IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 12 , Iss 1, 2023

Predicting Drug Side Effects with Advanced Deep Learning: The DeepSide Approach

DR.K.NAGESWARARAO, PROFESSOR
DEPARTMENT OF CSE
Sai Tirumala NVR Engineering College, Narasaraopet
Mail Id: nageswararaokapu@yahoo.com

ABSTRACT

Unexpected adverse effects that cause clinical trial drug failures endanger participant health and cause large financial losses. Potentially, side effect prediction algorithms could guide the development of new drugs. The LINCS L1000 dataset provides a wealth of cell line gene expression data disrupted by different drugs and establishes a foundation of knowledge for context-specific factors. The state-of-the-art approach uses just the high-quality trials in LINCS L1000, discarding a large proportion of the tests to employ context-specific information. Our study's goal is to make the most of this data in order to enhance prediction performance. Our experiments involve five deep learning architectures. Among multi-layer perceptron-based architectures, we demonstrate that a multi-modal design produces the best prediction performance when employing drug chemical structure (CS) and the full set of drug altered gene expression profiles (GEX) as modalities. We believe that the CS is generally more informative than the GEX. Using only the SMILES string representation of the drugs, a convolutional neural network-based model achieves the best results, outperforming the state-of-the-art by 13:0% in macro-AUC and 3:1% in micro-AUC. We also show that the model can predict drug-side effect combinations that were not included in the ground truth side effect dataset but have been reported in the literature.

1. INTRODUCTION

Computational methods hold great promise for mitigating the health and financial risks of drug development by predicting possible side effects before entering into the 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 drug-protein interactions, protein-protein interactions (PPI), activity in metabolic networks, 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 the 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 (i.e., drug-protein interactions, etc.) and are not cell or condition (i.e., dosage) specific. To address this issue ,Wang et al. (2016) utilize the data from the LINCS L1000 project [32]. This project profiles gene expression changes in numerous human cell lines after treating them 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 the gene expression profiles provide

context-dependent information for the side-effect prediction task. While the LINCS dataset contains a total of 473,647 experiments for 20,338 compounds, their method utilizes only the highest quality experiment for each drug to minimize noise. This means that most of the expression data are left unused, suggesting a potential room for improvement in the prediction performance. Moreover, their framework performs feature engineering by transforming gene expression features to enrichment vectors of biological terms. In this work, we investigate whether the incorporation of gene expression data along with the drug structure data can be leveraged better in a deep learning framework without the need for feature engineering.

In this study, we propose a deep learning framework, Deep Side, for ADR prediction. Deep Side uses only (i) in vitro gene expression profiling experiments (GEX) and their experimental meta data (i.e., cell line and dosage - META), and (ii) the chemical structure of the compounds (CS). Our models train on the full LINCS L1000 dataset and use the SIDER dataset as the ground truth for drug - ADR pair labels [13]. We experiment with five architectures: (i) a multi-layer perceptron (MLP), (ii) MLP with residual connections (Res MLP), (iii) multi-modal neural net- works (MMNN. Concat and MMNN. Sum), (iv) multi-task neural network (MTNN), and finally, (v) SMILES convolutional neural network (SMILES Conv).

We present an extensive evaluation of the above-mentioned architectures and investigate the contribution of different features. Our experiments show that CS is a robust predictor of side effects. The base MLP model, which uses CS features as input, produces _11% macro-AUC and _2% micro- AUC improvement over the state-of-the-art results provided in [32], which uses both GEX (high quality) and CS features. The multi-modal neural network model, which uses CS, GEX and META features and uses summation in the fusion layer (MMNN. Sum) achieves 0:79 macro-AUC and 0:877 micro-AUC which is the best result among MLP based approaches. We also find out that when the chemical structure features are fully utilized in a complex model like ours, it overpowers the information that is obtained from the GEX dataset. The convolutional neural network that only uses the SMILES string representation of the drug structures achieves the best result among all the proposed architectures with provides 13:0% macro-AUC and 3:1% micro-AUC improvement over the state-of-the-art algorithm.

2. EXISTING SYSTEM

A drug-drug interaction (DDI) is defined as an association between two drugs where the pharmacological effects of a drug are influenced by another drug. Positive DDIs can usually improve the therapeutic effects of patients, but negative DDIs cause the major cause of adverse drug reactions and even result in the drug withdrawal from the market and the patient death. Therefore, identifying DDIs has become a key component of the drug development and disease treatment.

In this study, an existing system, develops a method to predict DDIs based on the integrated similarity and semi-supervised learning (DDI-IS-SL). DDI-IS-SL integrates the drug chemical, biological and phenotype data to calculate the feature similarity of drugs with the cosine similarity method. The Gaussian Interaction Profile kernel similarity of drugs is also calculated based on known DDIs. A semi-supervised learning method (the Regularized Least Squares classifier) is used to calculate the interaction

possibility scores of drug-drug pairs. In terms of the 5-fold cross validation, 10-fold cross validation and de novo drug validation, DDI-IS-SL can achieve the better prediction performance than other comparative methods. In addition, the average computation time of DDI-IS-SL is shorter than that of other comparative methods. Finally, case studies further demonstrate the performance of DDI-IS-SL in practical applications.

Disadvantages

• The complexity of data: Most of the existing machine learning models must be able to accurately interpret large and complex datasets to detect an accurate Drug Side Effect.

• Data availability: Most machine learning models require large amounts of data to create accurate predictions. If data is unavailable in sufficient quantities, then model accuracy may suffer.

• Incorrect labeling: The existing machine learning models are only as accurate as the data trained using the input dataset. If the data has been incorrectly labeled, the model cannot make accurate predictions.

3. PROPOSED SYSTEM

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. Such residual connections have been shown to reduce the vanishing gradient problem to a large extend [7].

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.

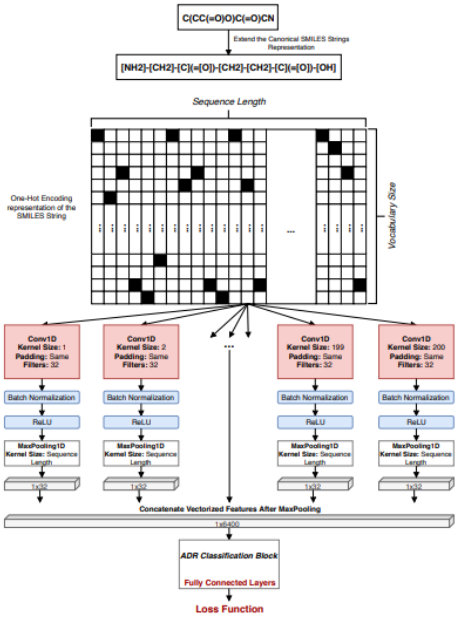
Multi-task neural network (MTNN) our multitask learning (MTL) based architecture aims to take the side effect groups obtained from the taxonomy of ADReCS into account. For this purpose, the approach defines shared and task-specific MLP sub-network blocks. The shared block takes the concatenation of GEX and CS features as input and outputs a joint embedding. Each task-specific sub-network then converts the joint embedding into a vector of binary prediction scores for a set of inter-related side-effect classes.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 12 , Iss 1, 2023

Advantages

- The proposed system implemented many ml classifies for testing and training on datasets.
- The proposed system developed Convolutional neural networks (CNN) which are known to provide a powerful way of automatically learning complex features in vision tasks to find an accurate accuracy on the datasets.

4. SYSTEM ARCHITECTURE



5. MODULES

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Train & Test Drug Data Sets, View Trained and Tested Drug Datasets Accuracy in Bar Chart, View Trained and Tested Drug Datasets Accuracy Results, View Drug Side Effect Prediction Type, Find Drug Side Effect Prediction Type Ratio, Download Predicted Data Sets, View Drug Side Effect Prediction Type Ratio Results, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN,PREDICT DRUG SIDE EFFECT TYPE, VIEW YOUR PROFILE.

6. ALGORITHIM

Gradient boosting

Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees.[1][2] When a decision tree is the weak learner, the resulting algorithm is called gradient-boosted trees; it usually outperforms random forest.A gradient-boosted trees model is built in a stage-wise fashion as in other boosting methods, but it generalizes the other methods by allowing optimization of an arbitrary differentiable loss function.

Logistic regression Classifiers

Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name *logistic regression* is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name *multinomial logistic regression* is usually reserved for the case when the dependent variable has three

or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practical use of the procedure is similar.

Logistic regression competes with discriminant analysis as a method for analyzing categorical-response variables. Many statisticians feel that logistic regression is more versatile and better suited for modeling most situations than is discriminant analysis. This is because logistic regression does not assume that the independent variables are normally distributed, as discriminant analysis does.

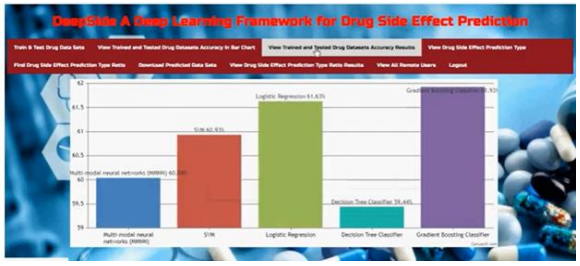
This program computes binary logistic regression and multinomial logistic regression on both numeric and categorical independent variables. It reports on the regression equation as well as the goodness of fit, odds ratios, confidence limits, likelihood, and deviance. It performs a comprehensive residual analysis including diagnostic residual reports and plots. It can perform an independent variable subset selection search, looking for the best regression model with the fewest independent variables. It provides confidence intervals on predicted values and provides ROC curves to help determine the best cutoff point for classification. It allows you to validate your results by automatically classifying rows that are not used during the analysis.

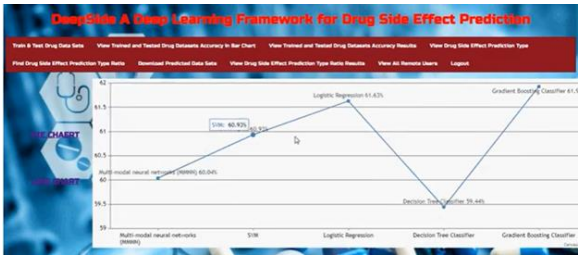
SVM

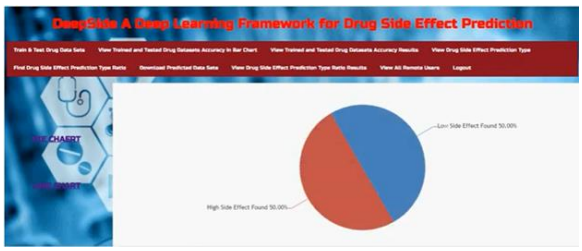
In classification tasks a discriminant machine learning technique aims at finding, based on an *independent and identically distributed* (*iid*) training dataset, a discriminant function that can correctly predict labels for newly acquired instances. Unlike generative machine learning approaches, which require computations of conditional probability distributions, a discriminant classification function takes a data point x and assigns it to one of the different classes that are a part of the classification task. Less powerful than generative approaches, which are mostly used when prediction involves outlier detection, discriminant approaches require fewer computational resources and less training data, especially for a multidimensional feature space and when only posterior probabilities are needed. From a geometric perspective, learning a classifier is equivalent to finding the equation for a multidimensional surface that best separates the different classes in the feature space.

SVM is a discriminant technique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyperplane parameter—in contrast to *genetic algorithms* (*GAs*) or *perceptrons*, both of which are widely used for classification in machine learning. For perceptrons, solutions are highly dependent on the initialization and termination criteria. For a specific kernel that transforms the data from the input space to the feature space, training returns uniquely defined SVM model parameters for a given training set, whereas the perceptron and GA classifier models are different each time training is initialized. The aim of GAs and perceptrons is only to minimize error during training, which will translate into several hyperplanes' meeting this requirement.

7. SCREEN SHOTS







8. CONCLUSION

Pharmaceutical medication development is a lengthy and challenging procedure. During the drug development process, unexpected adverse drug reactions (ADRs) have the potential to stop or restart the entire process. Therefore, anticipating the negative effects of the medicine during the design phase is essential. We use our Deep Side framework, which integrates chemical structure with context-related (gene expression) information, to forecast ADRs, taking into consideration variables such as dosage, time interval, and cell line. By using GEX and CS as integrated features, the proposed MMNN model achieves higher accuracy than models that only use chemical structure (CS) fingerprints. The claimed accuracy is noteworthy given that all we are trying to do is forecast the adverse effects that are independent of the condition. Finally, the SMILES Conv model outperforms all other approaches when convolution is used to the SMILES representation of drug chemical structure.

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.compbiolchem.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