Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 16, Issue. 1: 2025
ISSN : **1906-9685**

DEEP SIDE: A DEEP LEARNING FRAMEWORK FOR DRUG SIDE EFFECT PREDICTION

Mrs. J. KUMARI[1], SHAIK. NURAIN[2]

#1 Assistant Professor Department of Master of Computer Applications e-Mail: j@qiscet.edu.in
#2 Pursuing M.C.A e-Mail: j@qiscet.edu.in
QIS COLLEGE OF ENGINEERING & TECHNOLOGY
Vengamukkapalem(V), Ongole, Prakasam dist., Andhra Pradesh- 523272

Abstract

Drug failures due to unforeseen adverse effects during clinical trials pose significant health risks for participants and lead to substantial financial losses. Side effect prediction algorithms have the potential to guide the drug design process and mitigate these risks. The LINCS L1000 dataset provides a vast resource of cell line gene expression data perturbed by different drugs and serves as a knowledge base for context-specific features. The state-of-the-art approach that aims to use context-specific information relies only on the high-quality experiments in LINCS L1000 and discards a large portion of the available data. In this study, our goal is to boost prediction performance by utilizing the dataset to its full extent. We experiment with five deep learning architectures. We find that a multi-modal architecture produces the best predictive performance among multi-layer perceptron-based models when drug chemical structure (CS) and the full set of drug-perturbed gene expression profiles (GEX) are used as modalities. Overall, we observe that the CS modality is more informative than the GEX. A convolutional neural network-based model that uses only the SMILES string representation of the drugs achieves the best results, providing a 130% improvement in macro-AUC and a 31% improvement in micro-AUC over the state-of-the-art. We also show that the model is able to predict side effect–drug pairs that are reported in the literature but were missing from the ground truth side effect.

Introduction:

Computational methods hold great promise for mitigating the health and financial risks of drug development by predicting possible side effects before entering clinical trials. Several learning based methods have been proposed for predicting the side effects of drugs based on various features such as: chemical structures of drugs [25, 1, 23, 8, 19, 34, 17, 9, 2, 5], drug–protein interactions [35, 33, 8, 19, 34, 17, 37, 2, 15, 36], protein–protein interactions (PPI) [8, 9], activity in metabolic networks [38, 26],

pathways, phenotype information, and gene annotations [8]. In parallel to the above-mentioned approaches, recently, deep learning models have been employed to predict side effects: (i) [31] uses biological, chemical, and semantic information on drugs in addition to clinical notes and case reports; and (ii) [4] uses various chemical fingerprints extracted using deep architectures to compare side effect prediction performance. While these methods have proven useful for predicting adverse drug reactions (ADRs—used interchangeably with drug side effects), the features they use are solely based on external knowledge about the drugs (e.g., drug–protein interactions) and are not cell- or condition specific (i.e., dosage). To address this issue, Wang et al. (2016) utilized data from the LINCS L1000 project [32]. This project profiles gene expression changes in numerous human cell lines after treatment with a large number of drugs and small-molecule compounds. By using the gene expression profiles of the treated cells, [32] provides the first comprehensive, unbiased, and cost-effective prediction of ADRs. The paper formulates the problem as a multi-label classification task. Their results suggest that gene expression profiles provide context-dependent information for the side effect prediction task

Multi-layer perceptron (MLP) Our MLP [22] model takes the concatenation of all input vectors and applies a series of fully-connected (FC) layers. Each FC layer is followed by a batch normalization layer [10]. We use ReLU activation [16], and dropout regularization [27] with a drop probability of 0:2. The sigmoid activation function is applied to the final layer outputs, which yields the ADR prediction probabilities. The loss function is defined as the sum of negative log- probabilities over ADR classes, i.e. the multi-label binary cross-entropy loss (BCE). An illustration of the architecture for CS and GEX features is given in this system. Residual multi-layer perceptron (ResMLP) The residual multi-layer perceptron (ResMLP) architecture is very similar to MLP, except that it uses residual-connections across the fully-connected layers. More specifically, the input of each intermediate layer is element-wise added to its output, before getting processed by the next layer. This effectively allows deeper architectures, therefore, potentially learning more complex and parameter-efficient feature extractors. Multi-modal neural networks (MMNN) The multi modal neural network approach contains distinct MLP sub-networks where each one extract features from one data modality only. The outputs of these sub-networks are then fused and fed to the classification block. For feature fusion, we consider two strategies: concatenation and summation. While the former one concatenates the domain-specific feature vectors to a larger one, the latter one performs element-wise summation. By definition, for summation based fusion, the domain-specific feature extraction sub-networks have to be designed to produce vectors of equivalent sizes. We refer to the concatenation and summation based MMNN networks as MMNN.Concat and MMNN.Sum, respectively.

1. DEEPSIDE: Enhancing Drug Side Effect Prediction with Deep Learning

Authors: K. Priyanka and Dr. Dhanaraj Cheelu (2025)

Description: Compares five deep learning architectures using chemical structure (CS) and gene expression (GEX) modalities from LINCS L1000 data. Finds that a convolutional neural network using SMILES strings outperforms prior approaches.

Metrics: Achieves +13.0% macro-AUC and +3.1% micro-AUC improvement over the state-of-the-art

2. Modeling Polypharmacy Side Effects with Graph Convolutional Networks (Decagon)

Authors: Marinka Zitnik, Monica Agrawal, Jure Leskovec (2018)

Description: Constructs a multimodal graph combining protein–protein interactions, drug–protein interactions, and polypharmacy side effects. Applies a novel graph convolutional neural network to predict exact side effects from drug pairs.

Metrics: Outperforms baselines by up to 69% in accuracy for predicting polypharmacy side effects

3. A Deep Learning Approach to the Prediction of Drug Side-Effects on Molecular Graphs

Authors: Pietro Bongini, Elisa Messori, Niccolò Pancino, Monica Bianchini (2022)

Description: Leverages recurrent graph neural networks to exploit molecular graph structure of drugs for multi-label side effect prediction. Builds a knowledge graph–based dataset from open data sources.

Metrics: Reports improved classification performance across multiple metrics and parameters compared to prior predictors (specific metrics implied but not detailed in abstract)

4. Drug Side Effect Prediction with Deep Learning Molecular Embedding in a "Graph-of-Graphs" Domain (MolecularGNN)

Description: Introduces a "graph-of-graphs" molecular embedding approach to predict drug side effects, comparing to prior models such as DruGNN, DeepSide, and others.

Metrics: Shows significantly higher F1-score and improvements in recall; introduces PatR (precision at recall 90%), observing increases in AUC, F1-score, recall, and PatR

5. A Deep Learning Framework for Multi-Drug Side Effects Prediction with Drug Chemical Substructure (DLMSE)

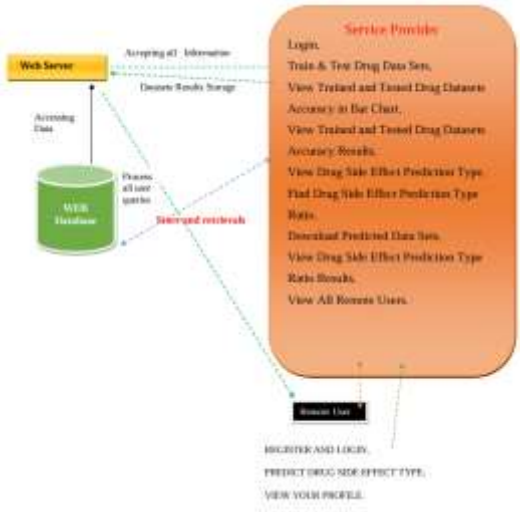
Authors: Muhammad Asad Arshed, Shahzad Mumtaz, Omer Riaz, Waqas Sharif, Saima Abdullah (2022)

Description: Proposes DLMSE, a multi-label deep learning model that uses chemical structure to predict multiple side effects per drug, focusing on dizziness, allergy, and headache.

Metrics: Achieves accuracy of 0.9494 in predicting the three chosen side effects

Paper / Model	Key Innovation	Reported Metrics
DEEPSIDE	SMILES-based CNN using multi-modal data	+13.0% macro-AUC, +3.1% micro-AUC
Decagon	GCN on multimodal graph (PPI, DDI, drug targets)	Up to 69% improvement over baselines
Molecular GNN	Graph-of-graphs molecular embedding	Higher F1, recall improved AUC
GNN on Graphs*	Multi-label GNN on molecular graphs	Improved classification across metrics
DLMSE	CS-based multi-label side effect predictor	94.94% accuracy

System Architecture:



Module Description

1. Input Module:

• Description: This module handles the input of necessary data, including SMILES strings representing the chemical structure of drugs and gene expression profiles (GEX) from datasets like LINCS L1000.

• Functionality: o Accepts and validates drug structure input. o Loads gene expression data. o Supports manual or automated data input. 2. Dataset Processing Module

• Description: Responsible for cleaning, transforming, and preparing the input data for training. • Functionality: o Normalizes gene expression values. o Converts SMILES strings to vector representations (e.g., molecular fingerprints or embeddings). o

Splits data into training, validation, and test sets 3. Model Training and Prediction Module

• Description: Core module where deep learning models are trained to predict drug

side effects. • Functionality: Supports multi-modal input (chemical structure + gene expression). o Trains models like MLP, CNN, or multi-modal architectures. o Generates multi-label predictions of side effects. o Evaluates performance using metrics like macro-AUC and micro-AUC.

Storage and Retrieval Module (Store and Retrievals)

• Description: Manages storing and retrieving datasets, trained models, and prediction results. • Functionality: Saves trained model checkpoints and logs.

Retrieves past results and performance metrics. o Allows reloading models for future use.

5. Evaluation & Visualization Module

Description: Provides tools to assess the model's performance and visualize the results for analysis.

Functionality: Calculates evaluation metrics (e.g., accuracy, macro/micro-AUC, precision, recall). Plots training loss, ROC curves, confusion matrices. Compares multiple model performances.

6. User Interface:

• Description: If a front-end or basic UI is provided, this module allows users to interact with the system easily.

• Functionality: o Interface to upload input files (SMILES or GEX). o View predicted side effects. o Access visual results and performance summary.

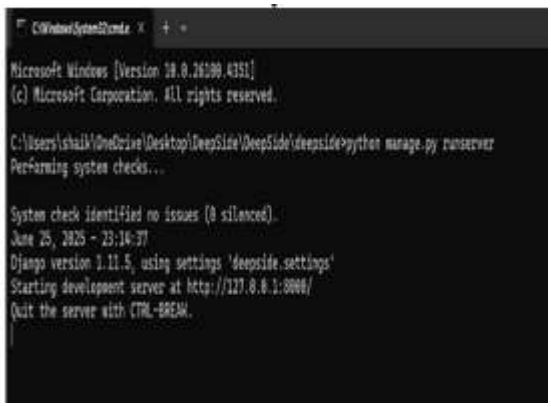
7. Logging and Monitoring Module

• Description: Tracks system performance, errors, and workflow steps for transparency and debugging.

• Functionality: Logs data loading, model training, and inference events. Captures errors and warnings during execution. Stores system usage stats for audit or debugging.

Implementation:

Click cmd to start the Python server and access the page below.



In above screen python server started and now open browser and enter URL as http://127.0.0.1:8000/ and press enter key to get below page

In above screen user is login and after login will get below page

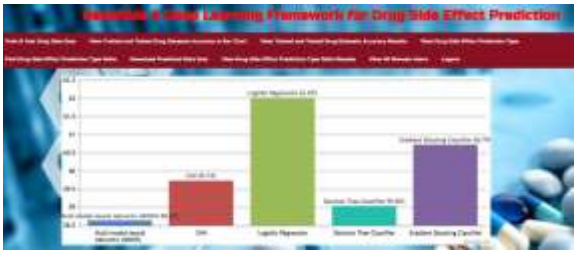
In above screen enter "UId and drug name and enter condition" with other drugs

After entering the details website will predict in high/low ratio as below page

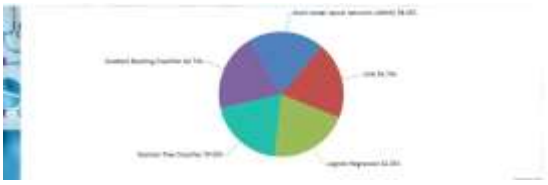
Predicted as "HIGH SIDE EFFECT FOUND" in downside can see. Remote User can only predict the side effect Service Provider can see more as below

Service Provider can see info about all remote user and other data such as below pages

Performance metrics



View trained and Tested Drug datasets accuracy in "BAR CHART"



View trained and Tested Drug datasets accuracy in "Pie chart"

CONCLUSION AND FUTURE ENHANCEMENT

The pharmaceutical drug development process is a long and demanding process. Unforeseen ADRs that arise at the drug development process can suspend or restart the whole development pipeline. Therefore, the a priori prediction of the side effects of the drug at the design phase is critical. In our Deep Side framework, we use context-related (gene expression) features along with the chemical structure to predict ADRs to account for conditions such as dosing, time interval, and cell line. The proposed MMNN model uses GEX and CS as combined features and achieves better

accuracy performance compared to the models that only use the chemical structure (CS) finger- prints. The reported accuracy is noteworthy considering that we are only trying to estimate the condition-independent side effects. Finally, SMILES Conv model outperforms all other approaches by applying convolution on SMILES representation of drug chemical structure. Future Enhancement

1. Integration of Multi-Omics Data: o Incorporate additional biological data such as proteomics, metabolomics, or epigenetics to improve prediction accuracy and model robustness.

2. Real-Time Drug Monitoring: o Develop APIs or a web-based interface that integrates real-world post-market surveillance data to continuously update and refine predictions.

3. Explainability and Interpretability: o Use explainable AI (XAI) techniques such as SHAP or LIME to interpret the model's predictions, helping researchers understand why a drug causes a certain side effect.

4. Transfer Learning and Pre-trained Models: o Implement transfer learning using pre-trained models on related biological tasks to improve performance on smaller datasets or rare side effects.

5. Adverse Interaction Prediction: o Extend the framework to predict drug–drug interactions and their combined side effects, which is critical for polypharmacy scenarios.

6. Mobile and Cloud Deployment: o Deploy the model as a lightweight mobile or cloud-based application for clinical and research use in real-time settings.

REFERENCES

1. Atias, N., Sharan, R.: An algorithmic framework for predicting side e_ects of drugs. Journal of Computational Biology 18(3), 207{218 (2011)

2. Bresso, E., Grisoni, R., Marchetti, G., Karaboga, A.S., Souchet, M., Devignes, M.D., Sma•_l-Tabbone, M.: Integrative relational machine-learning for understanding drug side-e_ect pro_les. BMC bioinformatics 14(1), 207 (2013)

3. Cai, M.C., Xu, Q., Pan, Y.J., Pan, W., Ji, N., Li, Y.B., Jin, H.J., Liu, K., Ji, Z.L.: Adrecs: an ontology database for aiding standardization and hierarchical classi_cation of adverse drug reaction terms. Nucleic acids research 43(D1), D907{D913 (2014)

4. Dey, S., Luo, H., Fokoue, A., Hu, J., Zhang, P.: Predicting adverse drug reactions through interpretable deep learning framework. BMC Bioinformatics 19 (12 2018). https://doi.org/10.1186/s12859-018 2544-0

5. Dimitri, G.M., Li_o, P.: Drugclust: A machine learning approach for drugs side effects prediction. Computational Biology and Chemistry 68, 204 { 210 (2017). https://doi.org/https://doi.org/10.1016/j.co mpbiolchem.2017.03.008

6. Groopman, J.E., Itri, L.M.: Chemotherapy-induced anemia in adults: incidence and treatment. Journal of the

National Cancer Institute 91(19), 1616{1634 (1999)

7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2016)

8. Huang, L.C., Wu, X., Chen, J.Y.: Predicting adverse side e_ects of drugs. BMC genomics 12(5), S11 (2011)

9. Huang, L.C., Wu, X., Chen, J.Y.: Predicting adverse drug reaction pro_les by integrating protein interaction networks with drug structures. Proteomics 13(2), 313{324 (2013)

10. Io_e, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)

11. Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A., et al.: Pubchem substance and compound databases. Nucleic acids research 44(D1), D1202{D1213 (2015)

12. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classi_cation with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097{1105 (2012)

13. Kuhn, M., Letunic, I., Jensen, L.J., Bork, P.: The sider database of drugs and side e_ects. Nucleic acids research 44(D1), D1075{D1079 (2015)

14. Landrum, G., et al.: Rdkit: Open-source cheminformatics (2006) 15. Lee, W., Huang, J., Chang, H., Lee, K., Lai, C.: Predicting drug side e_ects using data analytics and the integration of multiple data sources. IEEE Access 5, 20449{20462 (2017). https://doi.org/10.1109/ACCESS.2017.2755045

AUTHOR

Mrs. Jasti kumari is an Assistant Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. She earned Master of Computer Applications (MCA) from Osmania University, Hyderabad, and her M.Tech in Computer Science and Engineering (CSE) from Jawaharlal Nehru Technological University, Kakinada (JNTUK). Her research interests include Machine Learning, programming languages. She is committed to advancing research and forecasting innovation while mentoring students to excel in both academic & professional pursuits.

Shaik.Nurain is an MCA Scholar, Dept. of MCA, In QIS College of Engineering & Technology, Ongole. His areas of interest are

Machine Learning, Deep Learning.