INTRODUCTION

The most common definition, provided by the WHO, for an adverse drug reaction (ADR) is “A response to a drug which is noxious and unintended and which occurs at doses normally used in man for prophylaxis, diagnosis, or therapy of disease, or the modification of physiological function or noxious and unintended responses to drugs, when they are administered in their normal recommended dosages.”

This study focuses on the prediction of ADRs caused by the drug–drug interaction (DDI) of two-drug combinations. Apart from contributing to various productive drug design strategies such as drug repurposing (Zhou et al., 2015), coadministered drugs can exhibit synergistic DDIs (Liu et al., 2017), which comprises a new ADR that may be associated with either of the drugs or the aggravation of an existing ADR. In this study, we have proposed an artificial neural network (ANN) that predicts this specific subclass of ADRs using transcriptomic data, compound chemical fingerprint, and GO ontologies.

Abstract

Adverse drug reactions (ADRs) are pharmacological events triggered by drug interactions with various sources of origin including drug–drug interactions. While there are many computational studies that explore models to predict ADRs originating from single drugs, only a few of them explore models that predict ADRs from drug combinations. Further, as far as we know, none of them have developed models using transcriptomic data, specifically the LINCS L1000 drug-induced gene expression data to predict ADRs for drug combinations. In this study, we use the TWOSIDES database as a source of ADRs originating from two-drug combinations. 34,549 common drug pairs between these two databases were used to train an artificial neural network (ANN), to predict 243 ADRs that were induced by at least 10% of the drug pairs. Our model predicts the occurrence of these ADRs with an average accuracy of 82% across a multifold cross-validation.

KEYWORDS

ADR, adverse drug reaction, artificial intelligence, artificial neural network, drug–drug interaction, LINCS L1000, transcriptomic data, TWOSIDES dataset
1.1 Prevalence and salient statistics

Older studies implicate ADRs to be one of the top 10 causes of fatality in the United States (Lazarou et al., 1998), while more recent studies report that they account for anywhere for 2.7%–15% of hospitalizations (Liu et al., 2017; Miguel et al., 2012). Drug–drug interactions account for up to 30% of ADRs (Iyer et al., 2014) and are known to affect clinical outcomes in 80% of those observed in cancer patients (Beijnen & Schellens, 2004). While these statistics are for the United States, they are known to be important in all healthcare systems. In France, DDIs are estimated to be the cause of hospitalization in the elderly in as high as 2%–5% of the cases, and up to 1% in the general population (Létinier et al., 2019).

As a further example, observational studies conducted in a single tertiary care hospital over a period of 6 months reported 598 drug-related problems, of which 55.5% were due to drug interactions in 332 patients, of which 39 were putative ADRs (George et al., 2015). Another study focusing on geriatric patients, conducted over a period of one year, observed the incidence of polypharmacy in 22.9% of the patients and 29.5% of the patients were exposed to potentially inappropriate medication (since they were on multiple drugs) (Rajeev et al., 2018). Similarly, studies conducted in several other parts of India have estimated the incidence of suspected ADRs to be nearly 2%–3%, among hospitalized patients. Globally, the prevalence of ADRs to the proportion of in-patient hospitalizations was reported to be 2.3%, 4.8% and 7.3% for England, Germany, and the United States, respectively (Mulchandani & Kakkar, 2019). The administration of poly-pharma compounds increases the risk of adverse events due to drug–drug interaction (Iyer et al., 2014). If each of the drugs contributes to the cumulative probability of the observed ADR, the overall risk due to multiple drugs increases. To compound this problem, it is believed that 90% of ADRs go unreported (Raj et al., 2019).

1.2 Artificial intelligence in predicting ADRs caused by DDIs

Computational approaches to predict adverse drug reaction are based on the guilt-by-association rule. The predictions are based on the similarity of molecular structures, molecular circuits, and common targets and pathways. In general, these models process multidimensional data that are correlated with multidrug adverse events, and are able to collectively use integrated, heterogeneous databases. These analyses can potentially result in robust and comprehensive models that are precursors to experimental ADR studies that reduce cost and time to perform these studies experimentally, flag interactions that can be fatal, and significantly reduce losses from post-marketing surveillance (Chen et al., 2016). Artificial intelligence (AI) algorithms have been used to both detect and classify ADRs for both single- and two-drug conditions (see Lee and Chen (2020) for an exhaustive review of previous studies for all combinations of the above). Most of these detect and quantify the probability of interaction/interaction strength of DDIs, as opposed to a very small number of studies that assign a specific causal relation between the ADR and the DDI in question. We found three studies that use artificial intelligence/machine learning-related algorithms to predict associations of ADRs with DDI pairs. DeepDDI (Ryu et al., 2018) uses a deep neural network (DNN) to predict 86 specific ADRs using structural representation (SMILES) of the drug pairs using DrugBank’s DDI (Wishart et al., 2018). Decagon (Zitnik et al., 2018) developed a graph convolution neural network that modeled the prediction of polypharmacy as a multirelational link prediction problem between the drugs, the target, and the side-effects. They use STITCH (Szklarczyk et al., 2015) as a source for drug–protein interactions, and SIDER (Kuhn et al., 2016) to retrieve side-effects of individual drugs, as well as two-drug combinations. Finally, Zheng et al. (2018) built a machine learning algorithm based on what they term as highly credible negative samples. They use drug representation that combines fingerprints, protein targets, pathway enrichment, and substituents with ADR labels from the TWOSIDES dataset (Tatonetti et al., 2012). The analysis was designed such that each ADR was predicted individually using binary classification techniques that, historically, were mostly used for DDI detection (Lee & Chen, 2020).

1.3 Motivation and scope

In this study, we present an ANN to predict adverse drug reactions observed when two drugs are coadministered. The independent variable is an encoding of the drug pair, and the labels for training the ANN are obtained from the TWOSIDES dataset (Tatonetti et al., 2012). The encoded drug dataset used was previously calculated and utilized to predict single-drug ADR (Wang et al., 2016). We have used the same encoding (see Section 2.1), and juxtaposed the signatures of two drugs that constitute the drug pair whose DDI type we are predicting. In their scheme, which we have also adopted in this study, each drug is represented by its chemical fingerprint, differential gene expression (DEG) signature and an extension of the response data that represent the enrichment (per drug) for Gene Ontology gene sets. Specific to drug interactions, it should be noted that fingerprints embed structural information in the drug pair vector and provide an encoding that enables the algorithms to perceive functional group similarities between the two drugs. Nonetheless, using structural characterization of compounds alone can fail in cases where the structural similarities exist, but the
side-effects differ (Miotto et al., 2018). Wang et al. (2016) have used the LINCS L1000 (Subramanian et al., 2017) perturbagen (including compounds) responses encapsulated by transcriptomic data to predict single-drug-induced ADRs.

TWOSIDES dataset (Tatonetti et al., 2012) has collated adverse events caused by drug pairs that can be clearly distinguished from adverse events caused by the individual constituents. The differentiation is either in terms of a completely new type or a significant enhancement of an existing adverse event when compared to what is caused by a single drug. In effect, in this study and others (Liu et al., 2017; Zheng et al., 2018) the TWOSIDES dataset is used as a source of verified ADRs for combination of two drugs.

While previous studies that predict DDI types (ADRs) from DDI information have proposed that using drug expression data is one of the important and natural next steps (Ryu et al., 2018), there is no study till date that uses it to predict DDI-ADR causal relationships. The use of transcriptomic data supplements the largely dominant target-based method in predicting phenotypic effects. We combine the reported advantages of using multilabel classification to predict multidrug reactions, in combination with drug-induced differential gene expression data (as opposed to only drug-centered features), for prediction of ADRs originating from multidrug prescriptions.

1.4 Artificial neural network for multiclass predictions

In this study, we use an artificial neural network (Bishop C., 1995) to predict ADRs of two-drug combinations. The artificial neural network (ANN) is an example of supervised learning. In this framework, a classifier attempts to model the relationship between an input and its corresponding output by iteratively minimizing the "error," which is a function of the difference between the predicted output and desired output, over the training set. The layers sandwiched between the input and output layers are called hidden layers. The ANN offers great flexibility through a number of tunable parameters, for example, the number of hidden layers, number of nodes in each layer, error function, and optimization method (Reed & Marks, 1999). This facilitates modeling complex relationships between the input and various output classes. An ANN with 10 or more hidden layers is called a deep neural network (DNN) (Goodfellow et al., 2016), with a far greater number of tunable parameters and multiple architectures, making it a powerful classifier. Recent research has demonstrated the efficacy of DNN in predicting single-drug ADRs (Uner et al., 2019), ADRs for drug–drug interactions (Ryu et al., 2018) and possible ADRs of a new drug (Wang et al., 2020). In this study, we use a lighter-weight ANN with multidimensional features that span chemical structure representation, GO ontology terms, and drug-induced gene expression (Wang et al., 2016) to predict ADRs induced by two-drug combinations.

2 METHODS AND MATERIALS

The prediction model for ADRs of various drug pair combinations was structured as a multilabel classification task. The data sources, and the components in the respective resources used in this study are presented in Table S1. The models were built in Intel® Xeon(R) Silver 4114 CPU @ 2.20 GHz (40 cores), and 251.5 GiB RAM machine running on Ubuntu 16.04 LTS. The data processing and ANN model creation were done in a Python environment using various libraries, namely ScikitLearn (v0.22.2.post1) (Pedregosa et al., 2011), TensorFlow (v2.1.0) (Abadi et al., 2016), and Keras (v2.3.1) (Chollet, 2015). The dataset for drug representation is made public (Shankar, 2020a). The dataset for choosing best hyperparameters is made public (Bhandari, 2020a).

2.1 Drug pair representation and the associated ADRs

LINCS L1000 data (Subramanian et al., 2017) have differential expression data for a total of 20,338 perturbagens. Whang et al. (Wang et al., 2016) have used this to encode/represent drugs using three components (Table S1). In our study, this representation was directly downloaded from their paper (Prediction of Drug Side Effects, 2016; Wang et al., 2016). Further, the representation for a drug pair was formed by concatenating individual drug encodings. Thus, each input vector has 11,164 (Figure S1b) features.

Drug–drug–ADR associations to train our models are extracted from TWOSIDES database (Tatonetti et al., 2012). ADRs in the TWOSIDES data are extracted from Adverse Event Reporting System (AERS) from the FDA website. The categorization of these ADRs uses standardized MedDRA terminology. It comprises 868,221 associations between 59,220 pairs of drugs that collectively exhibit an aggregate of 1,301 adverse events. The drug names from LINCS L1000 metadata (called pert_iname) were used to map the records against the drug names in TWOSIDES. Subsequently, the associated perturbagen ID (pert_id) was used to merge the records from the Wang et al. dataset. While 206,792,742 drug pairs had transcriptomic data available from L1000 dataset, only 34,549 pairs had both expression and DDI-ADR information available from TWOSIDES data (see Figure 1).

Some of the ADRs are under-represented in the dataset, and will tend to skew the model. We decided to have a cutoff based on the minimum number of samples that need to manifest a particular ADR for it to be considered in the study. We
made empirical tests across five different percentages ranging from 10% through 90% at 20% intervals. To elucidate, for the 10% case, at least 10% of the 32,549 input samples is expected to present an ADR for it to be selected in the study. One-hot encoded vectors were used to represent the presence/absence of ADRs collectively associated for a particular drug pair.

2.2 | ANNs for the prediction of ADRs for drug pair combinations

Several machine learning models were tested, and a summary of methods and results is presented in the Appendix S1 (Section S3). ANNs were chosen based on performance evaluation. In this model, prediction of ADRs for two-drug combinations has been posited as a multilabel classification problem, since each drug pair can be associated with multiple ADRs. Recent literature has demonstrated the efficacy of artificial neural networks toward predicting ADRs associated with single-drug and drug–drug interactions, even with high dimensional and voluminous data encodings (Lee & Chen, 2020). The comparatively smaller dataset we use in this study prompted us to use lighter-weight ANNs that avoids the challenges associated with high computational complexity. An ANN is a network of nodes or artificial neurons distributed across different layers. A representation of the architecture used in this study is shown in Figure 2. The first layer comprises the input layer, which is the drug pair encodings in this study. In a trained ANN, each input is multiplied by an optimal weight. These weighted inputs are accumulated at each node in the subsequent layer and passed through a non-linear function, called the activation

FIGURE 1 Dataset preparation

FIGURE 2 Architecture of the Artificial Neural Network used for the neural model
function. This activation function ensures the weighted linear combination of inputs at any node lies within a specified range such as [0,1] or [-1,1]. The output of one layer serves as the input to the next, culminating in the final output layer. Through weights associated with the interconnections of nodes between layers, the network attempts to model the inherently complex boundaries between output classes in the input feature space. The weights in the network are adjusted iteratively in the training phase through a learning mechanism called backpropagation (Bishop C., 1995).

2.3 | Selection of ANN hyperparameters

The tuning of various parameters is typically empirical (Reed & Marks, 1999). In our model, the ANN has an input layer comprising 11,164 nodes corresponding to the features that represent two-drug combinations, three hidden layers, and an output layer that comprises 243 nodes corresponding to the 243 possible ADRs shortlisted for this study (Figure 2). A brief description of the parameters of the ANN and the rationale behind the choice of specific parameter values is presented in Table 1, and complete details of all the experiments that were run to evaluate the optimal values of each of the hyperparameters are explicated in the Section S2 (Appendix S1).

2.4 | Interpreting the ANN output and evaluation of the model

After the training process of the ANN is completed, a validation set comprising data that the network has not seen in the training phase is used to evaluate the model. This process was done in two stages:

2.4.1 | Binary representation

The last layer of the ANN with the loss function maps the values of each of the 243 nodes in the output layer corresponding to the 243 ADRs to a value in (0,1). This represents the probability that the input that corresponds to a drug pair combination will present that ADR. The target, however, is a binary value that is either a 0 or a 1 that indicates the absence or presence, respectively, of each ADR. Thus, there is a need to map the probabilities at the output layer to binary predictions. This can be done using a “threshold,” that is, a value that sets any value less than itself to 0 and greater than or equal to itself to 1. To find a suitable threshold, we first obtained the model’s prediction for each of the training samples. Then, for each ADR, we generated 100 thresholds between the minimum and maximum probabilities predicted for that ADR. From these 100 thresholds, we selected the one that maximized the F1 score for that ADR. These ADR-adaptive thresholds are then applied to the predictions on the validation set to obtain binary predictions. These binary predictions were compared with the ground truth to evaluate the model’s efficacy.

2.4.2 | Performance evaluation and prediction of ADR for unknown combination

The model’s efficacy is determined using cross-validation on four separate subsets of the data as follows: (a) From the original set of 32,549 drug pairs, 8,000 drug pairs were randomly

| TABLE 1 | Artificial neural network (ANN) parameters and their values |
|----------------|--------------------------|----------------|
| Parameter                  | Value                          | Description                                      |
| Number of hidden layers and nodes | Three hidden layers; the number of nodes in each layer is as follows: Layer 1:5,000 Layer 2:4,000 Layer 3:1,000 | The number of layers and nodes contribute to learning the complex map between the input and output layers. Since the input distribution is arbitrary, these parameters are best determined empirically (Reed & Marks, 1999) |
| Activation function | Rectified linear unit or ReLu | ReLu (x) = max(0,x) This ensures the output is non-negative (i.e., in the range [0, inf)) and is easy to compute (Glorot et al., 2011) |
| Optimizer                  | Stochastic gradient descent (SGD) | A computationally efficient approach to arriving at a minimum in the loss function (Bottou, 2012); having experimented with multiple options, we found SGD works best for our data (Choi et al., 2019) |
| Loss function              | Binary cross-entropy (also known as log loss) | A measure of how far away the ANN prediction is from the desired target output; well suited for multiclass problems (Janocha & Czarnecki, 2016) |
| Patience                   | 15 | The number of epochs for which the training should continue despite a decrease in the chosen performance metric to overcome local minima, harnessing the concept of overtraining (Amari et al., 1997) |
sampled and split to form four distinct subsets of 2,000 drug pairs each (say, S1, S2, S3, and S4) to be used for validation; (b) for each validation set Si, the corresponding training set was created by combining the 24,549 drug pairs that remain in the main dataset (Ds), with three subsets Sj (with j ≠ i) using the set Si with unknown drug pair combinations as the validation set; and (c) for each training set, the performance was evaluated on the corresponding validation set. The final performance is reported as an average of the performances on the four validation sets. The macro-averaging values of accuracy, RoC, and F1-score are used as evaluation metrics as implemented in the Keras API (look in the Supporting Information of Ryu et al. (2018) to find the equations of the above).

3 | RESULTS

3.1 | Comparison between the five input datasets constructed based on number of samples manifesting a particular ADR

Empirical studies to decide on the sample cutoff percentages that manifest ADRs show that using a 10% threshold resulted in optimal model prediction (see Section 2.1). Two hundred forty-three ADRs were present with a 10% cutoff, 165 with 30%, 68 ADRs with 50%, 18 with 17%, and three with 90% (see Figure S2k that compares average scores across four cross validated sets for each of the five partitions of data, Figure S2l that confirms the 10% partition yields the most number of ADRs, when a 0.65 cutoff is used for the F1-score, and dataset and detailed result present in Bhandari (2020a)). These results are consistent with the expectation that the power of discrimination reduces between the data samples with increase in the percentage of drug pairs that exhibit a particular ADR for it to be selected in the training set.

For 20,338 drugs in the L1000 dataset, 206,827,791 unique pairs were created, of which 34,549 drug pairs were common to both L1000 and TWOSIDES dataset. As detailed in Section 2.4, subsets of 32,549 were used to build the models and the remaining 2,000 combinations were used to predict the unknown ADRs. This was repeated for four such training and validation subsets, ensuring that each subset of 2,000 combinations used to test the model was randomly sampled and not overlapping with the corresponding training set. Of the 243 ADRs, respiratory distress syndrome had the least drug–drug associations (10% association), and decreased arterial pressure had the most number of drug–drug pairs associated with it (41% associations). The number of ADRs associated with a drug pair ranged from 0 to 195 out of 243 ADRs, and had an average of 43.19.

3.2 | Performance of the ANN model

The architecture of the final version of the ANN is presented in Table 1, and a detailed analysis of how we arrived at this configuration through empirical testing is in Section S2 (Appendix S1). The ANN yielded an average accuracy of 82% across all 243 ADRs. The top ADR predictions had a minimum RoC of 0.68, after which the values dropped by larger quantum. These top ADRs are presented in Table 2. The boxplot for the measured macro-evaluation metrics for all ADRs is presented in Figure S4. The performance of each of the ADR in terms of the evaluation metrics is provided as Appendix S1 (file name: Evaluation Metrics for the ANN Model.csv). The maximum and minimum of each of the evaluation metrics, corresponding to the most and least predictable ADRs for each of the performance measures, viz. accuracy, F1-score, and RoC, is reported in Table S4. Both the dataset and results are present in Shankar (2020b) and Shankar (2020c). Finally, this model was extensively compared against many previous studies, and against a logistic regression implementation. The results and details are presented in Sections S.5 and S.6 in the Appendix S1. As can be observed in Figures S5a,b and S6a,b, our model provides a competitive predictive tool when transcriptomic drug responses are available and the user seeks to explore drug–drug interactions using the same.

3.3 | ADR predictions

A set of 243 ADRs was identified, where each of those was manifest in at least 10% of the total samples. An F1-score >0.5 ensures positive discrimination of the model. With a threshold of 0.68 on the RoC, we obtain nine most predictable ADRs out of the 243 ADRs. The macro-averaged evaluation metrics after validating with four different datasets for the top-performing ADR is shown in Table 2 (for the

| ADR                               | Accuracy | Macro-F1 score | Macro-RoC  |
|-----------------------------------|----------|----------------|------------|
| Drug addiction                    | 0.864750 | 0.703959       | 0.702692   |
| Bruxism                           | 0.825875 | 0.702245       | 0.709496   |
| Mod                               | 0.813875 | 0.692815       | 0.698257   |
| Bursitis                          | 0.885625 | 0.691321       | 0.699877   |
| Fibromyalgia                      | 0.859375 | 0.690880       | 0.693449   |
| Migraine                          | 0.835000 | 0.689689       | 0.697851   |
| Arteriosclerotic heart disease    | 0.781500 | 0.687058       | 0.689559   |
| Acid reflux                       | 0.776000 | 0.686985       | 0.690390   |
| Elevated cholesterol              | 0.773250 | 0.685067       | 0.687172   |
3.4 | Use case of naproxen and goserelin

An extensive analysis of ADRs observed for multiple drug pairs considered is non-feasible, and beyond the scope of this work. Instead, we provide the complete list of results as a table in the Appendix S1 (file name: Drug_pair_output_all_subsets.xlsx). Domain experts can use this to investigate the drug or drug pair of interest. As with any predictive machine learning/data mining results, it is unreasonable to expect every ADR identified by the ANN to be significant (Zheng et al., 2016). As an exemplar of the type of analysis that may be helpful to researchers, we highlight the case of ADRs predicted for the naproxen–goserelin case since it achieved the highest accuracy score of 92%. As is the case with the results provided as an Excel document in the Appendix S1, each pair of the output data we have reported is from the test set from the respective partition of the data. While the performance metrics indicate the confidence in prediction for each drug pair, these tables add value to a domain experts' analysis by indicating other plausible ADRs, not reported by training dataset (TWOSIDES), albeit subject to validation experiments. Naproxen is a non-steroidal anti-inflammatory drug. It is used to treat acute pain and pain related to rheumatic diseases. Goserelin is a synthetic hormone that stops the production of the testosterone in men and may stimulate the growth of cancer cells. The pharmacokinetics, as reported by DrugBank, indicates that naproxen may decrease the excretion rate of Goserelin, which could, in turn, result in a higher serum levels (Wishart et al., 2018).

For this drug combination, a total of 16 ADRs were predicted to occur. Ten of 16 ADRs were also reported by the Tattonetti et al. dataset. All six of the other ADRs were reported by other databases as adverse events for at least one of the drugs in the combination. The resources used to confirm these observations are Drugs.com (Goserelin Side Effects: Common, Severe, & Long Term - Drugs.com, 2019; Naproxen Uses, Dosage, Side Effects, & Warnings - Drugs.com, 2020) and SIDER (Contributors to Wikimedia projects, 2010; Kuhn et al., 2016). Pain is reported to be caused by both drugs independently, but neuropathy was not reported anywhere in the literature. Figure 3 captures the detailed annotation for this drug combination. Both the figure and this format of presenting results were used in earlier studies (e.g., Zheng et al. (2018)).
time-consuming exercise (Murali et al., 2020). The scope of this project only covers the inclusion of ADRs resulting from DDIs reported in the TWOSIDES data table. Further improvements through enrichment of observed ADRs for drug pairs can be achieved by integrating the above with DrugBank's DDI (Wishart et al., 2018), all drug-drug side-effects reported in SIDER (Kuhn et al., 2016), and PAERS (Poleksic & Xie, 2019).

**CONFLICT OF INTEREST**

There is no conflict of interest among the authors.

**DATA AVAILABILITY STATEMENT**

The data that support the findings of this study are openly available in figshare at:

- https://doi.org/10.6084/m9.figshare.12602159 (Bhandari, 2020a).
- https://doi.org/10.6084/m9.figshare.12602150 (PerformanceOfFiveCutOffs, 2020).
- https://doi.org/10.6084/m9.figshare.12589547 (Bhandari, 2020b).
- https://doi.org/10.6084/m9.figshare.12582818 (Shankar, 2020b).
- https://doi.org/10.6084/m9.figshare.12582032 (Shankar, 2020d).
- https://doi.org/10.6084/m9.figshare.12579704 (Shankar, 2020a).
- https://doi.org/10.6084/m9.figshare.12581228 (Shankar, 2020e).

**ORCID**

David T. Okou https://orcid.org/0000-0001-6473-5012
Prashanth Athri https://orcid.org/0000-0002-5731-0291

**REFERENCES**

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., … Xiaoqiang, Z. (2016). TensorFlow: A System for Large-Scale Machine Learning, 12th {USENIX} Symposium on Operating Systems Design and Implementation (OSDI) ’16, Savannah, GA: USENIX Association. https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi

Amari, S., Murata, N., Muller, K.-R., Finke, M., & Yang, H. H. (1997). Asymptotic statistical theory of overtraining and cross-validation. *IEEE Transactions on Neural Networks*, 8(5), 985–996. https://doi.org/10.1109/72.623200

Beijnen, J. H., & Schellens, J. H. M. (2004). Drug interactions in oncology. *The Lancet Oncology*, 5(8), 489–496. https://doi.org/10.1016/S1470-2045(04)01528-1

Bhandari, I. (2020a). 3.1 Comparison between the five input datasets constructed based on number of samples manifesting a particular ADR. figshare. https://doi.org/10.6084/m9.figshare.12589547.v1

Bhandari, I. (2020b). 3.1 Comparison between the five input datasets constructed based on number of samples manifesting a particular ADR. figshare. https://doi.org/10.6084/m9.figshare.12589547

Bottou, L. (2012). Stochastic gradient descent tricks. In *Lecture notes in computer science* (pp. 421–436). https://doi.org/10.1007/978-3-642-33289-8_25

Chen, Y.-G., Wang, Y.-Y., & Zhao, X.-M. (2016). A Survey on computational approaches to predicting adverse drug reactions. *Current Topics in Medicinal Chemistry*, 16(30), 3629–3635. https://doi.org/10.2174/15680266166661603182013

Choi, D., Shallue, C. J., Nado, Z., Lee, J., Maddison, C. J., & Dahl, G. E. (2019). On empirical comparisons of optimizers for deep learning. (arXiv:1910.05446). https://arxiv.org/pdf/1910.05446.pdf

Chollet, F. (2015). Keras. GitHub repository. https://github.com/fchollet/keras

Chollet, F. (2015). *Deep learning*. MIT Press.

Goserelin Side Effects: Common, Severe, Long Term - Drugs.com. (2019). Drugs.com. Retrieved from https://www.drugs.com/sfx/goserelin-side-effects.html. June 29, 2020

Iyer, S. V., Harpaz, R., LePendu, P., Bauer-Mehren, A., & Shah, N. H. (2014). Mining clinical text for signals of adverse drug-drug interactions. *Journal of the American Medical Informatics Association: JAMIA*, 21(2), 353–362. https://doi.org/10.1136/ami2013-001612

Janocha, K., Czarnecki W. M. (2016). On Loss Functions for Deep Neural Networks in Classification. *Schedae Informaticae*, 25, 49–59. http://dx.doi.org/10.4467/20838476si.16.004.6185

Kuhn, M., Letunic, I., Jensen, L. J., & Bork, P. (2016). The SIDER database of drugs and side effects. *Nucleic Acids Research, 44*(D1), D1075–D1079. https://doi.org/10.1093/nar/gkv1075

Lazarou, J., Pomeranz, B. H., & Corey, P. N. (1998). Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies. *JAMA: the Journal of the American Medical Association*, 279(15), 1200–1205.

Lee, C. Y., & Chen, Y.-P.-P. (2020). Prediction of drug adverse events using deep learning in pharmaceutical discovery. *Briefings in Bioinformatics*, 1–18. https://doi.org/10.1093/bib/bbaa040

Létinier, L., Cossin, S., Mansiaux, Y., Arnaud, M., Salvo, F., Bezín, J., Thiessard, F., & Pariente, A. (2019). Risk of drug-drug interactions in out-hospital drug dispensings in France: Results from the DRUG-drug interaction prevalence study. *Frontiers in Pharmacology*, 10, 265–273. https://doi.org/10.3389/fphar.2019.00265

Liu, R., AbdulHameed, M. D. M., Kumar, K., Yu, X., Wallqvist, A., & Reifman, J. (2017). Data-driven prediction of adverse drug reactions induced by drug-drug interactions. *BMC Pharmacology and Toxicology*, 18(1), 44. https://doi.org/10.1186/s40360-017-0153-6

Miguel, A., Azvedo, L. F., Araújo, M., & Pereira, A. C. (2012). Frequency of adverse drug reactions in hospitalized patients: A systematic review and meta-analysis. *Pharmacoepidemiology and Drug Safety*, 21(11), 1139–1154. https://doi.org/10.1002/pds.3309
