Nguyen, Duc Anh; Nguyen, Canh Hao; Mamitsuka, Hiroshi

A survey on adverse drug reaction studies

Published in:
Briefings in Bioinformatics

DOI:
10.1093/bib/bbz140

Published: 01/01/2021

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Nguyen, D. A., Nguyen, C. H., & Mamitsuka, H. (2021). A survey on adverse drug reaction studies: Data, tasks and machine learning methods. *Briefings in Bioinformatics*, *22*(1), 164-177. https://doi.org/10.1093/bib/bbz140

Briefings in Bioinformatics
doi: 10.1093/bib/bbz140
Review article

Drug discovery

A survey on adverse drug reaction studies: data, tasks, and machine learning methods

Duc Anh Nguyen, Canh Hao Nguyen and Hiroshi Mamitsuka

Corresponding author. E-mail: ducanh@kuicr.kyoto-u.ac.jp

Abstract

Motivation: Adverse drug reaction (ADR) or drug side effect studies play a crucial role in drug discovery. Recently, with the rapid increase of both clinical and non-clinical data, machine learning methods have emerged as prominent tools to support analyzing and predicting adverse drug reactions. Nonetheless, there are still remaining challenges in ADR studies.
Results: In this paper, we summarized ADR data sources and review ADR studies in three tasks: drug-ADR benchmark data creation, drug-ADR prediction, and ADR mechanism analysis. We focused on machine learning methods used in each task and then compare performances of the methods on the drug-ADR prediction task. Finally, we discussed open problems for further ADR studies.
Availability: Data and code are available at https://github.com/anhnda/ADRPModels.
Supplementary: Supplementary materials are available at *Briefings in Bioinformatics* online.

Keywords: adverse drug reaction, ADR prediction, ADR mechanism, machine learning methods

Introduction

According to WHO, an adverse drug reaction (ADR) or drug side effect is a response to a medicine which is noxious and unintended, and which occurs at doses normally used in human [1]. In reports of 2011, ADRs accounted for nearly 6% of total hospitalizations in the USA, which cost billions of dollars and was responsible for significant patient morbidity and mortality [2, 3]. Therefore, the studies of ADRs are important in drug discovery.

The traditional methods for obtaining ADRs of drugs often use clinical trials or post-marketing surveillance reports [4]. However, these methods are costly and time-consuming. To deal with these disadvantages, machine learning methods integrating various kinds of ADR data sources are used to make inexpensive and fast predictions. These results provide potential ADRs and their mechanism analysis for further clinical verification to enhance ADR studies.

Data sources used in ADR studies consist of clinical and non-clinical data. The clinical data contains observations of ADRs from clinical treatments of patients. These observations have not only adverse drug reactions but also personal contexts, such as dosages of treatments, ages, genders, and diseases of patients. Since different patients can have different adverse drug reactions, such personal contexts can support to build personalized ADR prediction models.

The non-clinical data contains information of biological systems such as drug-protein interactions and biological processes. In fact, there are various possible mechanisms in ADRs, for example, by interactions of drugs with proteins, but the details of these mechanisms are still unknown [5, 6]. By integrating clinical data with non-clinical data, it is expected that the quality of ADR studies will be improved, and ADR mechanisms can be revealed.

Since there are different machine learning methods using various kinds of ADR data sources, an overview of current methods in ADRs is necessary. Table 1 summaries the most recent survey papers related to ADR studies. These studies often use either clinical data [7] or non-clinical data [8]. There is only one survey which uses both kinds of data [9], but there is no detailed analysis on methods, such as providing a taxonomy or conducting experiments to compare performances of methods. Recently,

Duc Anh Nguyen is currently a PhD student at the Bioinformatics Center in Kyoto University. His current research interests focus on machine learning and bioinformatics.
Canh Hao Nguyen is an assistant professor at the Bioinformatics Center, Institute for Chemical Research, Kyoto University, working on machine learning for graph data, with applications to biological networks.
Hiroshi Mamitsuka is a professor at the Bioinformatics Center, Institute for Chemical Research, Kyoto University, and a FiDiPro professor at the Department of Computer Science, Aalto University, working on machine learning, data mining and application to biology and chemistry.
Submitted: 07 Month 2019; **Received (in revised form)**: 09 Month 2019

Table 1. Recent surveys on ADR studies.

Paper	Task		Data		Method analysis
	Clinical data extraction	Drug-ADR prediction	Clinical data	Drug-ADR & non-clinical data	
Poloju et al., 2018 [7]	✓		✓		
Chen et al., 2016 [8]		✓		✓	✓
Ho et al., 2016 [9]	✓	✓	✓	✓	

there are new studies in ADRs with the emerging of using machine learning methods, leading to a need for a more detailed classification for these methods. Moreover, ADR studies are not only drug-ADR prediction [8] but also analyzing ADR mechanisms by revealing biological components associated with ADRs [10]. Motivated by this, we give a broader view of ADR studies containing ADR data sources and how computational tasks of ADRs use these kinds of data.

The contributions of our paper can be summarized as follows. 1) We summarize the ADR data sources containing both clinical and non-clinical data. 2) We summarize a wide range of drug descriptors used in ADR studies. 3) We analyze methods used in ADR studies in three main tasks: (i) drug-ADR benchmark data creation, (ii) drug-ADR prediction, and (iii) ADR mechanism analysis (We focus on papers on the main journals with the most numbers of papers on this topic such as Bioinformatics, BMC Informatics, Briefing in Bioinformatics, and Nucleic acid research, then we follow cited papers. Papers are collected up to Feb 2019.). In each task, we analyze data and commonly used machine learning methods. 4) We conduct an experiment to compare the drug-ADR prediction performances of eight commonly used methods.

The organization of the paper is as follows: Section 2 presents the data sources used in ADR studies. Section 3 details different kinds of drug descriptors that encode drug information. ADR studies with tasks and methods are detailed in Section 4. Finally, discussions on current ADR studies and open problems are presented in Section 5.

Data sources in ADR studies

In this section, we summarize commonly used data sources in ADR studies. Fig. 1 illustrates a hierarchical classification of data sources in ADR studies containing two groups: clinical and non-clinical data.

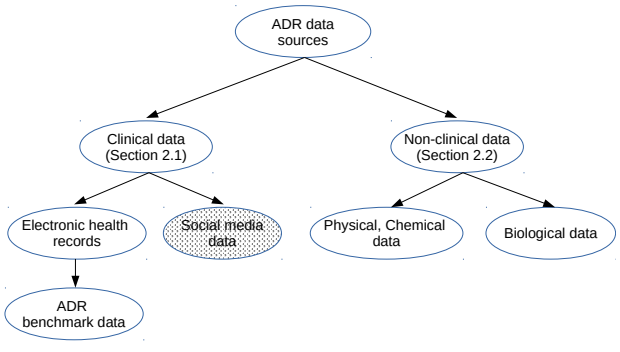

Fig. 1: Data source hierarchy in ADR studies.

Table 2. Commonly used clinical data sources.

Data sources	Personal context	Drug-ADR benchmark	
		Monopharmacy	Polypharmacy
FAERS [13]	✓		
OMOP-CDM [15]	✓		
SIDER [16]		✓	
Liu' dataset [17]		✓	
AEOLUS [18]		✓	
OFFSIDES [19]		✓	
TWOSIDES [19]			✓

Table 3. Commonly used non-clinical databases.

Database	Elements					Having Interactions
	Chemical / Drug	Protein / Gene	Pathway	ADR term	Disease	
DrugBank [20]	✓	✓				✓
PubChem [21]	✓					
PDB [22]		✓				
BindingDB [23]	✓	✓				✓
HPRD [24]		✓				✓
CTD [25]	✓	✓			✓	✓
KEGG [26]	✓	✓	✓		✓	✓
SuperTarget [27]	✓	✓				✓
ADReCS [28]				✓		
DART [29]	✓	✓		✓		✓
TTD [30]	✓	✓	✓		✓	✓
Bio2RDF [31]	✓	✓	✓	✓	✓	✓

Clinical data

Clinical data contains observations of ADRs in clinical treatments, which are often electronic health records (EHR) or records from adverse report systems. Each record contains drugs and observed ADRs. In addition, personal contexts such as demographic and dosage information are also stored. There is evidence that ADRs are different from different patients [11], therefore, these personal contexts are important to build personalized ADR prediction models [12].

Table 2 provides the commonly used clinical data sources. For personal contexts, it has FDA Adverse Event Reporting System (FAERS) [13] and Medical Outcomes Partnership Common Data Model (OMOP CDM) [14]. There are four main tables in FAERS: demographics, drug, therapy, and reaction. The demographics table describes patient information containing patient identification, age, gender, weight, location, and other related information. The amount and routes of drug administration with patient identifications come from the drug table, and the time of drug treatments is from the drug therapy table. The reaction table contains the drug adverse reactions with patient identifications.

OMOP CDM is a data model provided by Observational Health Data Sciences and Informatics [15], which is an international collaboration with the aim to create and apply data analytic solutions to a large number of observational health databases. There are four domains of OMOP CDM v5.0: standardized clinical data, standardized health system data, standardized health economics data, and standardized derived elements. Standardized clinical data contains the core information with clinical events and demographic information of patients. With OMOP CDM, millions of health records from different resources are transformed into pre-defined tables of the four domains, supporting further analysis [32].

FAERS was used to extract drug-ADR benchmark datasets, which contain reliable drug-ADR associations [16, 19]. SIDER, a common ADR

benchmark dataset for many ADR studies, was extracted from FAERS for ADRs caused by single drugs (monopharmacy) [16]. Liu' dataset [17] is a benchmark dataset extracted from SIDER into the binary format with additional drug information. AEOLUS is also a monopharmacy dataset extracted from FAERS, and has more drug-ADR associations than SIDER. Extracting from FAERS with a criteria of removing bias data, OFFSIDES for ADRs caused by single drugs, and TWOSIDES for ADRs caused by combinations of two drugs (polypharmacy) were created [19]. However, SIDER, AEOLUS, OFFSIDES, and TWOSIDES only contain two kinds of information: drugs and ADRs. As far as we know, there is no benchmark ADR data for academic research that contains personal contexts such as diseases, duration of drug treatments.

In recent years, data from social media such as Facebook, Twitter is another kind of data to analyze ADR. This social media data contains comments of patients during drug treatments. However, the tasks on this kind of data are mainly ADR identification using techniques of natural language processing [33–35], which are considered as data pre-processing steps, therefore, we do not cover them in this survey.

Non-clinical data

The non-clinical data contains information of chemical, physical, biological properties of drugs and biological systems, which can help revealing mechanisms of drugs and ADRs. In fact, ADRs are results of complex reactions of drugs with biological components. Some studies have shown that drug side effects can be the results of reactions of drug chemicals with proteins [5, 6, 36], which interrupts normal biological processes leading to abnormal reactions of human bodies. By using this kind of data, we can improve the performance of models and extract possibly associated biological components with ADRs.

Table 3 summarize the commonly used non-clinical databases in ADR studies in two aspects: elements in each database and interactions among elements existed or not. For example, ADReCS [28] is a database for only ADR term definitions, and KEGG [26] contains information of proteins, drugs, biological pathways, diseases, and interactions among them such as drugs with proteins targets. To link these databases, Bio2RDF [31] provides interconnections among elements of different databases.

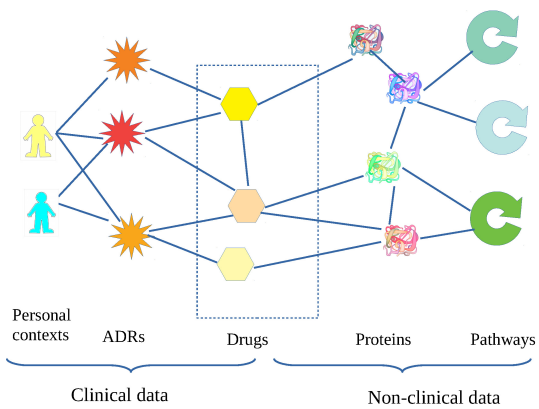

Fig. 2: A network for clinical and non-clinical data.

Finally, the connection between clinical and non-clinical data can be illustrated by a network in Fig. 2. The clinical data provides information of drug-ADR connections with personal contexts. The non-clinical data contains connections of drug-drug, drug-protein, protein-protein, and protein-biological pathway. This network is used to support some computational tasks represented in Section 4.

Drug descriptors

One possible way of encoding drugs is to use descriptors, which are physical, chemical, and biological characteristics of a drug. Since the quality of these descriptors impacts ADR prediction performances, the understanding of drug descriptors is a basic need. Fig. 3 presents a classification for drug descriptors. In general, drug descriptors can be categorized into two classes: physical or chemical descriptors (PC-descriptors) and biological descriptors (BIO-descriptors).

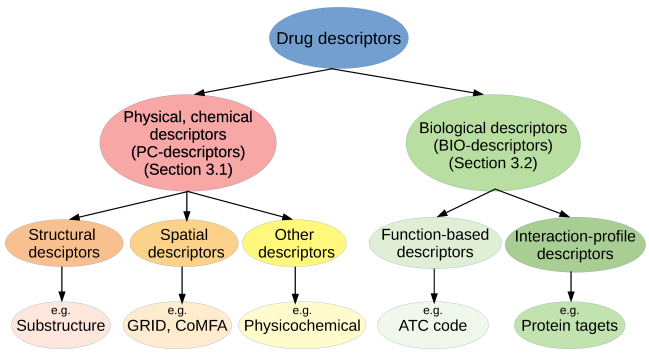

Fig. 3: Different kinds of drug descriptors.

Physical or chemical (PC) descriptors

The PC-descriptors describe the structure of drug molecules and their physical, chemical properties [37–39]. Based on their dimensionalities and properties, this class of descriptors can be divided into 3 subgroups: structural descriptors, spatial descriptors, and other miscellaneous descriptors.

Table 4. Two groups of structural descriptors implemented in CDK [40].

Group	Name	Number of descriptors
Variable-size	Daylight family [41]	-
Fixed-size	E-State fragments [42]	79
	Klekota-Roth [43]	4860
	MACCS keys [44]	166
	PubChem descriptors [21]	881
	CDK substructures [40]	307

The structural descriptors describe features of molecular structures such as atom counters, atom pairs, rings, and other substructures. Table 4 presents two groups of structural descriptors (fingerprints) implemented in Chemistry Development Kit (CDK) [40]: variable-size and fixed-size groups. The former group generates substructures from a given set of molecules, in which the number of substructures can be changed depending on the provided molecule set [41]. In contrast, the latter group uses pre-defined substructures, for example, MACCS keys and PubChem descriptors. An illustration of PubChem descriptors is shown in Fig. 4. The PubChem descriptors contain pre-defined 881 bits, which are divided into seven sections with corresponding bits. For instance, bit 308, which belongs to section 3 of simple atom pairs, indicates the existence of O-H connection.

The spatial descriptors describe spatial properties of drug molecules. In PubChem 3D database [45], 3D conformers descriptors of molecules are used. These descriptors are calculated by OMEGA [46], a tool

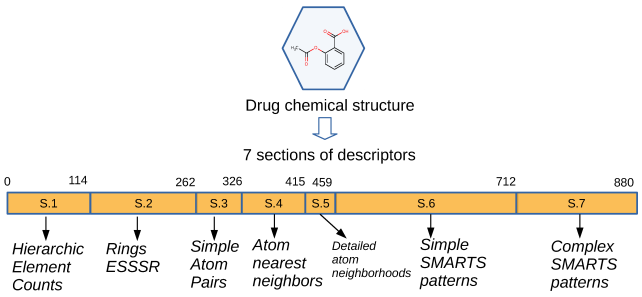

Fig. 4: Seven sections in PubChem descriptors.

published by OpenEye. Molecular interaction fields (MIFs) are another kind of spatial descriptors for drugs. MIFs describe spatial variation of the interaction energy between a molecular target and a chosen probe. Probes are small molecules representing common interactions such as hydrophobic, hydrogen bond donors and acceptors [47]. Some well known MIFs are GRID [48], VolSurf [49], CoMFA [50], and MetaSite [51]. Fig. 5 illustrates the idea of GRID descriptors. A molecule is put into a cube with grids. An empirical energy function will be used to calculate the interaction field of each cell at position (x, y, z) of the cube. The energy function is defined by:

$$E_{xyz} = \sum E_{lj} + \sum E_{el} + \sum E_{hb}$$

where E_{lj}, E_{el}, and E_{eb} are the Lennard-Jones function, the electronic function, and the hydrogen bound function, respectively [48].

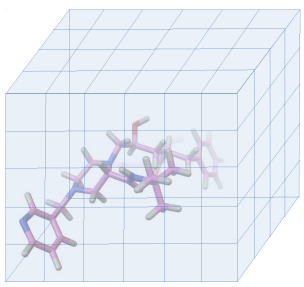

Fig. 5: A molecule with 3D GRID.

Other miscellaneous descriptors such as physicochemical properties of drugs also affect the action of drugs. Lipophilicity [52, 53] impacts solubility, absorption, distribution, membrane penetration and plasma protein binding of drugs. Hydrogen bond [54] is another physical property of electrostatic attraction, which takes two out of five Lipinski's rules [55]. Size/geometric features of drugs such as molecular weight and atom counters can also reflect drug properties.

Biological (BIO) descriptors

The BIO-descriptors describe biological properties of drugs, which can be classified into two subgroups: function-based descriptors and interaction-profile descriptors. The function-based descriptors describe purposes of drugs in therapy. ATC code [56], which is a classification system for drugs based on therapeutic properties, is a typical example of function-based descriptors.

The interaction-profile descriptors describe associated biological components of drugs containing protein targets and associated biological pathways of drugs [17, 57]. These interaction-profile descriptors are taken from the databases having drug interaction information in Table 3, such as DrugBank [20], BindingDB [23], and Bio2RDF [31].

ADR studies: tasks, data and methods

In this section, we summarize three main computational tasks in ADR studies: (i) drug-ADR benchmark data creation, (ii) drug-ADR prediction, and (iii) ADR mechanism analysis. Fig. 6 provides an overview of ADR studies of these three tasks. In each task, we analyze objectives, data, and commonly used methods. Main notations used in the following subsections are described in Table 5.

Table 5. Main notations

Notation	Description
$i \in \{1, \ldots, d\}$	a drug index in a set of given d drugs
$j \in \{1, \ldots, s\}$	an ADR index in a set of given s ADRs
$\mathbf{x}_i \in \mathbb{R}^e$	a descriptor vector of size e of drug i
$\mathbf{X} = [\mathbf{x}_1 \ldots \mathbf{x}_d]^\mathsf{T} \in \mathbb{R}^{d \times e}$	a descriptor matrix of given d drugs, T is the transpose operator.
$y_{i,j} \in \mathbb{R}$	an association score of drug i and ADR j
$\mathbf{y}_i = [y_{i,1} \ldots y_{i,s}]^\mathsf{T} \in \mathbb{R}^s$	a vector for association scores of drug i with s ADRs
$\mathbf{Y} = [\mathbf{y}_1 \ldots \mathbf{y}_d]^\mathsf{T} \in \mathbb{R}^{d \times s}$	a given drug-ADR association score matrix
$\mathbf{h}_i \in \mathbb{R}^m$	a vector of size m representing associated biological components of drug i
$\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_d]^\mathsf{T} \in \mathbb{R}^{d \times m}$	a given drug-biological component matrix

Task 1: drug-ADR benchmark data creation

ADR clinical data contains millions of records with redundant information, for example, some records contain similar information. Creating a drug-ADR benchmark dataset is a necessary task in ADR studies. It helps other studies in evaluating performances of new methods and comparing with existing methods.

Table 6. Contingency table for Fisher's exact test.

		Number of records of drugs	
		drug i	other drugs
ADR j	Yes	n_1	n_3
	No	n_2	n_4

In ADR studies, benchmark data is extracted from clinical records to retrieve reliable drug-ADR associations, which are pairs of drugs with corresponding ADRs. However, drug-ADR pairs have different levels of association significance in clinical records . Some pairs of drug-ADR rarely appear in the clinical records, leading to their low association significance. In addition, some records often contain a combination of more than one drug, making the verification of drug-ADR associations difficult.

To check the significance of drug-ADR associations, association rule mining or statistical significance tests can be applied [58]. We will briefly explain a typical significance test, Fisher's exact test [59]. Consider drug

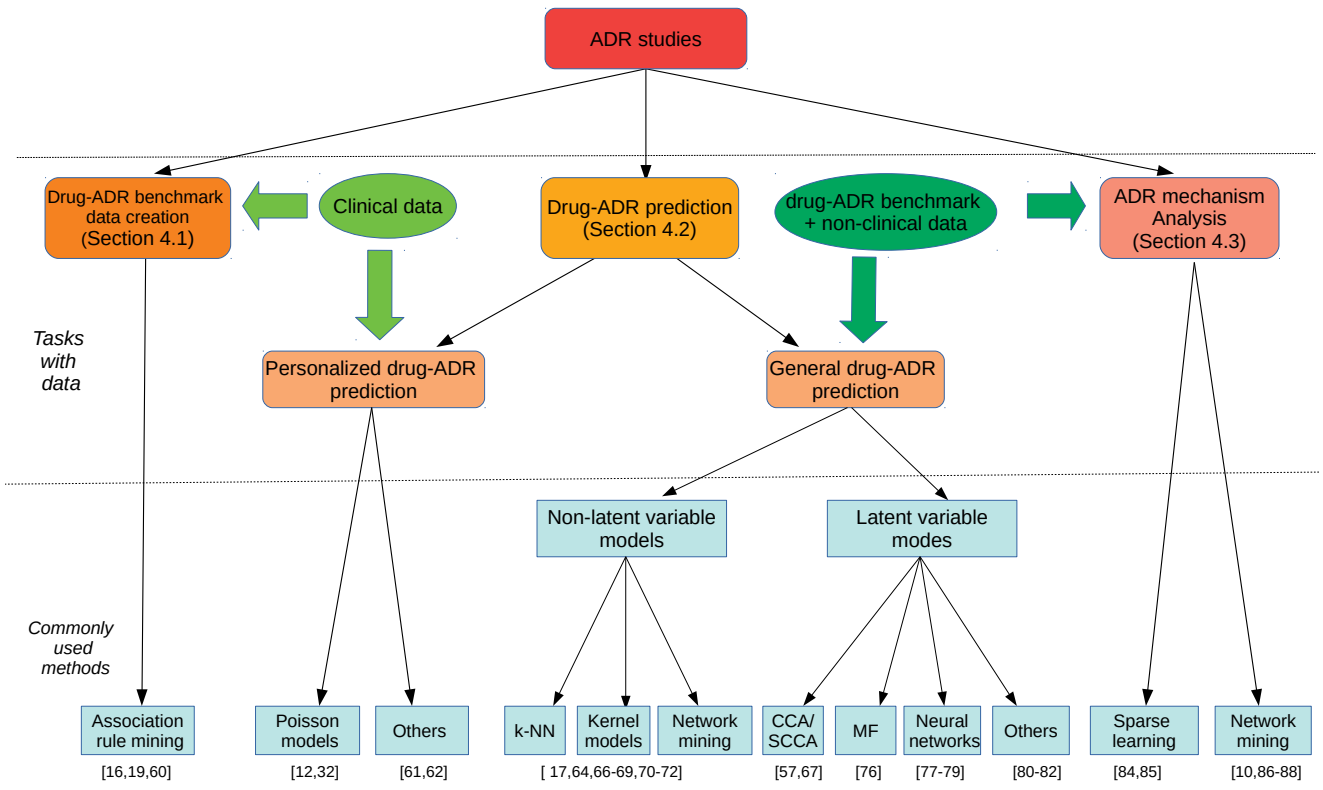

Fig. 6: Computational tasks of ADR studies: Data and commonly used methods.

i and ADR j in a clinical database, the association information of drug i and ADR j is stored into a contingency table as in Table 6. In this table, n_1 denotes the number of records containing ADR j of drug i, while n_2 is that of the other drugs. The number of records that do not contain ADR j of drug i is n_3, and that of the other drugs is n_4. The Fisher's exact test evaluates the significance of the association of drug i and ADR j by a p-value:

$$p = \frac{(n_1+n_2)!(n_3+n_4)!(n_1+n_3)!(n_2+n_4)!}{n_1!n_2!n_3!n_4!(n_1+n_2+n_3+n_4)!}$$

This technique was used on FAERS to extract SIDER, a monopharmacy ADR benchmark dataset used in a large number of ADR studies [16, 60]. The technique was also used to extract OFFSIDES for monopharmacy ADR, which are ADRs of drugs that do not appear in the drug's package insert, and TWOSIDES for polypharmacy ADRs of drug-drug interactions [19].

Task 2: drug-ADR prediction

Predicting ADRs of drugs, or drug-ADR association scores, is a main objective of ADR studies. Depending on the personal context information is used or not, studies in drug-ADR prediction can be divided into two classes: personalized drug-ADR prediction and general drug-ADR prediction. In the following subsections, we analyze machine learning methods according to each class.

Personalized drug-ADR prediction

The personalized drug-ADR prediction uses personal contexts taken from clinical data with information such as dosages of treatments, gender and age of each patient. Therefore, the prediction result will be different among patients even with the same drugs. For this prediction, we focus on methods using Poisson models, which are commonly used models for personalized drug-ADR prediction.

i. Poisson models

The aim of using Poisson models is to predict the probabilities of the numbers of occurrences of ADRs during drug treatments. It is assumed that these numbers follow Poisson distributions with expectations depending on the taken drugs [12, 32]. For simplicity, considering a patient p in a drug treatment, the probabilities of numbers of occurrences of s ADRs $\mathbf{\Phi}(\tilde{\mathbf{y}}|\tilde{\mathbf{x}}) \in \mathbb{R}^s$ are calculated by:

$$\mathbf{\Phi}(\tilde{\mathbf{y}}|\tilde{\mathbf{x}}) = [P(\tilde{y}_1|\phi_1(\tilde{\mathbf{x}})) \dots P(\tilde{y}_j|\phi_j(\tilde{\mathbf{x}})) \dots P(\tilde{y}_s|\phi_s(\tilde{\mathbf{x}}))]^\mathsf{T}$$

where $\tilde{\mathbf{x}} = [\tilde{x}_{p,1} \dots \tilde{x}_{p,i} \dots \tilde{x}_{p,d}]^\mathsf{T}$ is a vector indicating drugs taken by patient p during the treatment, $\tilde{\mathbf{y}} = [\tilde{y}_1 \dots \tilde{y}_j \dots \tilde{y}_s]^\mathsf{T}$ is a vector denoting the numbers of occurrences of s ADRs, and $P(\tilde{y}_j|\phi_j(\tilde{\mathbf{x}}))) = \phi_j(\tilde{\mathbf{x}})^{\tilde{y}_j} e^{-\phi_j(\tilde{\mathbf{x}})}/\tilde{y}_j!$ is the Poisson distribution for the number of occurrences of ADR j with expectation $\phi_j(\tilde{\mathbf{x}})$. A commonly used formulation of ϕ_j is:

$$\phi_j(\tilde{\mathbf{x}}) = \exp\left(\theta_{p,j} + \sum_{i=1}^{d} \tilde{x}_{p,i}.w_{i,j}\right)$$

where $\theta_{p,j}$ is a parameter depending on the patient, leading to differences in ADR occurrences of different patients, and $w_{i,j}$ is a parameter used as a weight for the association of drug i and ADR j [32]. This formulation shows a multiplicative contribution of each drug to the expectation of the number of occurrences of each ADR.

However, the existing Poisson models have a limitation in terms of integrating other information such as weights, genders of patients and also non-clinical data.

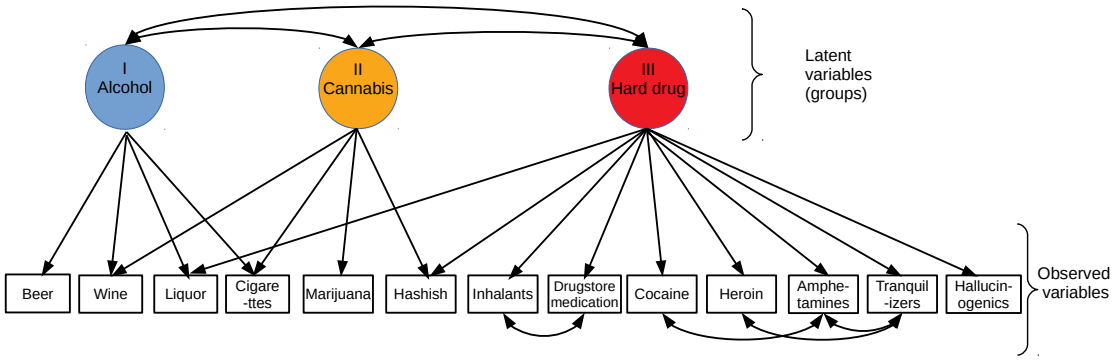

Fig. 7: An example of latent variables with thirteen psychoactive substances [63].

ii. Other methods

There are other methods that were used to combine drugs with personal contexts into medical case vectors. A feature-based similarity method was proposed to learn weights for these medical case vectors with the idea was to distinguish cases having an ADR with cases not having the ADR [61]. These medical case vectors were also used as inputs for a classification problem [62].

General drug-ADR prediction

In contrast to personalized drug-ADR prediction, general drug-ADR prediction predicts drug-ADR association scores without using personal contexts. A common approach for this class is to combine knowledge of drugs from non-clinical data to enrich drug information and apply machine learning methods to build drug-ADR prediction models. As presented in Section , drug information is described by various types of drug descriptors. The drug-ADR prediction models receive the drug descriptors as the inputs and output all corresponding ADRs.

In this study, we consider general drug-ADR prediction as a multi-label classification problem such that each ADR is a label and each drug can have many labels [64, 65]. The prediction models calculate the association scores, which are real numbers, of each drug with all labels. The final labels of the drug are selected from these scores by a ranking method. In detail, a drug-ADR prediction model is formulated as a function $\mathbf{f} : \mathbb{R}^e \rightarrow \mathbb{R}^s$, where e is the number of descriptors and s is the number of ADRs . Given a drug with descriptor vector $\mathbf{x} \in \mathbf{R}^e$, the model predicts drug-ADR association scores with s ADRs: $\mathbf{f}(\mathbf{x}) \in \mathbb{R}^s$.

We further classify the models into two classes: *i. non-latent variable models* and *ii. latent variable models*. Latent variables are ones that are not directly observed or measured, and needed to infer from observed data. Fig. 7 presents an example of latent variables from a study on finding patterns of psychoactive substances used in adolescents [63]. There are thirteen psychoactive substances from beer to hallucinogenics, which are observed variables. In addition, there are some correlated pairs of substance usages, for example, cocaine and amphetamines. The study suggested that these substances can be grouped into three groups: alcohol, cannabis, and hard drug. The patterns of substance usage will be taken from these three groups, which are called latent variables.

A latent variable model is a model that contains latent variables obtained from observed ones. In application to drug-ADR prediction, latent variables of drugs can be interpreted as groups of drug descriptors that are highly correlated with each other. The representations of drugs in the space created by latent variables are called latent vectors.

In the following contents, we first describe the formulation for function $\mathbf{f}$ according to models of the two classes: *non-latent variable models* and *latent variable models*, which are based on the criteria that latent vectors of drugs are learned or not. Then we present an experiment to compare prediction performances of these models.

i. Non-latent variable models

In non-latent variable models, drug descriptors are used to predict drug-ADR associations without learning drug latent vectors. We present three typical methods: *a. k nearest neighbors*, *b. kernel methods* and *c. mining networks of drug-ADR*.

a. k nearest neighbors

The idea of using k nearest neighbors (k-NN) is that drugs having similar descriptor vectors tend to have similar ADRs [17, 64, 66–68]. Suppose that there is a similarity measure $sim : \mathbb{R}^e \times \mathbb{R}^e \rightarrow \mathbb{R}$, for example, cosine similarity. To predict drug-ADR association scores $\mathbf{f}(\mathbf{x})$, first the top k most similar drugs to $\mathbf{x}$ are identified resulting in a set of indices of the similar drugs $T(\mathbf{x}, k)$. Then the drug-ADR association scores are calculated by:

$$\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}) \dots f_j(\mathbf{x}) \dots f_s(\mathbf{x})]^\mathsf{T}.$$

where f_j is a weighted average function:

$$f_j(\mathbf{x}) = \sum_{i \in T(\mathbf{x},k)} w_i(\mathbf{x}) y_{i,j}, \quad j \in \{1 \dots s\},$$

with weights w_i are obtained from drug similarities, for example:

$$w_i(\mathbf{x}) = \frac{sim(\mathbf{x}, \mathbf{x}_i)}{\sum_{i' \in T(\mathbf{x},k)} sim(\mathbf{x}, \mathbf{x}_{i'})} \tag{1}$$

Some extensions of KNN were also applied, for example the linear neighborhood similarity method (LNSM) [69]. In LNSM, the similarity weights are calculated such that a drug descriptor vector is a linear combination of descriptor vectors of the neighbor drugs with corresponding similarity weights.

b. Kernel methods

The idea of using kernel methods, for example, support vector machines (SVMs), is to use classification functions calculated from kernel functions in the form of inner products of drug descriptor vectors [17, 57, 67, 70]. To predict drug-ADR association scores $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}) \dots f_j(\mathbf{x}) \dots f_s(\mathbf{x})]^\mathsf{T}$, the kernel methods use the following form for f_j:

$$f_j(\mathbf{x}) = g(\sum_{i=1}^{d} w_{i,j} y_{i,j} K(\mathbf{x}, \mathbf{x}_i)), \quad j \in \{1 \dots s\}$$

where g is a function, for example, a sign function. $K : \mathbb{R}^e \times \mathbb{R}^e \rightarrow \mathbb{R}$ is a kernel function, for example, a radial basis function (rbf): $K(\mathbf{x}, \mathbf{x}_i) =$

$\exp(-\frac{(\mathbf{x}-\mathbf{x}_i)^\mathsf{T}(\mathbf{x}-\mathbf{x}_i)}{2\delta^2})$ with a hyperparameter δ, and $w_{i,j}$ is a parameter used as a weight for the association of drug i and ADR j.

Different from k-NN, the kernel methods learn weights from a training process, which depends on both drugs and ADRs, while weights in k-NN are calculated only from drug similarities.

c. Mining networks of drug-ADR

Consider a drug-ADR network $G = (V, E)$, where V is a set of nodes of d drug and s ADRs: $V = \{v_1, \ldots, v_i, \ldots, v_d\} \cup \{\nu_1, \ldots, \nu_j, \ldots, \nu_s\}$, and E is a set of edges of drug nodes-ADR nodes for known ADRs of drugs and drug nodes-drug nodes for drug similarities. The idea of mining this network is that if a drug and an ADR in the network are well-connected, they possibly have a high association score [71–73]. This approach can be formulated in two steps:

1. Calculate partial connection scores $r(v_i, \nu_j) \in \mathbb{R}^l$ of each pair of drug node v_i and ADR node ν_j using l different measures on G. A commonly used measure is Jaccard index [71, 73]. Let $N_i = \{v|(v, v_i) \in E\}$ be a set of neighbor nodes of drug node v_i, and $N_j = \{v|(v, \nu_j) \in E\}$ be that of ADR node ν_j, the partial connection score calculated by Jaccard index is: $|N_i \cap N_j|/|N_i \cup N_j|$, where $|\ .\ |$ denotes the cardinality of a set. Some other measures such as Dice index and Adamic/Adar index were also applied [73]. Random walk [74] was also applied to calculate r [75].
2. Calculate drug-ADR association scores $\mathbf{f}(\mathbf{x})$ of a drug with descriptor vector $\mathbf{x}$. Let $v(\mathbf{x})$ be the corresponding node in G of the drug. The association scores are obtained by:

$$\mathbf{f}(\mathbf{x}) = [f(r(v(\mathbf{x}), \nu_1)) \ldots f(r(v(\mathbf{x}), \nu_j)) \ldots f(r(v(\mathbf{x}), \nu_s))]^\mathsf{T}$$

where f was often a binary function [72] or a logistic regression function (LR) [71]. In addition, random forest (RF) was also applied to f [73].

However, a problem with mining drug-ADR networks is sparsity that there are too few edges between drugs and ADRs, for example, in SIDER dataset, the edge density is 0.017. This makes the prediction less effective since there is only a small number of ADRs predicted for each drug.

ii. Latent variable models

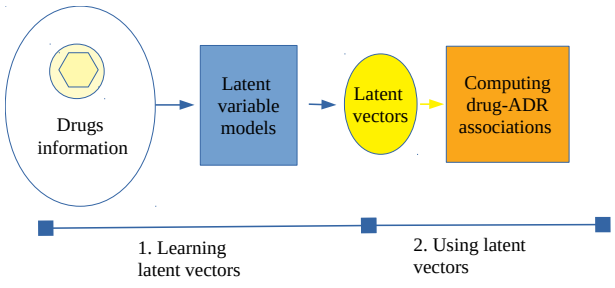

Fig. 8: Learning latent variables and using latent variables.

In latent variable models, drug-ADR association scores are calculated by using drug latent vectors learned from drug descriptors. Fig. 8 illustrates two stages of using latent models: learning latent vectors of drugs and then using these latent vectors for prediction. It is expected that latent vectors can remove redundant information from drug descriptors, for example, unnecessary descriptors. In addition, calculating with latent vectors of small size can reduce complexity of high-dimensional data. In this paper, we briefly describe three commonly used latent variable models (canonical correlation analysis, matrix factorization and neural networks), and some other miscellaneous models.

a. Canonical correlation analysis

The aim of using canonical correlation analysis (CCA) is to find weight vectors $\mathbf{a} \in \mathbb{R}^e$ and $\mathbf{b} \in \mathbb{R}^s$ such that the correlation of the projections of drug descriptor matrix $\mathbf{X}$ and drug-ADR association matrix $\mathbf{Y}$ is maximized [57]:

$$\arg\max_{\mathbf{a},\mathbf{b}} \frac{(\mathbf{Xa})^\mathsf{T}(\mathbf{Yb})}{\sqrt{(\mathbf{Xa})^\mathsf{T}(\mathbf{Xa})}\sqrt{(\mathbf{Yb})^\mathsf{T}(\mathbf{Yb})}}.$$

The first pair of $(\mathbf{Xa}, \mathbf{Yb})$ is called the first pair of canonical variables (latent variables). The remaining pairs of canonical variables have an additional constraint that they are uncorrelated with existing pairs of canonical variables. c pairs of weight vectors $\mathbf{a}$ and $\mathbf{b}$ form two weight matrices: $\mathbf{A} \in \mathbb{R}^{e\times c}$ and $\mathbf{B} \in \mathbb{R}^{s\times c}$, respectively.

The latent vector of a drug with descriptor vector $\mathbf{x}$ is calculated by: $\mathbf{z}(\mathbf{x}) = \mathbf{A}^\mathsf{T}\mathbf{x}$. Drug-ADR association scores $\mathbf{f}(\mathbf{x})$ are obtained by minimizing the distance of latent vectors:

$$\mathbf{f}(\mathbf{x}) = \arg\min_{\mathbf{y}\in\mathbb{R}^s} \left\|\mathbf{z}(\mathbf{x}) - \mathbf{B}^\mathsf{T}\mathbf{y}\right\|.$$

where $\|\ .\ \|$ is a norm, for example, Euclidean norm.

Sparse canonical correlation analysis (SCCA), a variant of CCA, was also applied to predict drug-ADR association scores [67]. In SCCA, L_1 regularization is applied to columns of $\mathbf{A}$ and $\mathbf{B}$, leading to their sparsity.

b. Matrix factorization

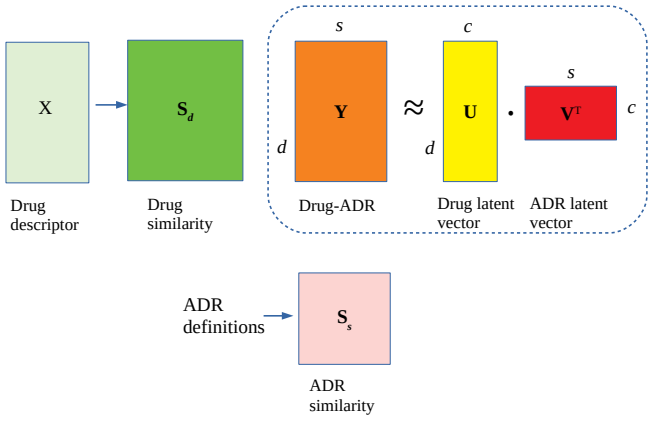

Fig. 9: An illustration of matrix factorization.

The idea of using matrix factorization (MF) is illustrated in Fig.9 [76]. It is assumed that drugs and ADRs share c unknown latent variables. Then the drug-ADR association matrix $\mathbf{Y}$ is decomposed into two matrices of latent vectors of drugs and ADRs in the space of latent variables: $\mathbf{U} \in \mathbb{R}^{d\times c}$ and $\mathbf{V} \in \mathbb{R}^{s\times c}$, such that $\mathbf{Y} \approx \mathbf{U}\mathbf{V}^\mathsf{T}$. Supposing there is a drug similarity matrix $\mathbf{S}_d \in \mathbb{R}^{d\times d}$ calculated from drug descriptors matrix $\mathbf{X}$, and an ADR similarity matrix $\mathbf{S}_s \in \mathbb{R}^{s\times s}$ calculated from ADR definitions, the objective function is:

$$\arg\min_{\mathbf{U},\mathbf{V}} \left\|\mathbf{Y} - \mathbf{U}\mathbf{V}^\mathsf{T}\right\| + \Re(\mathbf{U}, \mathbf{V}, \mathbf{S}_d, \mathbf{S}_s),$$

where the first part is the error from matrix factorization, and the second one is the regularization for $\mathbf{U}$ and $\mathbf{V}$ given $\mathbf{S}_d$ and $\mathbf{S}_s$, for example, Laplacian regularization.

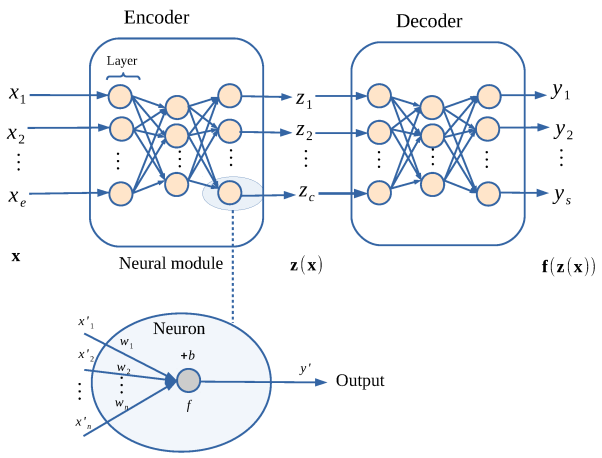

Fig. 10: An illustration of a neural network.

To calculate drug-ADR association scores $\mathbf{f}(\mathbf{x})$, first k-NN is applied to calculate a new latent vector $\mathbf{z}(\mathbf{x})$ from the existing drug latent vectors:

$$\mathbf{z}(\mathbf{x}) = \sum_{i \in T(\mathbf{x},k)} w_i(\mathbf{x})\mathbf{u}_i$$

where $\mathbf{u}_i \in \mathbb{R}^c$ is the latent vector of drug i such that $\mathbf{u}_i^\mathsf{T}$ corresponds to the i^{th} row of $\mathbf{U}$, $T(\mathbf{x}, k)$ is the set of indices of the top k most similar drugs to $\mathbf{x}$, and $w_i(\mathbf{x})$ are similarity weights defined in Equation 1.

Then, the drug-ADR association scores are obtained by:

$$\mathbf{f}(\mathbf{x}) = \mathbf{V}\mathbf{z}(\mathbf{x})$$

Different from CCA, MF only focuses on $\mathbf{Y}$ to learn latent vectors and uses $\mathbf{X}$ as additional information which can be omitted from the regularization part. Meanwhile, CCA requires both $\mathbf{X}$ and $\mathbf{Y}$ to obtain latent vectors.

c. Neural networks

Neural networks, which are machine learning models featured by the ability to learn non-linear relationships, were applied to predict drug-ADR association [77–79]. Fig. 10 illustrates this technique in detail. The basic components of neural networks are *neurons*. Each neuron receives an input vector $\mathbf{x}' = [x'_1\ x'_2 \dots x'_n]^\mathsf{T}$ and outputs a value y' by a function: $y' = f(\mathbf{w}^\mathsf{T}\mathbf{x}' + b)$, where b is a bias, $\mathbf{w} = [w_1\ w_2 \dots w_n]^\mathsf{T}$ is a weight vector, and f is an activation function, for example, a sigmoid function, making non-linear combinations. A neural network module is composed of multiple layers of neurons that the output of each neuron of a layer is used as an input for neurons of other layers. The outputs of a neural module, for example, named *Encoder*, given an input vector $\mathbf{x}$ is denoted by *Encoder*$(\mathbf{x})$.

To predict drug-ADR association scores $\mathbf{f}(\mathbf{x})$, there are two steps to process:

1. Obtain the latent vector: $\mathbf{z}(\mathbf{x}) = $ *Encoder*$(\mathbf{x})$, where *Encoder* is a neural module receiving drug descriptor vector $\mathbf{x}$ as the input vector.
2. Predict drug-ADR association scores: $\mathbf{f}(\mathbf{x}) = $ *Decoder*$(\mathbf{z}(\mathbf{x}))$, where *Decoder* is a neural module receiving drug latent vector $\mathbf{z}(\mathbf{x})$ as the input vector.

An advantage of using neural networks is the ability to approximate any continuous function. If there is no hidden layer, neural networks become logistic regression functions. The architecture of neural networks can be more complex when changing connections of neurons and number of layers, for example, a multi-layer feedforward neural network (MLN) [78], or a deep convolutional neural network (DCN) [79]. These complex neural networks aim to approximate mapping functions from inputs to outputs better. However, the number of parameters in a neural network is often much larger than that of other models. This problem leads to increasing computational complexity and the potential for overfitting of neural networks.

d. Other methods

There are some miscellaneous methods to obtain latent vectors of drugs to predict drug-ADR associations, for example, mapping drugs into an ADR space [80, 81] and mapping drugs into a metabolic reaction space [82]. In mapping drugs into an ADR space, groups of highly correlated ADRs were extracted, then each drug was represented by a vector over these groups. In mapping drugs into a metabolic reaction space, flux variability analysis (FVA) was applied to represent drug-protein/gene interaction profiles by a vector over metabolic reactions [83], then these vectors were used to predict ADRs.

Performance comparison in general drug-ADR prediction

We conducted experiments to compare the general drug-ADR prediction performances on monopharmacy cases of eight machine learning models. There were four non-latent variable models: LNSM [65, 69], SVMs [17, 70], RF [17, 73] and LR[17, 71], and four latent variable models: CCA [57, 67], MF [76], MLN [78], DCN [79] (The convolutional network proposed in [77] addressed polypharmacy ADRs, so we do not compare.).

i.Experimental setups

We ran experiments with AEOLUS dataset [18], a monopharmacy dataset for drug-ADR prediction, which was also used in [65] (AEOLUS is the largest one among AEOLUS, SIDER, and Liu's datasets.). We only selected drugs appearing in DrugBank and ADRs occurring in more than 50 drugs. The final statistical information of the dataset is provided in Table 7, containing the number of drugs, the numbers of ADRs, the numbers of drug-ADR associations, the average, minimum, and maximum numbers of ADRs per each drug.

Table 7. Statistics of the used dataset.

Num drugs	Num ADRs	Num drug-ADRs	ADRs/drug		
			Avg.	Min	Max
1,385	2,707	605,121	445	1	2,703

Table 8. Statistics of the used drug descriptors.

Name	Source	Size
PCBio	Pubchem+Bio2RDF	7, 593
2DChem	PubChem	[N_ATOMS_OF_DRUG , 53]

In the experiments, we used PCBio and Chem2D as two kinds of drug descriptors with information presented in Table 8. PCBio descriptors are the combinations of PubChem descriptors taken from PubChem and chemical, physical, and biological descriptors taken from Bio2RDF. We extracted descriptors with information from DrugBank in Bio2RDF as in [68], and selected descriptors occurring in at least 3 drugs. 2DChem descriptors are drug chemical descriptors represented in the form of a matrix such that each row of the matrix corresponds to chemical features of an atom in a drug. To represent 2DChem descriptors, we extracted 53 chemical properties of each atom in the drug's molecule, hence each drug is represented in the form of a matrix that the number of rows equals to the number of atoms of the drug and the number of columns is 53 (see supplement materials). In our experiments, 2DChem descriptors are only used for DCN model [79], other models use PCBio descriptors.

Two commonly used metrics were selected to evaluate prediction performance: area under the ROC curve (AUC) and area under the precision-recall curve (AUPR) [57, 64, 67, 80].

We used ten-fold cross-validation for the experiment. The hyperparameters of each model were selected by grid searches to obtain the highest prediction performances. In detail, the number of neighbors for LNSM was 60, SVMs were run with a rbf kernel and the soft-margin hyperparameter was 1. RF was run with 80 estimators. CCA had 60 pairs of canonical variables, MF had 60 latent factors, MLF had two hidden layers with the sizes of 1000 and 800. DCN had the same architecture described in [79] with 4 convolutional and pooling layers. The detail of selecting hyperparamters is provided in the supplemental material.

We calculated the average computational time of each fold. The computational time was evaluated on a computer with Intel Core i7-6700 CPU and 16 GB RAM.

ii.Experimental results

Table 9. Performance comparison of drug-ADR prediction models on Aeolus dataset and PCBio descriptors. Results for AUC and AUPR contain mean and standard error values in the format $value \times 10^{-2}$.

	Models						
	Non-latent models				Latent models		
	LNSM	SVMs	RF	LR	CCA	MF	MLN
AUC ($\times 10^{-2}$)	86.07 ±0.56	89.26 ±0.47	86.82 ±0.41	89.00 ±0.40	64.51 ±1.05	87.13 ±0.03	**89.55** ±0.39
AUPR ($\times 10^{-2}$)	59.04 ±1.58	67.57 ±1.63	61.92 ±1.11	66.75 ±1.08	34.17 ±2.07	61.03 ±1.13	**68.70** ±1.23
Time (s)	73	22642	181	3658	317	**25**	186

The results of prediction performances and computational time are presented in Table 9. In addition, DCN with 2DChem descriptors achieved 73.80 ± 0.46 in AUC, 39.10 ± 0.63 in AUPR, and 4862(s) of computational time.

The results show that MLN is the model having the highest prediction performances in both AUC and AUPR (89.55×10^{-2} and 68.70×10^{-2}). SVMs are the second highest model with 89.26×10^{-2} and 67.57×10^{-2} for AUC and AUPR. In terms of computational time, MF is the fastest model, and SVMs are the slowest one. CCA and DCN are the two models having the lowest prediction performances.

We summarize the properties of the models in terms of linearity and dimensional reduction, and rank the performances of the models in AUC and computational time as in Table 10. This table shows that in balancing between prediction accuracy and computational time, two latent variable models, MLN and MF, are the two most promising ones. In addition, latent variable models learn latent representation vectors of small size for drugs, which are much smaller than the original size of the drug descriptor vectors. This dimensional reduction can help to remove redundant information from drug descriptors. We also can see that three out of the four highest AUC models are non-linear, suggesting that there are non-linear relationships between drug descriptors and ADRs.

Table 10. Summary of the models in terms of performance, non-linearity, and dimensional reduction.

Models		AUC ranking	Time ranking	Non-linearity	Dimensional reduction
Non-latent	LNSM	6	2		
	SVMs	2	8	✓	
	RF	5	3	✓	
	LR	3	6	✓	
Latent	CCA	7	4		✓
	MF	4	1		✓
	MLN	1	4	✓	✓
	DCN	8	7	✓	✓

Task 3: ADR mechanism analysis

The objective of this task is to reveal associated biological components such as proteins or pathways of ADRs. In this task, non-clinical data of drug-protein interactions, protein-pathways, chemical-pathways is combined with clinical data, usually drug-ADR benchmark data. There are two commonly used approaches for this task: *i. using sparse learning* and *ii. using network mining*.

i. Using sparse learning

In the sparse learning approach, the idea is to consider associated biological components of each drug as a feature vector, and then find associated features corresponding to ADRs. To do this, weight vectors over biological components and ADR are used with sparse constraints by applying $L1$ regularization. Remaining subsets with high weights of biological components and ADRs are associated with each other. We describe two studies using this approach with logistic regression and canonical correlation analysis.

Logistic regression with regularization was proposed to obtain associated biological pathways with each ADR [84]. To obtain pathways associated with ADR j, let $\mathbf{w}_j \in \mathbb{R}^m$ be weights over m pathways obtaining from:

$$\arg\min_{\mathbf{w}_j} \frac{1}{d} \sum_{i=1}^{d} \big(- y_{i,j} \log \frac{1}{1+\exp(-\mathbf{h}_i \cdot \mathbf{w}_j)} - (1-y_{i,j}) \log(1 - \frac{1}{1+\exp(-\mathbf{h}_i \cdot \mathbf{w}_j)}) \big) + \lambda_w \left\| \mathbf{w}_j \right\|_1 .$$

where λ_w is a regularization parameter.

$L1$ regularization $\left\| \mathbf{w}_j \right\|_1$ forces $\mathbf{w}_j$ to be a sparse vector. The corresponding pathways with high weights are associated with ADR j.

SCCA was applied to obtains subsets of correlation of ADRs and pathways [85]. By applying SCCA into two matrices $\mathbf{Y}$ and $\mathbf{H}$ of drug-ADR and drug-biological component, respectively, two sparse weight matrices $\mathbf{A} \in \mathbb{R}^{s \times c}$ and $\mathbf{B} \in \mathbb{R}^{m \times c}$ are obtained. The corresponding subsets of ADRs and pathways of each pair of $(\mathbf{a}_l, \mathbf{b}_l)$ with $l = 1 \ldots c$ are correlated.

ii. Using network mining

The idea of using networks of ADR-biological components is similar to mining drug-ADR networks for drug-ADR prediction. If a biological component and an ADR are well-connected in a network of biological component-ADR, they are highly associated with each other. The technique was used in [10, 86, 87] to build a protein-ADR network and discover associated proteins with each ADR. Dijkstra algorithm, a well known method to calculate the shortest paths in a graph, was used on the network of biological components-ADR to obtain associated biological pathways of ADRs [88].

Discussion

This survey addresses ADR-related studies in three aspects: data, drug descriptors, and tasks with corresponding methods. We divide data sources into clinical and non-clinical data. Clinical data contains important personal context information such as ADRs, diseases, dosages of treatments, and demographic information. Non-clinical data contains more detailed information of drugs and biological systems with chemical, physical properties of drugs, drug-protein interactions, and biological pathways.

We summarize the commonly used drug descriptors in ADR studies. In addition to traditional physical and chemical descriptors, many studies integrate biological descriptors of drugs to have better drug information.

There are three main tasks in ADR studies: creating drug-ADR benchmark data, drug-ADR prediction, and ADR mechanism analysis. Association rule mining is the commonly used method for creating drug-ADR benchmark data. The drug-ADR prediction task is classified into two classes: personalized drug-ADR prediction and general drug-ADR prediction. In the former class, Poisson models are widely used. In the latter class, the commonly used machine learning models can be categorized into non-latent variable models and latent variable models. The non-latent variable models predict drug-ADR without learning latent variables, while the latent variable models learn latent vectors of small size to represent drugs such that these latent vectors can help the prediction efficiently. The experimental results show that MLN is the model having the highest prediction performances, and the latent variable models have the potential for further development. In ADR mechanism analysis, using sparse learning and network mining are two commonly used approaches.

From this survey, we have three remarks on problems in current ADR studies as follows in current ADR studies as follows:

1) Most of drug-ADR prediction studies address monopharmacy cases in SIDER benchmark data. There are few studies proposed models for polypharmacy prediction, for example, predicting with TWOSIDES benchmark data [77], in spite of the fact that most of significant ADRs come from drug combinations [19, 77].

2) ADR data sources are not effectively used. Recent ADR studies only use either clinical data without non-clinical data information or use ADR benchmark data and non-clinical data without personal context information. There are no studies that combine full clinical data with non-clinical data. In addition, current ADR benchmark data such as SIDER, OFFSIDES and TWOSIDES only contains drugs and ADRs, other personal context information still remains in original clinical records.

3) Machine learning models are mostly used as black boxes for drug-ADR prediction, since they only output association scores of drugs and ADRs. In ADR discovery, explaining ADR mechanisms is a big challenge. It is not only a problem of predicting corresponding ADRs of drugs but also how ADRs occur. However, predicting and revealing ADR mechanisms are now considered as two separate parts. Designing drug-ADR prediction models which reveal related information of ADR mechanisms seems to be an important topic.

In conclusion, the use of machine learning models in ADR studies is likely to develop in the future. Effectively using available data with suitable models still remains a big challenge. It is not only drug-ADR prediction is an important task, but also revealing ADR mechanisms is another task to concentrate on.

Key points

- Machine learning methods are prominent tools for ADR studies.
- There are three main tasks in ADR studies: creating ADR benchmark data, drug-ADR prediction, and ADR mechanism analysis.
- For drug-ADR prediction, latent variables models have the potential for further development.
- Remaining issues of ADR studies: 1) There are very few drug-ADR prediction models addressing polypharmacy ADR, 2) ADR data sources are not effectively used, and 3) Drug-ADR prediction models lack the ability to explain ADR mechanisms.

Funding

D.A.N. has been supported in part by Otsuka Toshimi Scholarship Foundation. C.H.N. has been supported in part by MEXT Kakenhi 18K11434. H.M. has been supported in part by JST ACCEL [grant number JPMJAC1503], MEXT Kakenhi [grant numbers 16H02868 and 19H04169], FiDiPro by Tekes (currently Business Finland) and AIPSE by Academy of Finland.

References

[1]World Health Organization: WHO (1972). Definitions. `http://www.who.int/medicines/areas/quality_safety/safety_efficacy/trainingcourses/definitions.pdf` (10 Sep 2019, date last accessed).

[2]Poudel, D. R., Acharya, P., Ghimire, S., *et al.* (2017). Burden of hospitalizations related to adverse drug events in the usa: a retrospective analysis from large inpatient database. *Pharmacoepidemiology and drug safety*, **26**(6), 635–641.

[3]Weiss, A., Elixhauser, A., Bae, J., *et al.* (2013). Origin of adverse drug events in us hospitals, 2011. *HCUP Statistical Brief*, **158**.

[4]Hoots, B. E., Xu, L., Kariisa, M., *et al.* (2018). 2018 annual surveillance report of drug-related risks and outcomes–united states. *CDC National Center for Injury Prevention and Control*.

[5]Rieder, M. J. (1994). Mechanisms of unpredictable adverse drug reactions. *Drug Safety*, **11**(3), 196–212.

[6]Mann, R. D. and Andrews, E. B. (2007). *Pharmacovigilance*. John Wiley & Sons.

[7]Poloju, N. and Muniganti, P. (2018). Adverse drug reaction detection using data mining approaches: A survey. *Internatinal journal of recent trends in engineering and research*, **2018**.

[8]Chen, Y. G., Wang, Y. Y., and Zhao, X. M. (2016). A survey on computational approaches to predicting adverse drug reactions. *Current topics in medicinal chemistry*, **16**(30), 3629–3635.

[9]Ho, T. B., Le, L., Thai, D. T., *et al.* (2016). Data-driven approach to detect and predict adverse drug reactions. *Current pharmaceutical design*, **22**(23), 3498–3526.

[10]Wang, X., Thijssen, B., and Yu, H. (2013). Target essentiality and centrality characterize drug side effects. *PLoS computational biology*, **9**(7), e1003119.

[11]Alberti, P. and Cavaletti, G. (2014). Management of side effects in the personalized medicine era: chemotherapy-induced peripheral neuropathy. In *Pharmacogenomics in Drug Discovery and Development*, pages 301–322. Springer.

[12]Bao, Y., Kuang, Z., Peissig, P., *et al.* (2017). Hawkes process modeling of adverse drug reactions with longitudinal observational data. In *Machine learning for healthcare conference*, volume 2017, pages 177–190. Proceedings of Machine Learning Research.

[13]U.S. Food and Drug Administration (2019). Questions and answers on fda's adverse event reporting system (faers). `https://www.fda.gov/drugs/surveillance/fda-adverse-event-reporting-system-faers`, 10 Sep 2019, date last accessed.

[14]Stang, P. E., Ryan, P. B., Racoosin, J. A., *et al.* (2010). Advancing the science for active surveillance: rationale and design for the observational medical outcomes partnership. *Annals of internal medicine*, **153**(9), 600–606.

[15]Hripcsak, G., Duke, J. D., Shah, N. H., *et al.* (2015). Observational health data sciences and informatics (ohdsi): opportunities for observational researchers. *Studies in health technology and informatics*, **216**, 574.

[16]Kuhn, M., Letunic, I., Jensen, L. J., *et al.* (2015). The sider database of drugs and side effects. *Nucleic acids research*, **44**(D1), D1075–D1079.

[17]Liu, M., Wu, Y., Chen, Y., *et al.* (2012). Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. *Journal of the American Medical Informatics Association*, **19**(e1), e28–e35.

[18]Banda, J. M., Evans, L., Vanguri, R. S., *et al.* (2016). A curated and standardized adverse drug event resource to accelerate drug safety research. *Scientific data*, **3**, 160026.

[19]Tatonetti, N. P., Patrick, P. Y., Daneshjou, R., and Altman, R. B. (2012). Data-driven prediction of drug effects and interactions. *Science translational medicine*, **4**(125), 125ra31–125ra31.

[20]Wishart, D. S., Knox, C., Guo, A. C., *et al.* (2007). Drugbank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic acids research*, **36**(suppl_1), D901–D906.

[21]Kim, S., Thiessen, P. A., Bolton, E. E., *et al.* (2015). Pubchem substance and compound databases. *Nucleic acids research*, **44**(D1), D1202–D1213.

[22]Berman, H. M., Westbrook, J., Feng, Z., *et al.* (2000). The protein data bank. *Nucleic acids research*, **28**(1), 235–242.

[23]Liu, T., Lin, Y., Wen, X., *et al.* (2006). Bindingdb: a web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic acids research*, **35**(suppl_1), D198–D201.

[24]Keshava Prasad, T., Goel, R., Kandasamy, K., *et al.* (2008). Human protein reference databaseâŁ"2009 update. *Nucleic acids research*, **37**(suppl_1), D767–D772.

[25]Davis, A. P., Murphy, C. G., Saraceni-Richards, C. A., *et al.* (2008). Comparative toxicogenomics database: a knowledgebase and discovery tool for chemical–gene–disease networks. *Nucleic acids research*, **37**(suppl_1), D786–D792.

[26]Kanehisa, M. and Goto, S. (2000). Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, **28**(1), 27–30.

[27]Günther, S., Kuhn, M., Dunkel, M., *et al.* (2007). Supertarget and matador: resources for exploring drug-target relationships. *Nucleic acids research*, **36**(suppl_1), D919–D922.

[28]Cai, M.-C., Xu, Q., Pan, Y.-J., *et al.* (2014). Adrecs: an ontology database for aiding standardization and hierarchical classification of adverse drug reaction terms. *Nucleic acids research*, **43**(D1), D907–D913.

[29]Ji, Z. L., Han, L. Y., Yap, C. W., *et al.* (2003). Drug adverse reaction target database (dart). *Drug safety*, **26**(10), 685–690.

[30]Chen, X., Ji, Z. L., and Chen, Y. Z. (2002). Ttd: therapeutic target database. *Nucleic acids research*, **30**(1), 412–415.

[31]Belleau, F., Nolin, M.-A., Tourigny, N., *et al.* (2008). Bio2rdf: towards a mashup to build bioinformatics knowledge systems. *Journal of biomedical informatics*, **41**(5), 706–716.

[32]Simpson, S. E., Madigan, D., Zorych, I., *et al.* (2013). Multiple self-controlled case series for large-scale longitudinal observational databases. *Biometrics*, **69**(4), 893–902.

[33]Huynh, T., He, Y., Willis, A., *et al.* (2016). Adverse drug reaction classification with deep neural networks. In *Proceedings of COLING*. Coling, COLING.

[34]Lee, K., Qadir, A., Hasan, S. A., *et al.* (2017). Adverse drug event detection in tweets with semi-supervised convolutional neural networks. In *Proceedings of the 26th International Conference on World Wide Web*, pages 705–714. International World Wide Web Conferences Steering Committee, WWW.

[35]Emadzadeh, E., Sarker, A., Nikfarjam, A., and Gonzalez, G. (2017). Hybrid semantic analysis for mapping adverse drug reaction mentions in tweets to medical terminology. In *AMIA Annual Symposium Proceedings*, volume 2017, page 679. American Medical Informatics Association, PubMed Central.

[36]Ring, J. and Brockow, K. (2002). Adverse drug reactions: mechanisms and assessment. *European surgical research*, **34**(1-2), 170–175.

[37]Testa, B. and Kier, L. B. (1991). The concept of molecular structure in structure–activity relationship studies and drug design. *Medicinal research reviews*, **11**(1), 35–48.

[38]Todeschini, R. and Consonni, V. (2008). *Handbook of molecular descriptors*, volume 11. John Wiley & Sons.

[39]Grisoni, F., Ballabio, D., Todeschini, R., *et al.* (2018). Molecular descriptors for structure–activity applications: A hands-on approach. In *Computational Toxicology*, pages 3–53. Springer.

[40]Steinbeck, C., Han, Y., Kuhn, S., *et al.* (2003). The chemistry development kit (cdk): An open-source java library for chemo-and bioinformatics. *Journal of chemical information and computer sciences*, **43**(2), 493–500.

[41]Daylight Chemica Information System, Inc (2018). Daylight theory: Fingerprint. `http://www.daylight.com/dayhtml/doc/theory/theory.finger.html` (10 Sep 2019, date last accessed).

[42]Hall, L. H. and Kier, L. B. (1995). Electrotopological state indices for atom types: a novel combination of electronic, topological, and valence state information. *Journal of Chemical Information and Computer Sciences*, **35**(6), 1039–1045.

[43]Klekota, J. and Roth, F. P. (2008). Chemical substructures that enrich for biological activity. *Bioinformatics*, **24**(21), 2518–2525.

[44]Durant, J. L., Leland, B. A., Henry, D. R., and Nourse, J. G. (2002). Reoptimization of mdl keys for use in drug discovery. *Journal of chemical information and computer sciences*, **42**(6), 1273–1280.

[45]Bolton, E. E., Kim, S., and Bryant, S. H. (2011). Pubchem3d: conformer generation. *Journal of cheminformatics*, **3**(1), 4.

[46]Openeye scientific software (2018). OMEGA. `https://www.eyesopen.com/omega` (10 Nov 2018, date last accessed).

[47]Wermuth, C. G. (2011). *The practice of medicinal chemistry*. Academic Press.

[48]Goodford, P. J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *Journal of medicinal chemistry*, **28**(7), 849–857.

[49]Crivori, P., Cruciani, G., Carrupt, P. A., *et al.* (2000). Predicting blood-brain barrier permeation from three-dimensional molecular structure. *Journal of medicinal chemistry*, **43**(11), 2204–2216.

[50]Kubinyi, H. (1998). Comparative molecular field analysis (comfa). *The encyclopedia of computational chemistry*, **1**, 448–460.

[51]Cruciani, G., Carosati, E., De Boeck, B., *et al.* (2005). Metasite: understanding metabolism in human cytochromes from the perspective of the chemist. *Journal of medicinal chemistry*, **48**(22), 6970–6979.

[52]Young, R. C., Mitchell, R. C., Brown, T. H., *et al.* (1988). Development of a new physicochemical model for brain penetration and its application to the design of centrally acting h2 receptor histamine antagonists. *Journal of medicinal chemistry*, **31**(3), 656–671.

[53]Testa, B., Caron, G., Crivori, P., *et al.* (2000). Lipophilicity and related molecular properties as determinants of pharmacokinetic behaviour. *CHIMIA International Journal for Chemistry*, **54**(11), 672–677.

[54]van de Waterbeemd, H. and Kansy, M. (1992). Hydrogen-bonding capacity and brain penetration. *CHIMIA International Journal for Chemistry*, **46**(7-8), 299–303.

[55]Lipinski, C. A., Lombardo, F., Dominy, B. W., *et al.* (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, **23**(1-3), 3–25.

[56]WHO Collaborating Centre for Drug Statistics Methodology (2019). Guidelines for atc classification and ddd assignment. `https://www.whocc.no/filearchive/publications/2019_guidelines_web.pdf`, (10 Sep 2019, date last accessed).

[57]Yamanishi, Y., Pauwels, E., and Kotera, M. (2012). Drug side-effect prediction based on the integration of chemical and biological spaces. *Journal of chemical information and modeling*, **52**(12), 3284–3292.

[58]Montastruc, J.-L., Sommet, A., Bagheri, H., and Lapeyre-Mestre, M. (2011). Benefits and strengths of the disproportionality analysis for identification of adverse drug reactions in a pharmacovigilance database. *British journal of clinical pharmacology*, **72**(6), 905–908.

[59]Agresti, A. (1992). A survey of exact inference for contingency tables. *Statistical science*, **7**(1), 131–153.

[60]Kuhn, M., Campillos, M., Letunic, I., *et al.* (2010). A side effect resource to capture phenotypic effects of drugs. *Molecular systems biology*, **6**(1), 343.

[61]Yang, F., Yu, X., and Karypis, G. (2014). Signaling adverse drug reactions with novel feature-based similarity model. In *Bioinformatics and Biomedicine (BIBM), 2014 IEEE International Conference on*, pages 593–596. IEEE, IEEE.

[62]Chen, A. W. (2018). Predicting adverse drug reaction outcomes with machine learning. *International Journal Of Community Medicine And Public Health*, **5**(3), 901–904.

[63]Huba, G., Wingard, J. A., and Bentler, P. M. (1981). A comparison of two latent variable causal models for adolescent drug use. *Journal of Personality and Social Psychology*, **40**(1), 180.

[64]Zhang, W., Liu, F., Luo, L., *et al.* (2015). Predicting drug side effects by multi-label learning and ensemble learning. *BMC bioinformatics*, **16**(1), 365.

[65]Muñoz, E., Nováček, V., and Vandenbussche, P.-Y. (2017). Facilitating prediction of adverse drug reactions by using knowledge graphs and multi-label learning models. *Briefings in bioinformatics*, **20**(1), 190–202.

[66]Cao, D., Xiao, N., Li, Y., *et al.* (2015). Integrating multiple evidence sources to predict adverse drug reactions based on a systems pharmacology model. *CPT: pharmacometrics & systems pharmacology*, **4**(9), 498–506.

[67]Pauwels, E., Stoven, V., and Yamanishi, Y. (2011). Predicting drug side-effect profiles: a chemical fragment-based approach. *BMC bioinformatics*, **12**(1), 169.

[68]Muñoz, E., Nováček, V., and Vandenbussche, P.-Y. (2016). Using drug similarities for discovery of possible adverse reactions. In *AMIA Annual Symposium Proceedings*, volume 2016, page 924. American Medical Informatics Association.

[69]Zhang, W., Chen, Y., Tu, S., *et al.* (2016). Drug side effect prediction through linear neighborhoods and multiple data source integration. In *2016 IEEE international conference on bioinformatics and biomedicine (BIBM)*, pages 427–434. IEEE.

[70]Jahid, M. J. and Ruan, J. (2013). An ensemble approach for drug side effect prediction. In *Bioinformatics and Biomedicine (BIBM), 2013 IEEE International Conference on*, volume 2013, pages 440–445. IEEE, IEEE.

[71]Cami, A., Arnold, A., Manzi, S., and Reis, B. (2011). Predicting adverse drug events using pharmacological network models. *Science translational medicine*, **3**(114), 114ra127–114ra127.

[72]Lin, J., Kuang, Q., Li, Y., and et al. (2013). Prediction of adverse drug reactions by a network based external link prediction method. *Analytical Methods*, **5**(21), 6120–6127.

[73]Davazdahemami, B. and Delen, D. (2018). A chronological pharmacovigilance network analytics approach for predicting adverse drug events. *Journal of the American Medical Informatics Association*, **25**(10), 1311–1321.

[74]Tong, H., Faloutsos, C., and Pan, J.-Y. (2006). Fast random walk with restart and its applications. In *Sixth International Conference on Data Mining (ICDM'06)*, pages 613–622. IEEE, IEEE.

[75]Rahmani, H., Weiss, G., Méndez-Lucio, O., *et al.* (2016). Arwar: A network approach for predicting adverse drug reactions. *Computers in biology and medicine*, **68**, 101–108.

[76]Poleksic, A. and Xie, L. (2018). Predicting serious rare adverse reactions of novel chemicals. *Bioinformatics*, **1**, 8.

[77]Zitnik, M., Agrawal, M., and Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, **34**(13), i457–i466.

[78]Wang, C.-S., Lin, P.-J., Cheng, C.-L., *et al.* (2019). Detecting potential adverse drug reactions using a deep neural network model. *Journal of medical Internet research*, **21**(2), e11016.

[79]Dey, S., Luo, H., Fokoue, A., *et al.* (2018). Predicting adverse drug reactions through interpretable deep learning framework. *BMC bioinformatics*, **19**(21), 476.

[80]Dimitri, G. M. and Lió, P. (2017). Drugclust: a machine learning approach for drugs side effects prediction. *Computational biology and chemistry*, **68**, 204–210.

[81]Xiao, C., Zhang, P., Chaowalitwongse, W. A., *et al.* (2017). Adverse drug reaction prediction with symbolic latent dirichlet allocation. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. AAAI Press.

[82]Shaked, I., Oberhardt, M. A., Atias, N., *et al.* (2016). Metabolic network prediction of drug side effects. *Cell systems*, **2**(3), 209–213.

[83]Mahadevan, R. and Schilling, C. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic engineering*, **5**(4), 264–276.

[84]Wallach, I., Jaitly, N., and Lilien, R. (2010). A structure-based approach for mapping adverse drug reactions to the perturbation of underlying biological pathways. *PloS one*, **5**(8), e12063.

[85]Zheng, H., Wang, H., Xu, H., *et al.* (2014). Linking biochemical pathways and networks to adverse drug reactions. *IEEE transactions on nanobioscience*, **13**(2), 131–137.

[86]Chen, X., Liu, X., Jia, X., *et al.* (2013). Network characteristic analysis of adr-related proteins and identification of adr-adr associations. *Scientific reports*, **3**, 1744.

[87]Jiang, Y., Li, Y., Kuang, Q., *et al.* (2014). Predicting putative adverse drug reaction related proteins based on network topological properties. *Analytical Methods*, **6**(8), 2692–2698.

[88]Wang, J., Li, Z.-x., Qiu, C.-x., *et al.* (2011). The relationship between rational drug design and drug side effects. *Briefings in bioinformatics*, **13**(3), 377–382.