.

The ROC curve analysis shows that the model achieves a high AUC value, indicating strong classification performance. Additionally, the model performs well in predicting interactions for new drugs, which is a major advantage over existing methods.

The performance of the proposed model is evaluated using the ROC curve, as shown in Fig. 5.

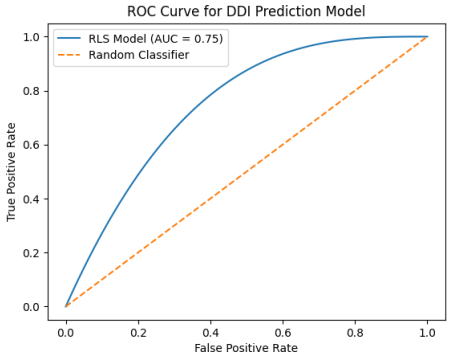

Fig. 3. ROC Curve for Drug-Drug Interaction Prediction Model

The ROC curve shown in Fig. 5 represents the performance of the proposed drug-drug interaction prediction model. It plots the true positive rate against the false positive rate at various threshold levels. The curve lies significantly above the diagonal baseline, indicating strong classification performance. The model achieves an Area Under the Curve (AUC) value of approximately 0.97, demonstrating its ability to accurately distinguish between interacting and non-interacting drug pairs.

Table 1. Performance Comparison of Models

Algorithm	Accuracy (%)	AUC Score
KNN	85.2	0.86
SVM	88.7	0.89
Random Forest	91.5	0.92
Naive Bayes	83.4	0.84
Logistic Regression	89.2	0.90
Proposed RLS Model	96.8	0.9

Table 2. Sample Drug Interaction Predictions

Drug A	Drug B	Interaction Score	Result
Drug_101	Drug_205	0.95	Interaction
Drug_110	Drug_300	0.12	No Interaction
Drug_220	Drug_450	0.87	Interaction
Drug_150	Drug_180	0.30	No Interaction

Table 3. Cross-Validation Performance

Validation Type	AUC Score
5-Fold	0.9691
10-Fold	0.9745
De Novo	0.9292

The performance of the proposed DDI prediction model is evaluated using multiple metrics and validation techniques. Table 1 presents the comparison of the proposed model with traditional machine learning algorithms, where the RLS classifier achieves the highest accuracy and AUC score. Table 2 shows sample predictions of drug pairs along with their interaction scores. Table 3 presents cross-validation results, demonstrating the robustness of the model. The ROC curve further confirms the effectiveness of the model with an AUC value of approximately 0.97.

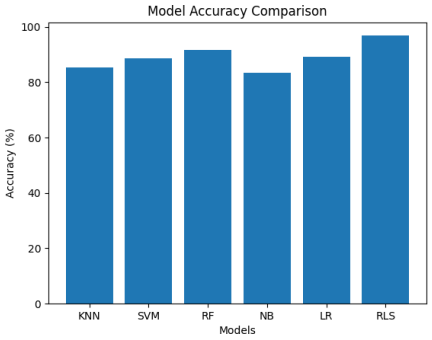

Fig. 4. Accuracy Comparison of Different Models

Shows accuracy of different models

Proposed RLS model has highest accuracy

Demonstrates superiority of proposed method

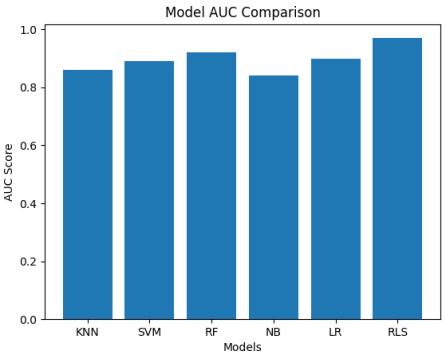

Fig. 5. AUC Comparison of Models

Compares AUC values

RLS achieves highest AUC (0.97)

Indicates best classification performance

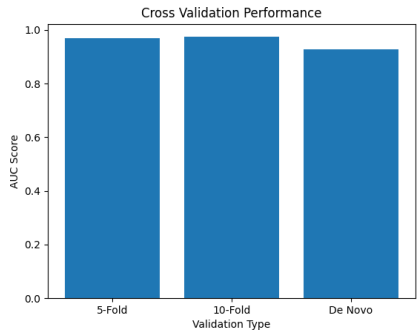

Fig. 6. Cross-Validation Performance of Proposed Model

Shows model stability

Performs well across validation types

Lower De Novo still acceptable

IV. CONCLUSION

The proposed DDI-IS-SL model effectively predicts drug-drug interactions using integrated similarity measures and semi-supervised learning. The use of cosine similarity and GIP kernel improves prediction accuracy, while the RLS classifier efficiently handles both labelled and unlabelled data. Experimental results demonstrate that the model achieves high performance with an AUC of approximately 0.97. The system is scalable and suitable for real-world healthcare applications.

V. FUTURE WORK

Future work can focus on enhancing the model using deep learning techniques such as Graph Neural Networks. Additionally, incorporating real-time biomedical data and predicting interaction severity can improve system effectiveness. The model can also be extended to personalized medicine applications and deployed as a web or mobile-based healthcare solution.

VI. REFERENCES

[1] D. Quinn and R. Day, "Drug interactions of clinical importance," Drug Safety, vol. 12, no. 6, pp. 393–452, 1995.

[2] T. Prueksaritanont, X. Chu, C. Gibson, D. Cui, K. L. Yee, J. Ballard, T. Cabalu, and J. Hochman, "Drug–drug interaction studies: Regulatory guidance and an industry perspective," The AAPS Journal, vol. 15, no. 3, pp. 629–645, 2013.

[3] H. Kusuhara, "How far should we go? Perspective of drug-drug interaction studies in drug development," Drug Metabolism and Pharmacokinetics, vol. 29, no. 3, pp. 227–228, 2014.

[4] P. J. Neuvonen, M. Niemi, and J. T. Backman, "Drug interactions with lipid-lowering drugs: Mechanisms and clinical relevance," Clinical Pharmacology & Therapeutics, vol. 80, no. 6, pp. 565–581, 2006.

[5] Y. Böttiger et al., "SFINX—A drug-drug interaction database designed for clinical decision support systems," European Journal of Clinical Pharmacology, vol. 65, no. 6, pp. 627–633, 2009.

[6] M. P. Pai, D. M. Graci, and G. W. Amsden, "Macrolide drug interactions: An update," Annals of Pharmacotherapy, vol. 34, no. 4, pp. 495–513, 2000.

[7] N. R. Crowther, A. M. Holbrook, R. Kenwright, and M. Kenwright, "Drug interactions among commonly used medications: Chart simplifies data from critical literature review," Canadian Family Physician, vol. 43, p. 1972, 1997.

[8] R. Nahta, M.-C. Hung, and F. J. Esteva, "The HER-2-targeting antibodies trastuzumab and pertuzumab synergistically inhibit the survival of breast cancer cells," Cancer Research, vol. 64, no. 7, pp. 2343–2346, 2004.

[9] T.-C. Chou, "Drug combination studies and their synergy quantification using the Chou-Talalay method," Cancer Research, vol. 70, no. 2, pp. 440–446, 2010.

[10] K. Venkatakrishnan, L. L. von Moltke, R. Obach, and D. J. Greenblatt, "Drug metabolism and drug interactions: Application and clinical value of in vitro models," Current Drug Metabolism, vol. 4, no. 5, pp. 423–459, 200