AUT Journal of Mathematics and Computing

AUT J. Math. Comput., 5(1) (2024) 45-59
https://doi.org/10.22060/AJMC.2023.21538.1091

Original Article

Adverse drug reaction prediction using voting ensemble training approach

Milad Besharatifard[a], Zahra Ghorbanali[a], Fatemeh Zare-Mirakabad[*a]

[a]*Computational Biology Research Center (CBRC), Department of Mathematics and Computer Science, Amirkabir University of Technology (Tehran Polytechnic), Iran*

ABSTRACT: Identifying and controlling adverse drug reactions (ADRs) is a challenging problem in the pharmacological field. For instance, the drug Rosiglitazone has been associated with adverse reactions that were only recognized after its release. Due to such experiences, pharmacists are now more interested in using computational methods to predict ADRs. The performance of computational methods is contingent upon the defined dataset. In some studies, the known drug-adverse reaction associations are regarded as positive while the unknown drug-adverse reaction associations are regarded as negative data. This consequently creates an unbalanced dataset, which can lead to inaccurate predictions from models and cause the classifiers to be flawed. We propose a framework named Adverse Drug Reaction using the Voting Ensemble Training Approach (ADRP-VETA) for ADR problem to overcome unbalanced dataset challenges. We construct the similarity vector of each drug with other drugs based on chemical structure as a drug feature. Also, the similarity vector of each ADR with other ADRs is computed based on the Unified Medical Language System (UMLS) as adverse reaction feature. With this approach, we can leverage the similarity of the features to more accurately capture the intricate relationships between drugs and adverse reactions. We compare ADRP-VETA to three state-of-the-art models and find that it outperforms them, achieving an AUC-ROC of 91% and an AUC-PR of 89.8%. Furthermore, we assess ADRP-VETA's ability to predict rare adverse reactions, and find that its AUC-ROC and AUC-PR are 83.3% and 92.2%, respectively. As a case study, we focus on the associations between liver-injury adverse reactions and three drugs.

Review History:
Received:29 June 2022
Revised:24 April 2023
Accepted:24 April 2023
Available Online:01 January 2024

Keywords:
Adverse drug reaction
Machine learning
Random forest
Rare adverse reactions
Unbalanced dataset

AMS Subject Classification (2010):
68T01; 68T07; 92-08

1. Introduction

An Adverse Drug Reaction (ADR) is a harmful or unintended reaction to a medication that occurs at doses normally used for treatment or prevention. ADRs can range in severity from mild symptoms to life-threatening events. Following the outbreak of the novel coronavirus (SARS-CoV-2), the World Health Organization (WHO) initially issued an emergency use authorization for the drug hydroxychloroquine. However, this authorization was later revoked due to the studies and tests that showed cardiotoxicity to be a rare adverse reaction of the drug [1]. This highlights the importance of detecting ADR prior to a drug being manufactured and made available to consumers [3, 31].

*Corresponding author.
E-mail addresses: milad1besharati@aut.ac.ir, z_ghorbanali@aut.ac.ir, f.zare@aut.ac.ir

This is an open access article under the CC BY-NC 2.0 license (https://creativecommons.org/licenses/by-nc/2.0/)

Sometimes, monitoring drugs after launch can reveal rare adverse reactions, as was the case with Rosiglitazone [29]. In May 2007, a clinical study found that taking Rosiglitazone significantly increased deaths caused by cardiovascular diseases. As a result, the drug was withdrawn in Europe and the United States imposed severe restrictions on its use in the same year [22]. Despite extensive laboratory studies conducted during various stages of drug discovery to identify potential adverse reactions, this approach has yet to be enough to adequately address the ADR problem. Therefore, there is an urgent need to diagnose ADRs accurately. To this end, researchers have become increasingly interested in approaching the ADR problem using computational methods. Nowadays, recommender systems [24] and machine learning methods [7, 21] have become popular computational models for addressing the ADR problem. A recommender system can predict whether users prefer an item based on their profiles [24]. This technique is also used in the ADR problem where drugs and their potential adverse reactions are treated as users and items, respectively. In other words, the recommender system can be used to predict potential ADRs of a given drug based on its profile. Galeano and Paccanaro [7] pioneered the use of collaborative filtering recommendation systems to predict potential adverse reaction for a new drug by leveraging known similar drugs. Poleksic et al. proposed the use of matrix factorization, a class of collaborative filtering recommendation systems, as a compressed sensing (CS) model to predict the unknown associations between drugs and adverse reactions [21]. The CS model is well-suited for dealing with sparse data, making it a particularly appropriate choice for the ADR problem, where the number of known drug-adverse reactions (positive data) is far lower than the unknowns (negative data). The proposed method presents a latent space for drugs and adverse reactions by computing a lower-dimensional vector that minimizes a predefined loss function in order to complete the drug-adverse reaction associations. In 2020, Guo et al. [8] utilized the Triple Matrix Factorization model to recover the drug-adverse reaction matrix and calculated the similarity between drugs and adverse reactions with different features. Recently, Shabani-Mashcool et al. [26] proposed a two-phases recommender system by applying fingerprints of drugs to address the ADR problem.

In addition to recommender systems, various machine learning methods can be used to effectively address ADR problem. Chen et al. [5] predicted the possible likelihood of each drug being associated with adverse effects. In this method, the similarity of two drugs based on drug-drug interaction and drug-protein interaction is calculated. Then, the algorithm uses the drugs with the same adverse reaction to calculate the drug-adverse reaction association score based on the drug relationships.

Zhao et al. [33] proposed an innovative approach to identify potential ADR by utilizing the similarity of drugs with different properties such as fingerprints, the two-dimensional structure of drugs, target proteins, ATC code and some features extracted from the STITCH database. Similarity vectors are considered as the input of the random forest classifier model. In addition, they considered adverse reactions that are associated with less than five drugs as rare. Rodrigues et al. [25] proposed a bayesian network approach to predict ADRs using 593 pharmaceutical care center reports. Dey et al. [6] introduced a model to convert two-dimensional or three-dimensional drug structures into numerical vectors using convolutional neural networks (CNNs) for each adverse reaction. Zhao et al. [32] utilized a multi-layer perceptron to estimate the frequency of ADRs after taking drugs using drug-drug and adverse reaction-adverse reaction similarities. Uner et al. [28] proposed a learning method to solve the ADR problem based on CNNs using the structural feature of drugs and gene expression characteristics.

Most of the aforementioned studies regarded unknown drug-adverse reaction associations as negative data; however, selecting appropriate negative samples is a major challenge in the area of ADR problem. Khan [10] limited the SIDER database [12] based on the ten adverse reactions with the maximum variance across the drugs. To differentiate between negative and positive data, the frequency of each adverse reaction is examined. Consequently, the drug-adverse reaction association is deemed positive or negative for ADR, if its frequency is above or below than 0.5, respectively. Zheng et al. [34] selected negative data based on the assumption that dissimilar drugs have fewer common adverse reactions. In 2020, Liang et al. [15] suggested a new approach for making negative data by random walking on drug-drug interaction networks. They removed rare adverse reactions associated with less than six drugs from the dataset. Zhang et al. [30] defined negative data based on drug-indication associations and applied a machine learning model to classify adverse reactions as either adverse or therapeutic. Table 1 shows a summary of the previous studies on the ADR problem.

Although previous models perform well in facing ADR problem, there are still challenges that need to be addressed:

- Most studies consider the unknown drug-adverse reaction as negative data. It causes two main drawbacks:
    - There is no evidence to support that the current status of the unknown drug-adverse reaction associations will remain unchanged and not develop into recognized associations in the future.
    - The number of the unknown drug-adverse reaction associations is too more than the known associations. So, the dataset is unbalanced. An unbalanced dataset causes the model to predict the minor class samples inaccurately.
- Although the challenge of experimental ADR identification is finding rare adverse drug reactions, some computational methods exclude rare ADRs to improve the performance of models.

References/Year	Drug feature	Adverse Reaction feature	Algorithm	#Drugs	#Adverse Reactions	Database for Adverse Reactions	Database for Drugs
2013[5]	Target protein Literature-Association	-	Machine Learning	835	100	-	SIDER STITCH
2017[10]	Fingerprint Indication Target protein	-	Machine Learning	667	10		SIDER Chembl
2018[21]	Fingerprint	UMLS semantic	Recommender System	1430	5868	SIDER MEDDRA	SIDER Pubchem
2018[33]	Fingerprint SMILES ATC code Target protein Literature association	-	Machine Learning	841	824	-	SIDER KEGG RDKit
2018[25]	ADR report	Causality categories	Machine Learning	-	2100	North Pharmacovigilance Center	-
2019[34]	Fingerprint Target protein ATC code Substituent	-	Machine Learning	917	500	-	SIDER DrugBank EMBL-EBI
2022[28]	Gene expression SMILES	-	Machine Learning	791	1052	-	SIDER Pubchem
2020[26]	Fingerprint Side effect profile	Drug profile	Recommender System	614	5596	SIDER Pubchem	SIDER
2020[15]	Fingerprint SMILES ATC code Target protein Literature association	-	Machine Learning	841	824	-	SIDER STITCH PubChem KEGG
2021[30]	Indication Target protein	CUI code	Machine Learning	3632	5589	SIDER DrugBank	SIDER

Table 1: Summary of reviewed papers.

We introduce ADRP-VETA[1] for ADR problem to face the above challenges. Here, we apply the similarity vector of each drug with other drugs based on fingerprint as a drug feature. Meanwhile, the similarity vector of each ADR with other ADRs is computed based on the Unified Medical Language System (UMLS) [17, 21] as adverse reaction feature. To overcome the negative data selection challenge, we define drug-indication associations as negative data similar to Zhang's approach [30] which is more accurate than considering all unknown pairs as negative. We consider the association between a drug and an adverse reaction as a negative sample when the drug can be indicated in the treatment of the adverse reaction. Moreover, to tackle the unbalanced dataset obstacle, we introduce a voting ensemble training approach to predict an output based on a frequent output of ensembles. In addition, some methods [15, 33] remove rare adverse drug reactions associated with less than five drugs to improve the performance of models. In this regard, we evaluate the performance of ADRP-VETA using three different strategies. The first strategy, classic, considers all rare and common adverse reactions to make the training and test sets for ADR problem. The second strategy, called common ADR, studies the performance of ADRP-VETA on common ADRs and excludes all rare adverse reactions for generating the test and training sets to predict ADRs. The third one, known as rare ADR, assesses the performance of ADRP-VETA for predicting rare ADRs. In this strategy, we use common ADRs to train the model and rare ADRs as the test set. Results show that the ADRP-VETA framework predicts rare adverse reactions more accurately than others. Finally, we focus on the associations between liver-injury adverse reactions and three drugs as a case study, which is not provided in SIDER database. Our model determines these associations, which are also confirmed by clinical studies.

2. Material and methods

This section introduces drug and adverse reaction datasets and then defines some notations to describe the adverse drug reaction (ADR) problem. Finally, the proposed framework named ADRP-VETA for ADR problem, is explained in more detail.

2.1. Datasets

To Address the ADR problem, we must extract drug and adverse reaction features as well as any known drug-adverse reaction associations from existing databases. This paper utilizes the SIDER database [12] for collecting adverse reactions (CUI codes), drugs (CID codes) and drug-adverse reaction associations. Moreover, if a drug treats an adverse reaction, we consider it as an indication association between the drug and adverse reaction. It should be noted that we remove overlapping drug-adverse reaction and drug-indication associations from the SIDER4.1 database. An example of such cases in the SIDER database is dexamethasone (CID: 3003), which can both cause an adverse reaction of eye disorder (CUI: 0015397) and treat the same disorder, thus creating two distinct drug-associations: a drug-adverse reaction association and a drug-indication association. Table 2 shows the details of the applied dataset, including the number of drugs, ADRs, drug-adverse reaction associations (positive data), and drug-indication associations (negative data). According to Table 2, the fifth column shows the number of known drug-adverse reaction associations (or positive data), 119040 samples, while the number of all possible drug-adverse reaction pairs is (1430*5562=) 7955090. Therefore, the large number of unknown associations (7836050) as negative samples results in the issue of unbalanced data. In addition, there is no evidence that these currently unknown associations will not become known in the future. To address this issue we utilize the Zhang's [30] idea to consider drug-indication associations as negative data to reduce the number of negative data to 10928 samples which is shown in the sixth column of Table 2.

Dataset	#Drugs	#Adverse reactions	#Indications	#Drug-Adverse reaction associations (P)	#Drug-Indication associations (N)
All ADRs in SIDER	1430	5562	1788	119040	10928
Rare ADRs in SIDER	1178	3318	605	5753	2122
Common ADRs in SIDER	1429	2244	1183	113287	8806

Table 2: Statistics of the applied dataset.

[1] Adverse Drug Reaction Prediction using Voting Ensemble Training Approach

2.2. Notations and definitions

This section describes the selected biological features of a drug and an adverse reaction. Moreover, we define some notations for the drug and adverse reaction representations.

2.2.1. Drug

A set of m drugs is denoted by $D = \{d_1, d_2, ..., d_m\}$, where $d_i \in D$ shows the i^{th} drug. It has been demonstrated that there is a correlation between certain chemical substructures and the occurrence of adverse drug reactions [35]. In this regard, we use chemical structure as a drug feature and display it by fingerprint representation. Thus, we extract fingerprint chemical structure from the PubChem database [11], where, each drug $d \in D$ is displayed as a binary vector $F^d = [f_1, ..., f_{881}]$ with length 881. Each $f_i \in \{0, 1\}$ represents the existence (1) or absence (0) of the i^{th} substructure descriptor associated with a specific chemical feature, respectively. To calculate the fingerprint similarity of drugs, we use the *Cosine similarity* (Cos) [9] criterion. For each drug $d \in D$, we define the similarity vector δ^d with length m, as follows

$$\delta^d = [Cos(d, d_1), Cos(d, d_2), ..., Cos(d, d_m)], d_i \in D, 1 \leq i \leq m, \tag{1}$$

where δ^d is defined based on the the similarity between fingerprint of drug d and drug $d_i \in D$.

2.2.2. Adverse reaction

The set of n adverse reactions is denoted by $A = \{a_1, a_2, ..., a_n\}$ where $a_j \in A$ shows the j^{th} adverse reaction. Moreover, we define a rare ADR as one that is associated with fewer than five drugs. So

$$A_{rare} = \{a_i | a_i \in A \text{ and there are less than five drugs associated to ADR } a_i\}. \tag{2}$$

We show the common ADRs as

$$A_{common} = A - A_{rare}. \tag{3}$$

Considering the fundamental assumption that drugs with similar properties will likely cause similar ADRs, it is imperative to take into account the semantic similarity and relatedness between adverse reactions when attempting to predict them. Here, we apply UMLS-similarity software to make a numerical vector for adverse reaction $a \in A$. The UMLS includes over 100 million medical terminologies with a unified and semantic network designed by the National Library of Medicine to support scientific research. The UMLS-similarity software measures the relatedness of ADRs and semantic similarity over MedDRA vocabulary [17]. The UMLS similarity of adverse reactions $a, a' \in A$ is computed using $UMLS(a, a')$. For each adverse reaction $a \in A$, we define the similarity vector α^a with length n, as follows

$$\alpha^a = [UMLS(a, a_1), UMLS(a, a_2), ..., UMLS(a, a_n)],\ a_i \in A,\ 1 \leq i \leq n. \tag{4}$$

2.3. Adverse drug reaction problem

We assume that $D = \{d_1, d_2, ..., d_m\}$ and $A = \{a_1, a_2, ..., a_n\}$ represent m drugs and n adverse reactions. In the ADR problem, biological features of drug $d \in D$ and adverse reaction $a \in A$ are given to the model. The primary goal of the ADR problem is to predict whether adverse reaction $a \in A$ will occur after taking drug $d \in D$. If the model predicts adverse reaction a associated with drug d, the output is one and otherwise zero.
For each drug d and adverse reaction a, we use the similarity vector δ^d and α^a as the numerical vector feature representations, respectively.

2.4. The voting ensemble training approach

This subsection introduces an approach called VETA to make the training and test sets. Most studies consider all the known drug-adverse reaction associations as positive data shown by

$$P = \{< a, d > | a \in A, d \in D, \text{and drug } d \text{ causes ADR } a\}, \tag{5}$$

and all the unknown associations as negative data. However, there is no evidence to suggest that the current unknown association between drugs and adverse reactions will not become known in the future. Furthermore, considering any unknown association as a negative instance leads to the problem of unbalanced dataset (see Table 2) and thus it is not a proper selection method. Therefore, instead of thinking of all the unknown associations as negative samples, we apply Zhang's method [30]. In this method, negative data is defined by

$$N = \{< a, d > | a \in A, d \in D, \text{and drug } d \text{ treats ADR } a\}, \tag{6}$$

Despite taking into account the negative samples based on drug-indication associations, the number of such associations is still lower than the number of drug-adverse reactions. To address this challenge, we randomly select 10% of negative samples and an equal number of positive samples to form the test set. Then, we use the under-sampling technique to randomly select a subset of the remaining positive data with the same size as the remaining negative samples to train the model. This approach is called one-to-one distribution. Although under-sampling solves the unbalanced data challenge, it does not consider all drug-adverse reaction associations. To tackle this obstacle, we introduce the voting ensemble training approach. In this manner, we cluster the positive data according to one-to-one distributions for the construction of positive training sets. If the number of these clusters is p, the model is trained p times on the negative data and every subset of clustered positive data. In this regards, the i^{th} training set is called ξ_i. For each test sample, the model predicts according to a vote on predictions of all trained models and makes the final decision. The details of generating training and test sets based on the voting ensemble training approach (VETA) are available as follows:

1. Sets P (positive samples) and N (negative samples) are given as inputs.
2. Set N_{test} is the negative test set with the size of $\frac{|N|}{10}$ and includes samples which are chosen randomly from N.
3. Set P_{test} is the positive test set and includes $|N_{test}|$ randomly chosen samples from P.
4. Set $N_{train} = N - N_{test}$ is the negative training set.
5. Set P_{train_i} is the i^{th} positive training set based on the one-to-one distribution approach and includes $|N_{train}|$ randomly chosen samples from set $P - P_{test} - (\bigcup_{j=1}^{i-1} P_{train_j})$. We assume that the number of these sets is equal to p, where $\bigcup_{i=1}^{p} P_{train_i} = P - P_{test}$ and $\bigcap_{i=1}^{p} P_{train_i} = \varnothing$.
6. Set $\epsilon = P_{test} \bigcup N_{test}$ is considered as the test set.
7. Set $\xi_i = P_{train_i} \bigcup N_{train}$ is considered as i^{th} training set.
8. Set $\tau = \{\xi_i | 1 \leq i \leq p\}$ is the voting ensemble training set.

Therefore VETA algorithm generates τ as a set of training sets named the voting ensemble training set.

2.5. ADRP-VETA framework

The task at hand is to predict adverse drug reactions prior to clinical testing phases. To issue this goal, the present paper proposes ADRP-VETA framework, which predicts the associations between drugs and adverse reactions. Fig.1 illustrates the ADRP-VETA framework visually. The primary steps of this framework are as follows:

- Extracting a numerical vector of drug $d \in D$, δ^d, based on the similarity of its fingerprint to the drugs in the training set (see (1)).
- Extracting a numerical vector of the adverse reaction $a \in$A, α^a, based on its UMLS-similarity to the ADRs in the training set (see (4)).
- Using the random forest [4] as a classifier to feed the concatenation of the vector δ^d and α^a (δ^d, α^a).
- Training the classifier on each training set ξ_i constructed by the VETA model to learn the associations between drugs and ADRs (see Fig 1-I).
- Predicting the association between each pair of drug and ADR as a test sample by i^{th} trained classifier named $\mathcal{R}_i$, $1 \leq i \leq p$ (see Fig 1-II).
- Selecting the prediction of trained classifier $\mathcal{R}_i$ on the test sample and voting on p trained classifiers (see Fig 1-III).

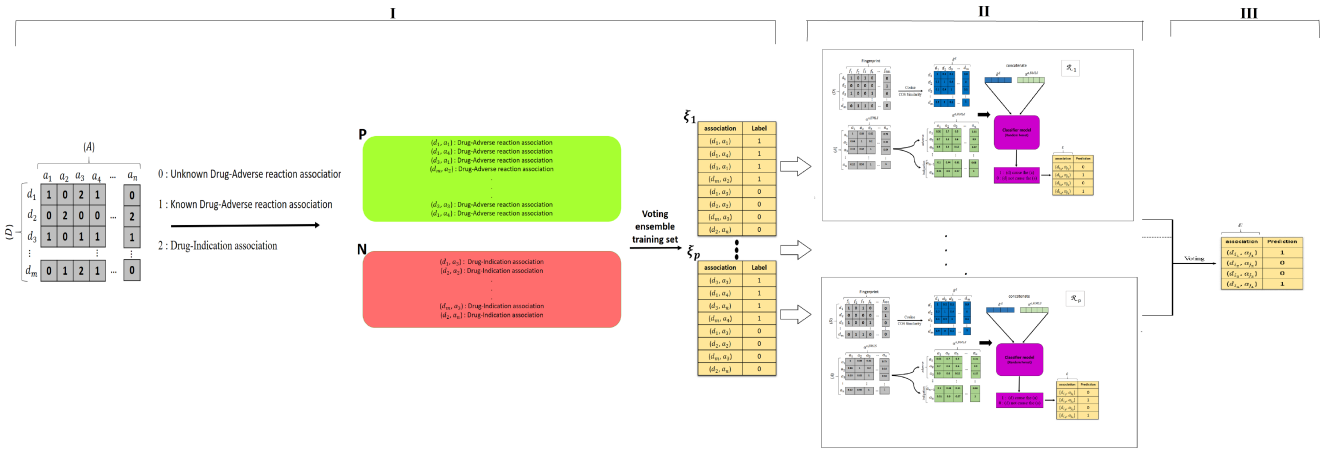


Figure 1: Outline of the ADRP-VETA frameworks.

3. The results and discussion

In this section, we evaluate the ADRP-VETA framework. The framework was implemented in MATLAB 2018b under Windows and Intel Core i5-2430M processor and 4GB of memory.
In the following, we introduce our selected evaluation criteria, then the framework parameters are explained. Next, we assess the performance of ADRP-VETA and compare it with related studies. Finally, we determine the effectiveness of our framework in predicting some case studies.

3.1. Evaluation criteria

We evaluate our framework using the area under the receiver operating characteristic curve ($AUC-ROC$) [13], the area under the precision-recall curve ($AUC-PR$) [27], accuracy (ACC), precision, recall, F1-score [23], specificity (SP) [2], and Matthews correlation coefficient (MCC) [16] criteria.
$AUC-ROC$ is obtained based on the false positive rate (FPR) and the true positive rate (TPR) under different classification thresholds, where

$$FPR=\frac{FP}{FP+TN}\ ,\ TPR=\frac{TP}{TP+FN},$$

and TP, FP, TN, and FN display true positive, false positive, true negative, and false negative, respectively (see Table 3).

Prediction	Definition
True Positive (TP)	the number of the known drug-adverse reaction associations predicted correctly by the model
False Positive (FP)	the number of drug-indication associations predicted wrongly by the model
True Negative (TN)	the number of drug-indication associations predicted correctly by the model
False Negative (FN)	the number of the known drug-adverse reaction associations predicted wrongly by the model

Table 3: The definition of TP, FP, TN and FN.

AUC-PR shows the relationship between sensitivity (recall) and positive predictive value (precision), where

$$precision=\frac{TP}{TP+FP}\ ,\ recall=\frac{TP}{TP+FN},$$

ACC indicates the rate of correct prediction to all predictions as below

$$ACC=\frac{TP+FN}{TP+FP+TN+FN}.$$

The F1-score is the harmonic mean of precision and recall

$$F1-score=\frac{2.TP}{2.TP+FP+FN}.$$

SP measures the probability of a negative test provided on TN. This score is defined as below

$$SP = \frac{TN}{TN + FP}.$$

Finally, MCC is applied to assess the quality of binary classification. The definition is as follows

$$MCC = \frac{(TP.TN) - (FP.FN)}{\sqrt{(TN + FN).(TN + FP).(TP + FN).(TP + FP)}}.$$

3.2. Parameters of the framework

ADRP-VETA is implemented in MATLAB 2018. The random forest classifier is located in the package Statistics and Machine Learning Toolbox [18] and has some hyperparameters that are changed according to the problem. Here, we refer to the three most important ones:

- "MinLeafSize" shows the minimum observations (samples) per leaf, which is essential in dividing the nodes in the decision trees. By default, this parameter is 1 for classification. A smaller number of "MinLeafSize" makes the model more prone to capturing noise in the training data.
- "NumPredictorsToSample" means the number of predictor or feature variables to select randomly for each decision split. By default, it equals the square root of the total number of variables for classification.
- "NumLearningCycles" variable represents the number of decision trees in the random forest.

Using trial and error, we define the $MinLeafSize = 1$, $NumPredictorsToSample = 6$ and $NumLearningCycles = 50$.

3.3. Model training and evaluations

To evaluate the effectiveness of ADRP-VETA, we perform three strategies: classic, common ADR-based, and rare ADR-based. The details of each strategy are available as follows.

3.3.1. Classic strategy

The classic strategy applies all rare and common adverse reactions. So, the positive P (see (5)) and negative N (see (6)) sets are given to the VETA model as input for making training and test sets to evaluate the ADRP-VETA performance.

3.3.2. Common ADR strategy

This strategy considers the positive data as:

$$P_{common} = \{< a, d > | a \in A_{common}, d \in D, \text{and drug } d \text{ causes ADR } a\},$$

and negative data as:

$$N_{common} = \{< a, d > | a \in A_{common}, d \in D, \text{and drug } d \text{ treats ADR } a\}.$$

where A_{common} is obtained based on (3).
Then we perform the VETA approach using P_{common} and N_{common} to generate training and test sets.

3.3.3. Rare ADR strategy

The rare ADR-based strategy utilizes P_{common} and N_{common} to train the ADRP-VETA model. Then, we apply rare adverse reaction associations to assess the model performance in predicting rare ADRs, where

$$P_{rare} = \{< a, d > | a \in A_{rare}, d \in D, \text{and drug d causes ADR a}\},$$

shows the positive test set and

$$N_{rare} = \{< a, d > | a \in A_{rare}, d \in D, \text{and drug d treats ADR a}\},$$

presents the negative test set, where A_{rare} is computed based on Eq. 2. Since the size of P_{rare} and N_{rare} are not equal, the test set is unbalanced.

3.4. The assessment of ADRP-VETA based on evaluation strategies

In this section, we assess the performance of our framework based on the explained strategies.

3.4.1. The assessment of ADRP-VETA based on the classic strategy

The classic strategy evaluates ADRP-VETA using all common and rare adverse reactions. According to Table 2, the drug-adverse reaction association set, the *positive set (P)*, includes 119040 pairs and the number of drug-indication associations, the *negative set (N)*, is equal to 10928. We randomly select $\frac{|N|}{10}$ (see Table 2), 1092 samples, from N as the test set for negative data. For the positive data test set, $\frac{|N|}{10}$ samples are chosen randomly from P. Then, we construct P_{train} and N_{train} from 117948 and 9836 samples, using the VETA approach, respectively. The size of positive data is significantly larger than that of negative data. To achieve a well-balanced dataset, it is essential to partition the positive data in a manner equal to the size of the negative data. The positive data is approximately 12 times larger than the negative data. Thus, we define the size of set τ as 12, $p = 12$. Therefore, the model is trained 12 times using the ADRP-VETA framework to predict each sample of $\epsilon = P_{test} \bigcup N_{test}$. This strategy achieves $ACC = 83.6\%$, $AUC - ROC$ =91.3% and $AUC - PR$ =89.9%. The other scores are available in Table 4. Moreover, Fig.2.a. illustrates the ROC curve for the ADRP-VETA framework to predict the test set based on each training set in the set τ.

ADRP-VETA	AUC-ROC	AUC-PR	ACC	Precision	Recall	F1-score	SP	MCC
Cluster 1 (ξ_1)	0.899	0.881	0.817	0.811	0.828	0.819	0.807	0.636
Cluster 2 (ξ_2)	0.904	0.890	0.824	0.816	0.836	0.826	0.812	0.648
Cluster 3 (ξ_3)	0.900	0.885	0.823	0.819	0.829	0.824	0.817	0.647
Cluster 4 (ξ_4)	0.898	0.881	0.818	0.807	0.826	0.820	0.811	0.637
Cluster 5 (ξ_5)	0.901	0.888	0.813	0.808	0.820	0.814	0.805	0.626
Cluster 6 (ξ_6)	0.898	0.885	0.816	0.814	0.820	0.817	0.812	0.632
Cluster 7 (ξ_7)	0.902	0.879	0.823	0.816	0.833	0.824	0.812	0.645
Cluster 8 (ξ_8)	0.904	0.885	0.828	0.819	0.842	0.830	0.814	0.656
Cluster 9 (ξ_9)	0.902	0.882	0.823	0.822	0.826	0.824	0.820	0.647
Cluster 10 (ξ_{10})	0.906	0.886	0.30	0.826	0.835	0.831	0.824	0.660
Cluster 11(ξ_{11})	0.902	0.884	0.827	0.822	0.834	0.828	0.819	0.653
Cluster 12 (ξ_{12})	0.899	0.882	0.823	0.815	0.837	0.826	0.810	0.647
Final output based on the voting	0.913	0.899	0.836	0.830	0.845	0.837	0.827	0.673

Table 4: The evaluation criteria of ADRP-VETA using classic strategy.

3.4.2. The assessment of ADRP-VETA based on the common ADR strategy

Since the limited associations of rare ADRs can reduce the performance of a model, some studies [33, 15] exclude rare ADRs and then perform the models. In this regard, we use the common ADR strategy to evaluate ADRP-VETA using common ADR associations. For this purpose, we consider P_{common} and N_{common} as the positive and negative sets. According to Table 2, the cardinality of P_{common} and N_{common} is 113287 and 8806, respectively. Then we randomly select $\frac{|N_{common}|}{10}$, 880 samples, from N_{common} as the test set for negative data. For the positive test set, $\frac{|N_{common}|}{10}$ samples are chosen randomly from P_{common}. The remained pairs are considered as N_{train} with 7926 negative pairs and P_{train} with 112407 positive associations. The size of positive data is about 15 times larger than the negative data. To make a balanced dataset, we group the positive data to match the size of the negative data. Thus, the positive data is grouped into 15 clusters, and the model is trained 15 times. The final output is reported based on voting all 15 predicted results. This strategy achieves $ACC = 83.1\%$, $AUC - ROC$ =91.4% and $AUC - PR$ =88.9%. The other scores are presented in Table 5. Moreover, Fig.2.b. shows the ROC curve for the ADRP-VETA framework using the common ADR strategy to predict the test set based on each training set.

ADRP-VETA	AUC-ROC	AUC-PR	ACC	Precision	Recall	F1-score	SP	MCC
Cluster 1 (ξ_1)	0.900	0.868	0.825	0.791	0.825	0.808	0.824	0.647
Cluster 2 (ξ_2)	0.901	0.870	0.824	0.784	0.836	0.809	0.814	0.648
Cluster 3 (ξ_3)	0.896	0.869	0.819	0.789	0.812	0.800	0.824	0.634
Cluster 4 (ξ_4)	0.901	0.870	0.820	0.782	0.828	0.804	0.814	0.639
Cluster 5 (ξ_5)	0.899	0.865	0.816	0.777	0.824	0.800	0.809	0.630
Cluster 6 (ξ_6)	0.905	0.877	0.826	0.790	0.833	0.810	0.821	0.651
Cluster 7 (ξ_7)	0.903	0.876	0.826	0.788	0.834	0.810	0.819	0.650
Cluster 8 (ξ_8)	0.897	0.867	0.817	0.783	0.815	0.799	0.818	0.631
Cluster 9 (ξ_9)	0.900	0.865	0.830	0.793	0.839	0.815	0.823	0.659
Cluster 10 (ξ_{10})	0.903	0.877	0.828	0.787	0.842	0.814	0.816	0.655
Cluster 11 (ξ_{11})	0.902	0.866	0.820	0.777	0.835	0.805	0.807	0.640
Cluster 12 (ξ_{12})	0.899	0.867	0.821	0.799	0.820	0.804	0.823	0.641
Cluster 13 (ξ_{13})	0.898	0.857	0.820	0.781	0.828	0.803	0.812	0.637
Cluster 14 (ξ_{14})	0.895	0.864	0.818	0.777	0.830	0.802	0.808	0.635
Cluster 15 (ξ_{15})	0.891	0.857	0.801	0.846	0.679	0.753	0.900	0.600
Final output based on the voting	0.914	0.889	0.831	0.797	0.836	0.816	0.829	0.661

Table 5: The evaluation criteria of ADRP-VETA using the common ADR strategy.

3.4.3. The assessment of ADRP-VETA based on the rare ADR strategy

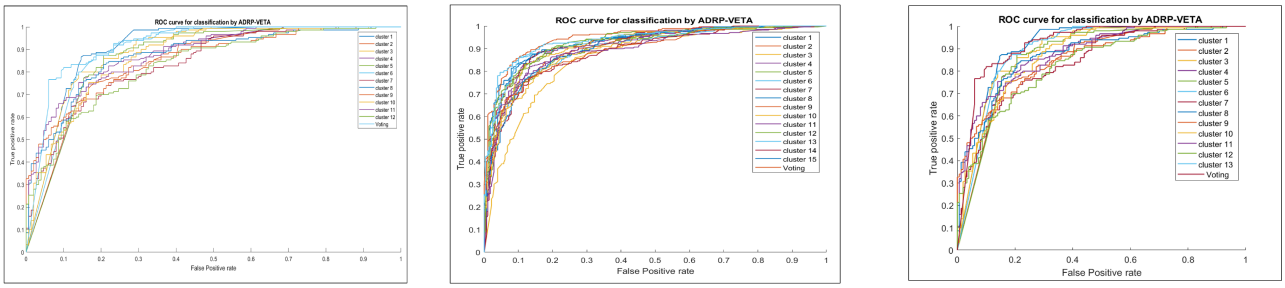
We investigate the efficacy of ADRP-VETA in predicting associations between drugs and rare adverse drug reactions, which other methods often overlook. For this purpose, we use the rare ADR strategy which trains the model using common ADRs. In this manner, P_{common} and N_{common} are considered as P_{train} and N_{train}, respectively. The size of the positive data, $|P_{train}| = 113287$, is about 13 times larger than the size of negative data, $|N_{train}| = 8806$. Afterwards, we divide P_{train} into 13 clusters. Then the P_{rare} and N_{rare} are considered as P_{test} and N_{test}, respectively. According to Table 2, P_{test} includes 5753 drug-adverse reaction associations and N_{test} includes 2122 drug-indication associations. To predict the output of every sample $\epsilon = P_{test} \bigcup N_{test}$, we vote on all 13 predicted results. This strategy achieves $ACC = 74.4\%$, $AUC - ROC = 83.3\%$ and $AUC - PR = 92.2\%$. The other scores are presented in Table 6. Moreover, Fig.2.c shows the ROC curve for the ADRP-VETA framework using the rare ADR strategy to predict test set based on each training set.

ADRP-VETA	AUC-ROC	AUC-PR	ACC	Precision	Recall	F1-score	SP	MCC
Cluster 1 (ξ_1)	0.812	0.912	0.727	0.900	0.715	0.793	0.760	0.428
Cluster 2 (ξ_2)	0.814	0.912	0.723	0.891	0.707	0.788	0.767	0.425
Cluster 3 (ξ_3)	0.809	0.911	0.722	0.885	0.713	0.790	0.749	0.416
Cluster 4 (ξ_4)	0.806	0.909	0.710	0.889	0.689	0.776	0.766	0.406
Cluster 5 (ξ_5)	0.818	0.913	0.730	0.898	0.712	0.794	0.780	0.441
Cluster 6 (ξ_6)	0.815	0.912	0.729	0.898	0.710	0.793	0.780	0.440
Cluster 7 (ξ_7)	0.827	0.920	0.733	0.897	0.718	0.797	0.777	0.445
Cluster 8 (ξ_8)	0.812	0.911	0.725	0.892	0.710	0.791	0.768	0.430
Cluster 9 (ξ_9)	0.816	0.913	0.726	0.891	0.712	0.791	0.763	0.428
Cluster 10 (ξ_{10})	0.818	0.915	0.740	0.894	0.731	0.804	0.764	0.448
Cluster 11 (ξ_{11})	0.818	0.913	0.722	0.896	0.700	0.787	0.780	0.431
Cluster 12 (ξ_{12})	0.812	0.913	0.720	0.886	0.707	0.787	0.754	0.415
Cluster 13 (ξ_{13})	0.811	0.910	0.701	0.900	0.665	0.765	0.799	0.413
Final output based on the voting	0.833	0.922	0.744	0.903	0.727	0.816	0.788	0.465

Table 6: The evaluation criteria ADRP-VETA for rare ADR.

3.5. Comparison with Related Studies

We compare ADRP-VETA framework with the two machine learning models, Zhao [33] and Zheng [34], that predict drug-adverse reaction associations using similarity-based methods and classic classifiers. In addition, we compare ADRP-VETA with the logistic regression model introduced by Zhang's method [30], our negative data is define as drug-indication associations similar to this model. It should be noted that all these models apply SIDER database as the primary dataset. Table 7 illustrates the values of the evaluation criteria for each method. The advantage



a) ROC curve of ADRP-VETA using classic strategy

b) ROC curve of ADRP-VETA using common strategy

c) ROC curve of ADRP-VETA using rare strategy

Figure 2: ROC curve of ADRP-VETA using different strategies.

of ADRP-VETA over other models is applying a voting ensemble training approach to make the final decision on the test set samples. Moreover, it assesses the performance of the model on rare and common ADRs by exploiting different strategies, separately.

Model	Drug feature	Adverse reaction feature	AUC-ROC	AUC-PR
ADRP-VETA, classic	Fingerprint	UMLS	0.910	0.898
ADRP-VETA, common ADR	Fingerprint	UMLS	**0.913**	**0.899**
ADRP-VETA, rare ADR	Fingerprint	UMLS	0.833	0.922
Zheng et al. [34]	Fingerprint, Target protein, Substituent, ATC code	-	0.908	0.542
Zhang et al. [30]	Target Protein, Drug Bank ID	CUI code	0.87	-
Zhao et al. [33]	ATC code, literature (STITCH), Target protein	Drug profile	0.801	-

Table 7: Comparison of the proposed model with other related studies.

The ADRP-VETA framework is more efficient than Zheng et al.[34] model, while this model has been trained on each adverse reaction separately and applied more features.
Our approach as a classifier outperforms the Zhang's model which is a regressor as evidenced by the impressive results obtained while utilizing Zhang's technique for selecting negative data [30]. Finally, we evaluate the performance of the ADRP-VETA framework against the model proposed by Zhao et al. [33]. This model considered ADRs associated to less than five drugs as rare, similar to ADRP-VETA. In this manner, we can compare the corresponding results of ADRP-VETA and Zhao's model based on three evaluation strategies. The visual comparison of Zhao's model and ADRP-VETA framework is available in Fig.3.
By the classic strategy, the performance of ADRP-VETA is significantly more accurate than Zhao's according to corresponding scores. Based on the common ADR strategy, ADRP-VETA and Zhao's method exhibit comparable performance. Nevertheless ADRP-VETA has a slight advantage in outperforming the latter. In the rare ADR strategy, Zhao et al. predicted 100 rare ADRs, including 548 associations, while ADRP-VETA predicted 3318 rare ADRs with 5753 associations. Since the considered rare ADRs by Zhao's model is less than ours, the Zhao's achieve higher scores in some criteria, including MCC and ACC, which are related to the classifier performance. However, correctly predicting positive samples is more critical than negatives in rare ADR strategies. The precision criterion shows the power of predicting drug-adverse reaction associations correctly. The SP indicates how the model can estimate drug-indication associations correctly. ADRP-VETA achieves better scores based on these criteria in predicting rare ADRs.

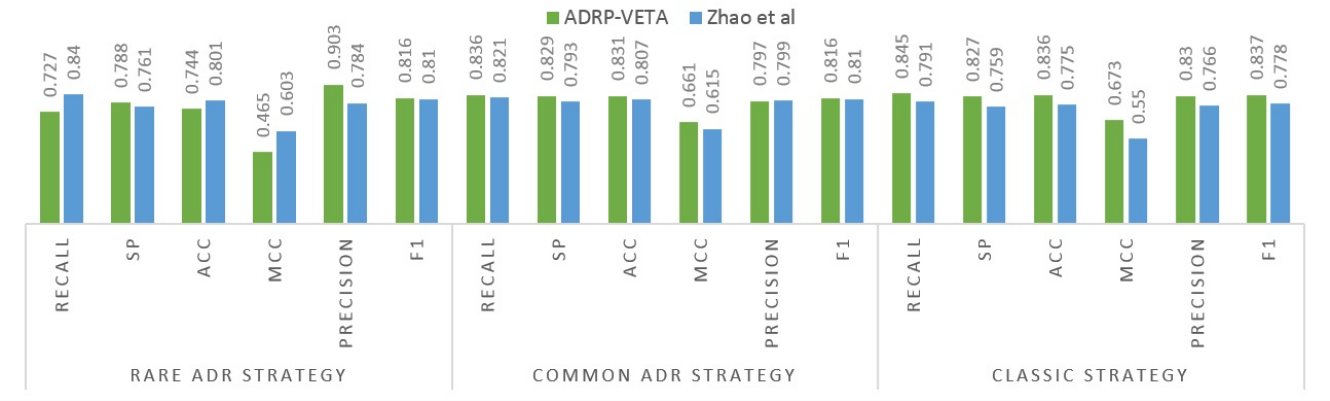


Figure 3: Comparison of the proposed model in three strategies with Zhao's model.

3.6. Discussion

We use Atorvastatin, Albendazole, and Crizotinib as case studies to test the performance of the ADRP-VETA model in predicting liver-injury. We utilized ADRP-VETA framework based on the classic strategy to predict the association between these drugs and liver-injury ADRs. Although the SIDER dataset does not provide information on the association between these drugs and the liver-injury, which has been clinically proven in the literature [20, 19, 14], our model is able to accurately estimate these associations (see Table 8). Each pair of drug and adverse reaction is given to p trained classifiers. Then, each classifier makes a probability score in range of 0 to 1 and decides whether the given input (drug-adverse reaction pair) is associated (if the score > 0.5) or not (otherwise)". Finally, we use the voting system of classifiers to predict the association for the given pair. ADRP-VETA predicted the associations of Atorvastatin-, Albendazole-, and Crizotinib-liver injury with the scores of 0.60, 0.57 and 0.79, respectively.

Drug	Adverse Reaction	Type	Evidence
Atorvastatin (CID: 2250)	liver-injury (CUI:C0160390)	liver disorder	drug-adverse reaction ([14])
Albendazole (CID: 2082)	liver-injury (CUI:C0160390)	liver disorder	drug-adverse reaction ([19])
Crizotinib (CID: 11597571)	liver-injury (CUI: C0160390)	liver disorder	drug-adverse reaction ([20])

Table 8: Adverse drug reactions obtained based on a case study by ADRP-VETA.

4. Conclusion

This study propose a framework called ADRP-VETA based on a random forest classifier and voting ensemble model to predict drug-adverse reaction associations. For this aim, a similarity vector is suggested by the drug-drug similarity score and the adverse reactions similarity as the drug and adverse reaction representations, respectively. As the performance of machine learning methods depends on the training data, similar to Zhang et al. [30], the drug-adverse reactions and drug indication are considered positive and negative data, respectively. The corresponding results of ADRP-VETA show its prediction power against other models with the $AUC - ROC = 0.91$ and $AUC - PR = 0.898$.

We conclude that using similarity vectors as drug and adverse reaction features, considering drug indications as negative data and training the model using the voting ensemble approach can enhance predicting drug-adverse reaction associations. ADRP-VETA reaches this achievement while unlike other studies that use complex models, it only uses the simple Random Forest classifier with less computational complexity.

In the future, we aim to assess the 3D structures of drugs to increase the performance of drug-adverse reaction

association prediction. In addition, applying drug-related clinical information can improve the performance of the model.

5. List of symbols

Notation	Description
D	Set of drugs
d_i	i^{th}drug of D
n	Number of adverse drug reactions
F^d	Fingerprint representation
$Cos(d_i, d_j)$	Cosine similarity of drug d_i and d_j
δ^d	Feature representation of drug d
A	Set of adverse drug reactions
a_j	j^{th} adverse drug reaction of A
m	number of drugs
$UMLS(a_i, a_j)$	UMLS-based similarity of adverse drug reaction a_i and a_j
α^a	Feature representation of adverse drug reaction a
P	Set of drug-adverse reaction associations as *positive set*
N	Set of drug-indication associations as *negative set*
τ	Voting ensemble training set
ϵ_i	i^{th}training set of τ
ϵ	The test set

References

[1] *Coronavirus (COVID-19) update: FDA revokes emergency use authorization for chloroquine and hydroxychloroquine.* `https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-revokes-emergency-use-authorization-chloroquine-and`, 2020. Accessed: September 2021.

[2] D. G. Altman and J. M. Bland, *Statistics notes: Diagnostic tests 1: sensitivity and specificity*, BMJ, 308 (1994), pp. 1552–1552.

[3] P. Bansal, A. Goyal, A. Cusick, S. Lahan, H. S. Dhaliwal, P. Bhyan, P. B. Bhattad, F. Aslam, S. Ranka, T. Dalia, L. Chhabra, D. Sanghavi, B. Sonani, and J. M. Davis, *Hydroxychloroquine: a comprehensive review and its controversial role in coronavirus disease 2019*, Annals of Medicine, 53 (2020), pp. 117–134.

[4] L. Breiman, *Bagging predictors*, Machine Learning, 24 (1996), pp. 123–140.

[5] L. Chen, T. Huang, J. Zhang, M.-Y. Zheng, K.-Y. Feng, Y.-D. Cai, and K.-C. Chou, *Predicting drugs side effects based on chemical-chemical interactions and protein-chemical interactions*, BioMed Research International, 2013 (2013), pp. 1–8.

[6] S. Dey, H. Luo, A. Fokoue, J. Hu, and P. Zhang, *Predicting adverse drug reactions through interpretable deep learning framework*, BMC Bioinformatics, 19 (2018).

[7] D. Galeano and A. Paccanaro, *A recommender system approach for predicting drug side effects*, in 2018 International Joint Conference on Neural Networks (IJCNN), 2018, pp. 1–8.

[8] X. Guo, W. Zhou, Y. Yu, Y. Ding, J. Tang, and F. Guo, *A novel triple matrix factorization method for detecting drug-side effect association based on kernel target alignment*, BioMed Research International, 2020 (2020), pp. 1–11.

[9] J. Han, M. Kamber, and J. Pei, *Data mining concepts and techniques third edition*, University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University, 2012.

[10] M. Khan, *Drug side-effect prediction using machine learning methods*, 2017.

[11] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, and E. E. Bolton, *PubChem 2019 update: improved access to chemical data*, Nucleic Acids Research, 47 (2019), pp. D1102–D1109.

[12] M. Kuhn, I. Letunic, L. J. Jensen, and P. Bork, *The SIDER database of drugs and side effects*, Nucleic Acids Research, 44 (2016), pp. D1075–D1079.

[13] R. Kumar and A. Indrayan, *Receiver operating characteristic (ROC) curve for medical researchers*, Indian Pediatrics, 48 (2011), pp. 277–287.

[14] N. Kuniyoshi, H. Miyakawa, K. Matsumoto, H. Tsunashima, K. Sekine, T. Tsujikawa, M. Mabuchi, S. Doi, and K. Kikuchi, *Detection of anti-mitochondrial antibodies accompanied by drug-induced hepatic injury due to atorvastatin*, Internal Medicine, 58 (2019), pp. 2663–2667.

[15] H. Liang, L. Chen, X. Zhao, and X. Zhang, *Prediction of drug side effects with a refined negative sample selection strategy*, Computational and Mathematical Methods in Medicine, 2020 (2020), pp. 1–16.

[16] B. W. Matthews, *Comparison of the predicted and observed secondary structure of t4 phage lysozyme*, Biochimica et Biophysica Acta (BBA) - Protein Structure, 405 (1975), pp. 442–451.

[17] B. T. McInnes, T. Pedersen, and S. V. Pakhomov, *UMLS-interface and UMLS-similarity: open source software for measuring paths and semantic similarity*, in AMIA annual symposium proceedings, vol. 2009, American Medical Informatics Association, 2009, pp. 431–435.

[18] N. Meinshausen, *Quantile regression forests*, Journal of Machine Learning Research, 7 (2006), pp. 983–999.

[19] S. Y. Moon, Y. H. Baek, and S. W. Lee, *Drug induced liver injury by prophylactic administration of albendazole*, The Korean Journal of Gastroenterology, 73 (2019), pp. 360–364.

[20] T. Ota, N. Masuda, K. Matsui, T. Yamada, N. Tanaka, S. Fujimoto, and M. Fukuoka, *Successful desensitization with crizotinib after crizotinib-induced liver injury in ROS1-rearranged lung adenocarcinoma*, Internal Medicine, 58 (2019), pp. 2651–2655.

[21] A. Poleksic and L. Xie, *Predicting serious rare adverse reactions of novel chemicals*, Bioinformatics, 34 (2018), pp. 2835–2842.

[22] K. B. Pouwels and K. van Grootheest, *The rosiglitazone decision process at FDA and EMA. what should we learn?*, International Journal of Risk & Safety in Medicine, 24 (2012), pp. 73–80.

[23] D. M. W. Powers, *Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation*, (2020).

[24] P. Resnick and H. R. Varian, *Recommender systems*, Communications of the ACM, 40 (1997), pp. 56–58.

[25] P. P. Rodrigues, D. Ferreira-Santos, A. Silva, J. Polónia, and I. Ribeiro-Vaz, *Causality assessment of adverse drug reaction reports using an expert-defined bayesian network*, Artificial Intelligence in Medicine, 91 (2018), pp. 12–22.

[26] S. Shabani-Mashcool, S.-A. Marashi, and S. Gharaghani, *NDDSA: A network- and domain-based method for predicting drug-side effect associations*, Information Processing & Management, 57 (2020), pp. 102357.

[27] H. R. Sofaer, J. A. Hoeting, and C. S. Jarnevich, *The area under the precision-recall curve as a performance metric for rare binary events*, Methods in Ecology and Evolution, 10 (2019), pp. 565–577.

[28] O. C. Uner, H. I. Kuru, R. G. Cinbis, O. Tastan, and E. Cicek, *DeepSide: A deep learning approach for drug side effect prediction*, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 20 (1) (2022), pp. 330–339.

[29] A. J. Wagstaff and K. L. Goa, *Rosiglitazone: : a review of its use in the management of type 2 diabetes mellitus*, Drugs, 62 (2002), pp. 1805–1837.

[30] F. Zhang, B. Sun, X. Diao, W. Zhao, and T. Shu, *Prediction of adverse drug reactions based on knowledge graph embedding*, BMC Medical Informatics and Decision Making, 21 (2021), pp. 1–11.

[31] H. Zhao, J. Wald, M. Palmer, and Y. Han, *Hydroxychloroquine-induced cardiomyopathy and heart failure in twins*, J. Thorac Dis., 10 (2018).

[32] H. Zhao, S. Wang, K. Zheng, Q. Zhao, F. Zhu, and J. Wang, *A similarity-based deep learning approach for determining the frequencies of drug side effects*, Briefings in Bioinformatics, 23 (2021).

[33] X. Zhao, L. Chen, and J. Lu, *A similarity-based method for prediction of drug side effects with heterogeneous information*, Mathematical Biosciences, 306 (2018), pp. 136–144.

[34] Y. Zheng, H. Peng, S. Ghosh, C. Lan, and J. Li, *Inverse similarity and reliable negative samples for drug side-effect prediction*, BMC Bioinformatics, 19 (2019).

[35] B. Zhou, X. Zhao, J. Lu, Z. Sun, M. Liu, Y. Zhou, R. Liu, and Y. Wang, *Relating substructures and side effects of drugs with chemical-chemical interactions*, Combinatorial Chemistry & High Throughput Screening, 23 (2020), pp. 285–294.

Please cite this article using:

Milad Besharatifard, Zahra Ghorbanali, Fatemeh Zare-Mirakabad, Adverse drug reaction prediction using voting ensemble training approach, AUT J. Math. Comput., 5(1) (2024) 45-59
https://doi.org/10.22060/AJMC.2023.21538.1091