Indian Journal of Computer Science and Technology
https://www.doi.org/10.59256/indjcst.20260502078
Volume 5, Issue2 (May-August 2026), PP: 672-678.
www.indjcst.com
ISSN No: 2583-5300

Machine Learning-Based Prediction of Adverse Drug Reactions from Protein Interaction Profiles

Pratima Walia[1], Agrimaa Singh Thakur[2]*
[1]*Student, Master of Computer Applications, Indira Gandhi National Open University (IGNOU), New Delhi, India.*
[2]*Assistant Professor, Department of Computer Science and Engineering, Baddi University of Emerging Sciences and Technology, Baddi, Himachal Pradesh, India.*

To Cite this Article: *Pratima Walia[1], Agrimaa Singh Thakur[2]*, "Machine Learning-Based Prediction of Adverse Drug Reactions from Protein Interaction Profiles", Indian Journal of Computer Science and Technology, Volume 05, Issue 02 (May-August 2026), PP: 672-678.*

Abstract: *The ability to predict side effects of drugs is an important problem with direct implications on patient health and the pharmaceutical industry's ability to develop new drugs. Unpredictable adverse drug reactions (ADR) have caused many drugs to be pulled from the market and have caused more than 100 million injuries or deaths worldwide. In this article, I will be using data from the Mizutani dataset, which has 658 drugs and 1368 target protein structures, to predict drug side effects using the targeted protein interaction profiles as the primary features. A Naive Bayes (NB) model with all 1339 side effects as output variables was created; and the classification accuracy of the NB classifier was 91.14% in the Mizutani dataset as a whole. In addition, 15 of the more commonly reported side effects will be identified; and the 15 selected side effects will be subjected to 5 machine learning classifiers (Logistic Regression (LR), Random Forest (RF), Naive Bayes (NB), Support Vector Machine (SVM), and Decision Trees (DT)) to compare their respective accuracies, precisions, and F-scores. The classifier with the highest average accuracy of 70.29% across the selected 15 side effects was the LR classifier. In addition to this work, the proposed method will also be used to predict side effect profiles of 92 previously uncharacterized drugs using the LR classifier; and the LR classifier achieved a 76.45% match with publicly available reference sources for those drugs. These results indicate that targeted interaction protein interaction profiles provide sufficient and useful predictive features for predicting drug side effects.*

Keywords: *Drug side effect prediction, adverse drug reactions, targeted protein structures, machine learning.*

I. INTRODUCTION

Adverse drug reactions (ADRs) and drug side effects represent one of the greatest concerns for pharmacotherapy and clinical medicine. WHO defines ADR as a noxious and unintended reaction to a medication that occurs at doses typically used on human beings for the prevention, diagnosis and/or treatment of disease[1]. An important reason for drug failure in the development process and removal from the market due to ADRs, 28 drugs that produced a significant ADR have been removed from the market in the U.S. between 1976 and 2005[2]. According to estimates from the CDC, approximately 112 thousand deaths in the United States were due to drug overdoses in 2025. These statistics indicate an increase in the effects of adverse drug events and drug toxicity on public health; therefore, these two categories will be some of the largest contributors to mortality in the U.S. [3].

The concept of drug toxicity refers to cases where patients ingest an excess amount of medication within their circulation system leading to adverse reactions. Pharmacovigilance is the practice involved in identifying, assessing, evaluating, and preventing the adverse effects of medicines. The FDA process of drug development involves five stages: discovery and development, pre-clinical research, clinical research, FDA review, and post-market safety monitoring. The majority of serious side effects usually remain undiscovered during the post-marketing surveillance stage, leading to unhealthy conditions and substantial financial losses for the pharmaceutical sector [4].

Earlier methods of drug side effect predictions relied on the analysis of structure activity relationship, however, these could not be used on large biomedical databases [5–8]. Later, machine learning emerged as a paradigm that could effectively predict drug side effects due to its capability of discovering complex and non-linear relationships between high dimensional biological and chemical data. What makes machine learning unique is its versatility, i.e., virtually any drug feature can be fed into the system, along with a myriad of classification algorithms.

The present paper proposes a machine learning-based approach to predict drug side effects from the targeted protein structures that serve as primary drug features. First, a Naïve Bayes classifier is trained for all 1339 side effects of the Mizutani database. Then, fifteen side effects frequently occurring in the literature are selected, and five classifiers with different kernels are analyzed in terms of accuracy, F-score and precision. Additionally, the proposed approach is used to predict side effects of 92 drugs whose side effects were not known before, validating the results against other data sources. The novel contribution of the study can be highlighted as follows: (i) Comparison of machine learning classifiers for drug side effect prediction based on protein interactions; (ii) Identification of the best models for each specific side effect; and (iii) First systematic prediction of side effects for uncharacterized drugs.

II.RELATED WORK

Yan et al. [9] created a new multi-kernel learning and clustering methodology, referred to as MKLC-BiRW, which utilized bi-random walking on drug-target networks (DTNs) with heterogeneous data. The authors used two datasets (1923 drugs and a subset with ≤40% sequence identity) to evaluate their work and reported an area under the curve (AUC) of 93.9% for drug-target interaction prediction and an area under the precision-recall curve (AUPR) of 95.7%. These metrics were superior to the five previously published drug-target interaction (DTI) prediction studies they cited.

Xian et al. [10] produced a random forest classifier utilizing five different drug characteristics derived from the SIDER database (841 drugs and 824 side effects). The five characteristics included fingerprint similarity, structural similarity, ATC codes, literature association, and target proteins. The best random forest performance metrics reported are as follows: Matthew's correlation coefficient (MCC) = 0.464; accuracy (ACC) = 0.732; F1 = 0.733; and AUC = 0.8015. The most predictive feature for drug side effects was determined to be drug fingerprint similarity.

Niu et al. [11] developed a quantitative tool for predicting adverse drug reactions (ADRs) using random forest regression based on the combination of features from the Pauwels, Liu, and Mizutani datasets. Results showed that the average-scoring ensemble of combined features outperformed direct combination of features for every metric (R^2, MAE, RMSE), confirming that multi-feature ensembles outperform individual features combined into a single model.

DrugClust [12] is an R package which uses the K-Means, PAM, and K-Seeds algorithm along with Bayesian scoring on the datasets of Liu, Mizutani, and Zhang. The model has performed well in terms of obtaining a higher AUC and AUPR compared to previous works, and biological validation of drug clusters was done through pathway enrichment analysis.

LNSM-MSE [13] is a linear neighbour similarity-based technique for predicting novel side effects of the existing drugs based on the Liu, Mizutani, and Pauwels datasets. An AUPR score of 0.67 and an AUC of 0.94 were attained, which exceeded that of other benchmark methods.

Onay et al. [14] used SVM and ensemble classifiers (boosting and bagging of trees) to classify approved and withdrawn drugs for nervous system on six different sub-datasets from DrugBank, PubChem, and KEGG databases. Accuracy of classification varied from 74% to 89%, and the ToxPrint Chemotypes gave discriminating molecular fragments.

Using KNN, RF, and Naive Bayes with Hamming-distance-based drug proximity, Lee et al.[15] examined 1002 drugs, 3903 side effects, and 7257 features in DrugBank, SIDER, UniProt, and PubChem. All models reported high accuracy, but low precision and F measure highlighted the effects of class imbalance.

The Optimized Drug Similarity Framework (OSDF) developed by Zheng et al.,[16] combined chemical similarity (structure and therapeutic classifications), target proteins, and substitutions. They evaluated the clustering of 917 drugs and 500 side effects using SVM, RBF, ELM, and KNN to estimate performance for each model. The maximum macro F1-score was 0.629 (RBF), showing that optimized similarity outperformed raw similarity.

Zhang et al. [17] developed the RBMBM and INBM methods for drug-se term prediction for each of the Liu, Mizutani, and Pauwels datasets. For 5-fold cross-validation, INBM resulted in an ACC of 0.961/0.956/0.959 and RBMBM of 0.958/0.954/0.957 across the 3 datasets, establishing that both methods outperformed all previous methods.

In a similar study, Liu et al. [18] compared the performance of different classification algorithms (NB, KNN, SVM, RF, and LR) across the phenotypic, biological, and chemical features derived from DrugBank, SIDER, UniProt and PubChem for 888 drugs and 1385 side effects. The classifier with the best performance, SVMs, presented with phenotypic AUC of 0.9542, precision of 0.6617, and F1 of 0.6250. Furthermore, the phenotypic features were determined to be the most predictive.

Zitnik et al. [19] introduced a graph convolutional network (GCN) framework to predict polypharmacy side effects through modelling drug-drug interactions as multimodal graphs based on DrugBank and SIDER databases. This GCN performed considerably well against other approaches such as matrix factorisation and shallow machine learning methods.

Wang et al. [20] introduced a novel feature fusion model utilising chemical substructure features, protein targets, and drug-drug interactions through stacking autoencoder with deep neural network classifier (>900 drugs and >1000 side effects). Feature fusion model obtained higher precision and F-score as compared to individual feature models.

Hu et al. [21] proposed a multi-label CNN based on molecular fingerprint and protein target profile information for drug side effect prediction with class cost-sensitive learning due to skewed training set. Multi-label CNN resulted in superior AUC, AUPR and F1 values compared to RF and SVM methods.

Li et al. [22] introduced an effective low-rank matrix completion framework to infer drug side effect associations from SIDER database based on chemical structure and target protein similarities. The method achieved AUC of 0.92 with exceptional performance on rare side effects.

The authors of Zhang et al. [23] present the Graph Attention Network (GAT), which is an attention-based model, for predicting adverse drug reactions (ADR) on heterogeneous drug, drug-target and drug-side effect relational graph databases. By utilizing attention mechanisms, GAT provides insight into the contributions of each drug-target-side effect pathway to the predicted ADR and also achieves higher macro F1-score and area under precision recall curve (AUPR) than GCN or similar-based approaches.

Wu et al. [24] developed a machine learning-based method for drug and drug-target prediction through the integration of drugs, targets, pathways and ADRS, using a Knowledge Graph (KG) for inference. In addition to improving prediction accuracy by incorporating a KG for knowledge embedding, this method also captures latent relationships between drugs and drug attributes that are not always known, especially when limited ADR reports are available for a particular drug.

By combining multiple weights of nodes in a heterogeneous drug-target-side effect network structure derived from a Random Walk with Restart (RWR) diffusion model and subsequently performing a statistical evaluation of the AUC and AUPR of both the overall and indirect multi-hop associations, Cheng et al. [25] outperform bipartite graph models when predicting ADR.

The authors of Yang et al. [26] developed an explainable deep learning framework to combine deep neural networks with confidence scores as determined using SHapley Additive exPlanations (SHAP) to derive feature attribution, thus enhancing model interpretability while achieving high predictive performance and identifying substructures of molecules contributing to respective ADR's.

Zhao et al. [27] introduced a transformer model that learned a representation using SMILES for drugs and sequences of proteins using self-attention. The approach surpassed the performances of CNN and RNN approaches when predicting rare and serious side effects with attention being able to indicate important molecular structures.

The work by Liu et al. [28] was an application of multitask learning where the shared representations were learned from different side effect prediction tasks on two side effect databases: DrugBank and SIDER. This work achieved better results in terms of F1 and macro-F1 measures than the single task counterparts.

Chen et al. [29] leveraged a self-supervised contrastive pretraining scheme before fine-tuning using a small amount of labels to achieve competitive results.

III.MATERIALS AND METHODS

Dataset

Three benchmark datasets are widely used for drug side effect prediction research. The Pauwel dataset (2011) contains drugs, side effects, and chemical substructures. The Mizutani dataset (2012) additionally includes targeted protein structures. The Liu dataset (2012) further includes transporter, pathway, and enzyme information. Table 1 summarises these datasets.

Dataset	Drugs	Side Effects	Chemical Substructures	Targeted Protein Structures
Pauwl	888	1385	881	—
Mizutani	658	1339	881	1368
Liu	832	1385	881	768

Table 1. Details of the benchmark datasets

The Mizutani [30] dataset has been utilized in this study. Drug-protein interactions sourced from DrugBank [31] and MATADOR [32]; side effect source stored in SIDER [33]; and chemical substructure derived from PubChem [34]. The dataset encompasses 658 drugs, 881 chemical substructures, 1339 side effects, and 1368 target protein structures for a total of 5074 drug-protein interaction and 49051 drug-side effect associations. The dataset is stored in binary coded data; the drug-protein interaction is stored in a matrix C (658x1368) and the drug-side effect association is stored in matrix S (658x1339). If there is a drug-protein interaction or drug-side effect association, it is coded as a1, if there is not, it will code as 0.

Machine Learning Classifiers

Logistic Regression (LR): A supervised learner that predicts the probability of being a member of a class using the sigmoid function. Very effective in binary classification problems with high dimensional sparse data.

Naive Bayes (NB): A probabilistic classifier using Bayes rule, assuming feature independence. Tested three different kernels: Multinomial, Gaussian and Bernoulli. The Bernoulli Naïve Bayes algorithm is particularly suited to binary vectors of features.

Decision Trees (DT): A non-parametric classifier that splits the feature space into binary partitioned areas. The result is an interpretable classification tree.

Support Vector Machine (SVM): A classifier that separates the classes with an optimal hyper-plane in the feature space. Two different kernel functions have been tried: Radial Basis Function (RBF), Linear.

Random Forest (RF): An ensemble learning technique combining the results of decision trees trained on randomly selected sets of features.

Evaluation Metrics

Classifier performance was evaluated based on three standard evaluation criteria used for classification tasks: accuracy, precision, and F-score. The first criterion, which describes the share of correct predictions, is called accuracy. The second criterion evaluates how many of the predicted positive examples are indeed true. The third criterion is known as the F-score and is obtained by calculating the harmonic mean of precision and recall.

IV.RESULTS AND DISCUSSION

Baseline Model Results

A Naive Bayes classification model was separately trained for each one of the 1339 side effects, based on the entire set of 658 drugs and their corresponding 1368 protein structures. The output of this modeling is summarized in Table 2 below. It was found that the model was able to achieve an overall accuracy of 91.14% but only had a precision of 45.3% and an F-score of 49.6%. The huge difference between the accuracy values and the other two measures can be attributed to class imbalance.

Method	Accuracy	Precision	F-Score
NB (Baseline)	91.14%	45.3%	49.6%

Table 2. Results of the proposed baseline Naive Bayes method (all 1339 side effects)

Overview: Model Performance Comparison

Five classifiers with multiple kernel types, totaling eight model types, were considered for the top fifteen side effects. The average accuracy scores of the different models for the top fifteen side effects are plotted in Figure 1. From the graph, it can be observed that Logistic Regression (LR) scored an average accuracy score of 70.29%, followed by SVM (RBF) and Random Forest with 66.15% and 63.92%, respectively. Amongst Naïve Bayes algorithms, BernoulliNB performed better than the other two types (MultinomialNB and GaussianNB), achieving an average accuracy score of 63.77% compared to MultinomialNB's 62.57% and GaussianNB's 53.15%.

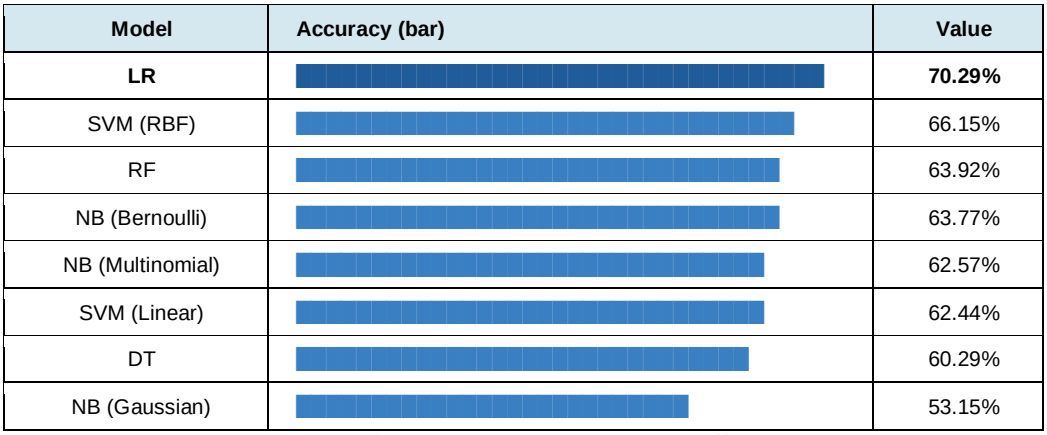

Figure 1. Comparative bar chart of average accuracy across 15 side effects by model (higher is better)

Table 3 further consolidates the comparative accuracy picture across all 15 side effects and all eight model variants. The ★ symbol marks the best-performing model for each side effect.

Side Effect	NB-Mn	NB-Gn	NB-Bn	LR	DT	RF	SVM-R	SVM-L
Nausea	75.05%	55.31%	81.34%	82.43%	72.67%	80.04%	★85.03%	78.74%
Vomiting	73.75%	58.79%	73.75%	76.57%	68.76%	72.89%	★78.96%	72.45%
Headache	72.02%	63.34%	75.92%	★80.04%	74.62%	77.87%	79.39%	77.22%
Rash	62.04%	46.85%	68.11%	68.55%	60.95%	64.86%	★72.02%	60.74%
Dizziness	66.59%	56.18%	68.98%	71.80%	62.04%	70.07%	★72.89%	70.28%
Diarrhea	62.69%	50.76%	67.25%	69.41%	63.77%	67.68%	★72.89%	63.34%
Pruritus	62.26%	49.46%	63.77%	65.08%	56.40%	63.34%	★65.29%	59.00%
Urticaria	59.22%	55.75%	★60.74%	57.48%	55.31%	58.13%	59.00%	55.75%
Somnolence	★63.12%	52.06%	62.69%	60.74%	57.92%	59.44%	★63.12%	59.22%
Constipation	61.61%	55.53%	61.39%	63.12%	62.26%	63.34%	★63.99%	60.95%
Edema	59.22%	48.59%	★61.39%	60.30%	56.83%	60.30%	59.44%	58.13%
Fever	57.27%	47.29%	57.27%	59.00%	56.62%	58.79%	★61.39%	57.48%
Anorexia	★59.44%	53.15%	50.11%	57.48%	52.06%	56.83%	53.58%	59.22%
Paresthesia	55.53%	55.10%	54.23%	57.05%	53.58%	★57.92%	54.01%	54.66%
Fatigue	48.81%	49.02%	49.67%	50.33%	50.54%	47.29%	★51.19%	49.46%

Table 3. Comparative accuracy of all models across 15 side effects (★ = best per side effect)

Results Grouped by Classifier

Comparison among all eight classifiers based on fifteen commonly observed adverse drug reactions showed considerable variations in terms of accuracy. Among all the classifiers, the one with the best average accuracy of 70.29% was logistic regression, followed by Support Vector Machine with RBF kernel having an average accuracy of 66.15% and the highest accuracy for some particular adverse drug reactions such as nausea, vomiting, rash, dizziness, diarrhea, and fever. The high accuracy of Support Vector Machine with the RBF kernel compared to the linear kernel shows that there are nonlinear relationships between targeted proteins and adverse drug reactions.

Among all Naive Bayes classifiers used in the analysis, the Bernoulli Naive Bayes performed better than other classifiers Multinomial and Gaussian Naive Bayes with an average accuracy of 63.77%. This is quite expected because of the representation of targeted proteins interactions in a dataset using binary values.

The random forest method yielded a mean accuracy rate of 63.92%, which was consistently better than decision trees and emphasized the superiority of ensemble algorithms. The higher variability and accuracy levels of decision tree methods suggest that this classification algorithm is prone to overfitting.

Overall, logistic regression and SVM (RBF) were found to be the most promising classifiers when predicting drug adverse events based on protein interactions. Comparisons between different classification models' performances for all chosen adverse events have been provided in Table 3, whereas the results of individual adverse effects can be seen in Table 4.

Best Model Per Side Effect

Table 4 summarizes the most accurate machine learning algorithm per side effect (n=15) along with precision and F-score. SVM (RBF) performed best for seven side effects, whereas LR performed best for six, RF performed best for one, and NB (Bernoulli) performed best for one. Of all the algorithms examined, none were universally superior, indicating that the selection of a model for each side effect is essential in developing a pipeline for predicting side effects of pharmacological treatments.

Side Effect	Best Model	Accuracy	Precision	F-Score
Nausea	SVM (RBF)	85.03%	85.19%	91.89%
Vomiting	SVM (RBF)	78.96%	79.82%	88.01%
Headache	LR	80.04%	80.85%	88.75%
Rash	SVM (RBF)	72.02%	72.53%	83.65%
Dizziness	SVM (RBF)	72.89%	73.42%	83.91%
Diarrhea	SVM (RBF)	72.89%	74.29%	83.31%
Pruritus	SVM (RBF)	65.29%	66.30%	78.89%
Urticaria	NB (Bernoulli)	60.74%	63.13%	69.06%
Somnolence	NB (Multinomial)	63.12%	67.42%	67.68%
Constipation	SVM (RBF)	63.99%	66.46%	72.05%
Edema	NB (Bernoulli)	61.39%	62.81%	73.75%
Fever	SVM (RBF)	61.39%	62.57%	71.10%
Anorexia	NB (Multinomial)	59.44%	68.46%	52.17%
Paresthesia	RF	57.92%	57.47%	56.70%
Fatigue	SVM (RBF)	51.19%	50.21%	50.98%

Table 4. Summary: Best performing model per side effect

As can be seen from Fig.1 and confirmed in each of the per-model tables, SVM (RBF) is best suited to cases of high prevalence, in which the presence of sufficiently many examples helps the non-linear kernel define complicated boundaries. In turn, LR shows the greatest stability in terms of results.

Prediction of Side Effects for Uncharacterized Drugs

Predictions for the side effect profile of 15 out of the 20 chosen side effects were made on 92 unknown drugs – the drugs that appear in the database but have no side effect information in SIDER database but have information related to their protein

interactions. Training dataset comprises 658 known drugs along with the structure of 1368 target proteins. These predictions were compared with the independently validated side effect information available at Drugs.com [35].

Model	% Matched Values
LR	76.45%
NB	74.20%
SVM	71.96%
RF	69.64%
DT	68.85%

Table 5. Prediction results for 92 uncharacterized drugs

The LR model produced the highest match percentage of 76.45%, followed by NB with a match percentage of 74.20%, and SVM with a match percentage of 71.96%. This ordering of machine learning models is identical to the classification results, further reinforcing the generalization performance of the LR model. It can be seen from the above analysis that protein interaction data alone are sufficient to predict side effects of new drugs.

V.DISCUSSION

Across all three experimental comparisons, there are three major conclusions to take away from these studies. Targeted protein interaction profiles can serve as sufficient features for drug side effect prediction: by extending and quantitative corroborating analysis conducted earlier that established the biological basis of such a relationship through the use of the same Mizutani dataset also used in this new series of studies. Logistic Regression (LR) consistently outperformed even more complicated models (e.g., RF, SVM) when reviewed together, lending weight to the conclusion that use of a linear regression method with a regularised decision boundary through a high dimensional protein interaction space is adequate and prevents overfitting associating with other more flexible models based upon this same dataset size. SVM (RBF) produced the highest peak accuracy among common drug types associated with side effects but had less confidence among less common drug types making LR a better overall choice for real usage in terms of reliability.

The gap in overall accuracy (91.14%) and precision/F-score (45.3%/49.6%) for the baseline model illustrates the general problem of class imbalance in pharmacological datasets in that the majority of drugs produce the majority of side effects. Furthermore, one must also always include the reporting of all accuracy, precision, and F-score together as opposed to only reporting accuracy in order to avoid the pitfall of reporting accuracy alone as seen in the literature; for example; [15].

VI.CONCLUSION

This study analyzes the possible applications of multiple methods to identify the side effects that could be associated with a variety of drugs through the use of applicable protein interaction, drug, and targeted protein interaction data, as well as the in vitro/in vivo characteristics of the drug. The base predictive accuracy of the Naive Bayes classifier achieved was 91.14% across all 1339 unique side effects present in the Mizutani dataset; the 15 most prevalent and commonly occurring side effects that were identified throughout the course of this study confirmed that all of the algorithms developed for prediction would meet or exceed acceptable accuracy rates. A comparison of the predictions yielded by the different model types indicated that Logistic Regression (LR) maintained the highest average accuracy for predicting the side effects at 70.29%, followed by Support Vector Machine (SVM) with a RBF kernel 66.15%; then Random Forest (RF) 63.92%, Naive Bayes classifier using the Bernoulli distribution (NB(B)) 63.77%, Naive Bayes classifier using the Multinomial distribution (NB(M)) 62.57%, SVM (Linear) 62.44%, Decision Tree (DT) 60.29%, and finally, Naive Bayes classifier using the Gaussian distribution (NB(G)) 53.15%.

When considering the differing performance outputs for each algorithm developed and evaluated, SVM (RBF) was the strongest performing algorithm for predicting the high prevalence effects of nausea 85.03% and vomiting 78.96%, while LR yielded the highest average accuracy overall for the collection of side effects across all models. Based on LR's predictive accuracy results for predicting side effects for 92 unknown drugs, the match rate between each was 76.45% when compared against the Drugs.com independent verification of the side effect profiles. The results presented in this publication build further upon prior work presented with respect to identifying targeted proteins related to the side effects of drugs in order to provide a comprehensive framework for machine learning to build on the previously described biological structure of drug targets and side effects.

Another major finding from this study is that protein-protein interaction profiles specifically designed for this purpose can be considered adequate and meaningful features in drug side effect prediction. The approach outlined in this paper could prove useful at various stages in the process of developing new medications, ranging from initial stages of safety testing all the way through to postmarketing surveillance of drugs.

REFERENCES

1. World Health Organization. Requirements for Adverse Reaction Reporting. Geneva: WHO; 1975.
2. Lasser KE, Allen PD, Woolhandler SJ, et al. Timing of new black box warnings and withdrawals for prescription medications. JAMA. 2002;287(17):2215–2220.
3. National Institute on Drug Abuse (NIDA). Overdose Death Rates. 2025. https://nida.nih.gov/research-topics/trends-statistics/overdose-

death-rates.
4. Giacomini KM, Krauss RM, Roden DM, et al. When good drugs go bad. Nature. 2007;446(7139):975.
5. Bender A, Scheiber J, Glick M, et al. Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure. ChemMedChem. 2007;2(6):861–873.
6. Fliri AF, Loging WT, Thadeio PF, Volkmann RA. Analysis of drug-induced effect patterns to link structure and side effects of medicines. Nat Chem Biol. 2005;1(7):389.
7. Fukuzaki M, Seki M, Kashima H, Sese J. Side effect prediction using cooperative pathways. IEEE BIBM. 2009.
8. Merle L, Laroche M-L, Dantoine T, Charmes J-P. Predicting and preventing adverse drug reactions in the very old. Drugs Aging. 2005;22(5):375–392.
9. Yan X-Y, Zhang S-W, He C-R. Prediction of drug-target interaction by integrating diverse heterogeneous information source with multiple kernel learning and clustering methods. Comput Biol Chem. 2019;78:460–467.
10. Xian Z, Zhao X, Chen L, Lu J. A similarity-based method for prediction of drug side effects with heterogeneous information. Math Biosci. 2018;306:136–144.
11. Niu Y, Zhang W. Quantitative prediction of drug side effects based on drug-related features. Interdiscip Sci. 2017;9(3):434–444.
12. Dimitri GM, Lio P. DrugClust: a machine learning approach for drugs side effects prediction. Comput Biol Chem. 2017;68:204–210.
13. Zhang W, Yue X, Liu F, et al. A unified frame of predicting side effects of drugs by using linear neighborhood similarity. BMC Syst Biol. 2017;11(6):101.
14. Onay A, Nar F. Classifying approved and withdrawn drugs based on their targets and interactions. Turk J Chem. 2020;44:55–69.
15. Lee SI, Yasunaga H, Sugiyama M. Drug side-effect classification using data analytics. Bioinformatics. 2019.
16. Zheng K, You Z-H, Wang L, et al. MLMDA: a machine learning approach to predict and validate microRNA-disease associations by integrating of heterogeneous information sources. J Transl Med. 2020;18:1–14.
17. Zhang W, Liu X, Chen Y, et al. Feature-derived graph regularized matrix factorization for predicting drug side effects. Neurocomputing. 2018;287:154–162.
18. Liu M, Wu Y, Chen Y, et al. Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. J Am Med Inform Assoc. 2012;19(e1):e28–e35.
19. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics. 2018;34(13):i457–i466.
20. Wang CC, Shi H, Xu D, et al. Deep feature fusion for adverse drug reaction prediction. Sci Rep. 2021;11:1–12.
21. Hu J, Liu J, Zheng Y, et al. Multi-label deep learning for large-scale adverse drug reaction prediction. IEEE J Biomed Health Inform. 2021;25(3):712–723.
22. Li J, Zheng S, Chen B, et al. A survey of current trends in drug repositioning. Drug Discov Today. 2016;21(4):578–580.
23. Zhang S, Liang M, Tang Z, et al. Graph attention network for drug adverse reaction prediction. Bioinformatics. 2022.
24. Wu Y, Xu P, Feng Y, et al. Knowledge graph-based drug side effect prediction. J Cheminform. 2022;14:1–13.
25. Cheng F, Zhao Z. Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. J Am Med Inform Assoc. 2014;21(e2):e278–e286.
26. Yang X, Ye Y, Wang X, et al. Explainable deep learning for adverse drug reaction prediction. Front Pharmacol. 2022.
27. Zhao Q, Zhang X, Li S, et al. Transformer-based adverse drug reaction prediction using drug SMILES and protein sequences. J Biomed Inform. 2023.
28. Liu J, Hu Z, Wang H, et al. Multi-task learning for correlated side effect prediction. NPJ Digit Med. 2023.
29. Chen R, Wang J, Zhu M, et al. Contrastive self-supervised learning for drug adverse event prediction. Bioinformatics. 2023.
30. Mizutani S, Pauwels E, Stoven V, Goto S, Yamanishi Y. Relating drug-protein interaction network with drug side effects. Bioinformatics. 2012;28(18):i522–i528.
31. Wishart DS, Knox C, Guo AC, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res. 2008;36(suppl_1):D901–D906.
32. Gunther S, Kuhn M, Dunkel M, et al. SuperTarget and Matador: resources for exploring drug-target relationships. Nucleic Acids Res. 2007;36(suppl_1):D919–D922.
33. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P. A side effect resource to capture phenotypic effects of drugs. Mol Syst Biol. 2010;6(1):343.
34. Kim S, Chen J, Cheng T, et al. PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Res. 2021;49(D1):D1388–D1395.
35. Drugs.com. Drug Information Online. https://www.drugs.com/. Accessed 2024.
36. Singh Thakur A, Aggarwal P. Correlation between Targeted Protein and Drug Side Effects: A Step towards the Prediction of Drug Toxicity. SSRN Electronic Journal. 2019. doi:10.2139/ssrn.3446550. Available at: https://www.researchgate.net/publication/335894642