Teaser Machine learning, especially deep learning, has the predictive power to predict adverse drug reactions, repurpose drugs and perform precision medicine. We provide a background of machine learning and propose a potential high-performance deep learning framework for its successful applications in these practices.

Machine learning on adverse drug reactions for pharmacovigilance

Chun Yen Lee and Yi-Ping Phoebe Chen

College of Science, Health and Engineering, La Trobe University, Melbourne, Australia

Introduction

An adverse drug reaction (ADR) is an unintended response to a drug that is noxious and the reaction has a causal relationship to the drug [1–5]. It is disturbing that ~3.6% of all hospital admissions are caused by ADRs [1–5]. Around 27% of ADR cases reoccur within 6 months of the initial event owing to unintentional re-dispensing or re-prescribing of the drugs. For hospitalized patients, 16.88% experienced ADRs with 6.7% having serious ADRs, with 0.32% being fatal [1–5]. The emergence of 'big healthcare data' characterized by massive volume, complexity and velocity has presented an intriguing opportunity for the study of digital pharmacovigilance. Social media, in addition to other conventional data platforms, has thus become an enabling source for the detection and prediction of ADRs to improve pharmacovigilance. Many existing research studies use different methodologies to detect ADRs through assessing the association between a drug and its ADRs. However, in reality, the occurrence of ADRs can be related to many causal factors, thus it is essential to identify the multifactors that cause ADRs.

There are many factors that contribute to the occurrence of ADRs. Of these, medication errors (MEs) caused by missing doses, performing the wrong administration techniques, equipment failure, inadequate monitoring and preparation errors are some of the contributing factors [22]. Approximately one-third of MEs resulted in ADRs [1,2,7–9]. Polypharmacy, which is the prescription of multiple medications, is also associated with increased risk of ADRs. Polypharmacy and inappropriate medication have been shown to contribute substantially to the burden of morbidity, hospitalization and death. Studies show that up to 50% of ADR-related hospitalizations were preventable by avoiding inappropriate prescribing [1,3,5].

Poor management and monitoring of ADRs has resulted in unnecessary hospitalization, morbidity and mortality [32]. Patients with ADRs are susceptible to many hospitalization-related complications such as cardiovascular and neurological disorders, nosocomial infections and deconditioning [10]. Although all drugs undergo extensive screening before they are approved by the FDA, there are still many occurrences of ADRs due to small clinical trial samples used during premarketing research and, on many occasions, patients with comorbid diseases are omitted from the trials. Because premarketing trials cannot fully cover diverse populations, post-marketing surveillance to identify and predict ADRs is necessary.

In this review, we present a brief overview of current data sources available for post-marketing surveillance, discuss the current computational methodologies applied for pharmacovigilance at the post-marketing stage, examine the causation of ADRs and make reference to the approaches used at the premarketing stage to arrive at a method to predict ADRs at the

Corresponding author: Chen, Y.-P. (phoebe.chen@latrobe.edu.au)

Chun Yen Lee received a Bachelor's degree in pharmacy from the University of Queensland, Brisbane, QLD, Australia, in 2014. She is currently pursuing her Master's degree at La Trobe University, Melbourne, Australia. She has been undertaking research on prediction of ADRs and DDIs in pharmacovigilance. Her current research interests include application of artificial intelligence in pharmacovigilance.

Yi-Ping Phoebe Chen is Professor and Chair at the Department of Computer Science and Information Technology, La Trobe University, Melbourne, Australia, since April 2010. Prof. Chen is the Chief Investigator of ARC Centre of Excellence in Bioinformatics. Phoebe received her BInfTech degree with First Class Honors and her PhD in computer science (bioinformatics) from the University of Queensland. Professor Chen has been doing multidiscipline research for >18 years. Her research focus is to find effective solutions for visualizing, integrating, analyzing and mining big data, complex structures and functions for scientific and biomedical applications. She has been working in many emerging areas such as bioinformatics, multimedia, scientific visualization, pattern recognition, health informatics, data mining, deep learning and databases. She has published >220 research papers, many of them appeared in top journals and at conferences such as *IEEE Transactions on Neural Networks and Learning Systems*, *IEEE Transactions on Biomedical Engineering*, *Pattern Recognition*, *IEEE Transactions on Information Technology in Biomedicine*, *Molecular Systems Biology*, *Aging Cell*, *Drug Discovery Today*, *Nucleic Acids Research*, *Journal of Medical Genetics*, *RNA*, *BMC Genomics*, *BMC Bioinformatics*, *Current Drug Metabolism*, *Data Mining and Knowledge Discovery*, *IEEE Transactions on Knowledge and Data Engineering*, *Information Systems*, *SIGMOD*, among others. She is the steering committee chair of the Asia-Pacific Bioinformatics Conference (founder) and International conference on Multimedia Modeling. She has been on the program committees of >100 international conferences, including top-ranking conferences such as ICDE, ICPR, ISMB, CIKM, among others. Phoebe is a senior member of IEEE and a senior member of ACM.

https://doi.org/10.1016/j.drudis.2019.03.003

post-marketing stage. With the goal of improving drug safety, we examine the latest methods for the prediction of ADRs after post-marketing drug surveillance. The systemic literature review shows that traditional post-marketing drug surveillance methods have been replaced by machine learning for the prediction of ADRs because machine learning methods are better suited for larger datasets. An examination of the drug discovery methods revealed that deep learning has recently been applied for drug discovery. In 2017, Zhang *et al.* [57] found that advances in computer hardware make very big data modeling possible and, when compared with machine learning approaches which have been applied to drug discovery for decades, deep learning methods are more powerful and efficient in dealing with massive amounts of complex tasks based on large, heterogeneous and high-dimensional datasets without the need for human input, which is a promising new technique for modern drug discovery. In light of the potential of deep learning for drug discovery in the new big data era, we anticipate the possibility of using deep learning for prediction of ADRs.

Data sources for post-marketing surveillance

Diverse data sources have been used by many researchers for post-marketing drug safety surveillance. Reliable data sources are important for researchers, healthcare providers and the broader community to obtain an insight into the safety profiles of drugs. There are two types of data sources: structured and unstructured, for post-marketing drug safety surveillance. The spontaneous adverse event reports collected by the health authorities under the voluntary reporting systems are the major sources for structured data. Some of the prominent spontaneous reporting systems (SRSs) include the adverse event reporting system (AERS) maintained by the FDA and VigiBase, which is managed by WHO [3]. The FDA has a post-marketing drug surveillance program called MedWatch to monitor the effects of drugs [2]. As soon as drugs are released for use, MedWatch allows the spontaneous reporting of ADRs by healthcare professionals and patients. All the reported adverse events are recorded as a part of the FDA AERS (FAERS) and are constantly monitored for statistically significant adverse drug event (ADE) reports [3]. However, spontaneous reports are often incomplete and inaccurate owing to voluntary effort [23,50,51]. It is regarded as being biased because many healthcare professionals tend to under-report ADR cases.

Banda *et al.* [3] provided a curated and standardized version of FAERS that removed duplicate case records. They applied standardized vocabularies with drug names mapped to RxNorm concepts and mapped outcomes to SNOMED-CT concepts. They even pre-computed a statistics summary on drug outcome relationships to enable standardized analyses to be performed using common vocabularies. One advantage of this resource over others is that it provides all source codes and this allows researchers to periodically refresh the data to keep themselves updated on the latest datasets released by the FDA. It was report that the availability of this public resource, along with source code, reduced the amount of time spent on performing data management on the source FAERS reports and eased the detection of ADRs [36]. Table 1 details the data sources, such as SRS, electronic health records (EHRs), scientific literature, social media and clinical narratives, for detecting ADRs.

In the past decade, pharmacovigilance has evolved from passive surveillance to active surveillance. In addition to FAERS, EHRs have also become a useful type of structured data sources [48]. The coded EHR data provide active and real-time post-marketing drug safety surveillance [34]. To minimize the occurrence of ADRs, active medication safety surveillance systems, which include the retrospective method and prospective method, are used in hospitals to prevent ADRs. Retrospective methods, which are conducted after the ADEs, depend on voluntary reporting or incident reporting after the patient is discharged from the hospital. This method includes nontargeted and targeted medical record reviews. For nontargeted medical record reviews, a detailed study of all patient data is performed. For targeted medical record reviews, a trigger tool is used to review a particular set of patients [15]. By contrast, prospective methods, which strive to capture events in real-time or close to real-time, usually rely on automated systems to prevent an event from becoming worse [35]. Information on ADRs is also widely available in unstructured data sources. Typical examples of unstructured data cover published biomedical literature including books, journals and papers, along with narrative clinical notes, online healthcare forums and other patient-generated data shared among networked communities using social media [48]. These new data sources have the potential to provide better drug safety surveillance, resulting in the more accurate detection of safety issues and, thus, have the ability to give more-timely regulatory actions to minimize or mitigate patient risks.

An overview of detection of ADRs at the post-marketing stage

The availability of very large quantities of ADR data provides opportunities for the detecting and prediction of ADRs. Several studies have presented methods to detect and predict ADRs using statistical and data mining techniques [42,54]. Banda *et al.* [2] used a proof-of-concept to explore the chemical space of drugs and establish correlation for ADR prediction. They continued to use the proof-of-concept method to prioritize drug–drug–event associations by mining multiple resources from public databases, spontaneous reports and the literature. Seeing that there was a large volume of drug-related information available in various social networks, Sarker and Gonzalez [44] utilized advanced natural language processing (NLP) and machine learning techniques for extracting useful features such as polarities of sentences, sentiment and topics. These features were then applied to the task of ADR relation extraction for performing automatic ADR detection and monitoring. They reported that combining lexical features from well-established research areas such as sentiment analysis and polarity classification could improve automatic classification of ADR mentions from social media text. According to authors, this scenario was prevalent in social media because patient sentiments during posting were strongly associated with the ADR of the drugs they were taking. They further emphasized that annotated data from distinct sources could be combined to perform ADR detection tasks from social media and multicorpus training could provide significant improvement in accuracy of ADR classification [37].

White *et al.* [42] used a large set of recently labeled ADRs and negative controls to evaluate the ability of search logs to accurately detect ADRs in advance of their publication. They leveraged the

TABLE 1

Data sources for detecting ADRs

Data sources	Justifications	Advantages	Limitations
Primary Data Sources:			
Spontaneous reporting systems (SRS) [1,8,19,24,27]	ADRs are reported to the regulatory authorities by healthcare practitioners, or an individual such as the patient. ADRs are then documented and readily accessed for study, research or review purposes	Highly relevant information collected from diverse international population Public can access to this platform such as FAERS[a] Information collected is wide and varies according to populations which acts as good source for research and study purposes	Some cases are not reported Incomplete data received owing to insufficient information given Some data are duplicated owing to the similar situations ADR reporting is subjected to bias because individual might put blame on medications instead of other reasons such as change in diet or different supplement intake
Electronic health record [1,2,10]	Inpatient information collected through diagnosis, administration of drug. Incidents such as missed dose, double dose or activties that happen in the hospital for each patient is recorded and documented	The information is accurate as detailed information such as diagnosis record, laboratory results, drug dosage are recorded Information collected is varied with different patients owing to different genetic background, different diet, different ethnicity, which gives a high quality for research and study purposes	Sample size is small owing to the limited patients in a hospital The information might not be sufficient for ADR detection owing to the limited patient and short admitted time of patients in the hospital Information are subjected to bias owing to the discrepancy that happens from human errors in the hospital
Scientific literature [1,2,8]	Scientific journals and literature written through peer-review related to ADRs. Information related to ADRs is collected from different studies and analyzed	Information has quality control through peer review Information collected from studies conducted by researchers, which provides a higher quality of the results	Assumptions are included Information is limited to a few studies Information can be biased from the studies collected Study is not of variety owing to limited study and might not be sufficient for getting high accuracy results related to ADRs
Social media [1,6,11,38]	Social media such as twitter, blog, forum related to health, usually written by non-experts. People tend to discuss their health information online by using social media platform	Information is varied and from different groups of populations across the globe Content is huge as growing number of patients are sharing their healthcare experience online, can provide important insight for clinical practioners	Sophisticated linguistic processing is required to account for colloguial language, grammatical or spelling errors. For example, people tend to describe their condition as 'flat out' instead of using the word 'fatigue'. This increases the complexity in interpreting the data related to ADRs Information is limited, because people tend not to share every detail online
Clinical narratives [1,2,8,9,41]	Literature written by healthcare professionals. They wrote the literature according to their experience and also the knowledge that they have	Literature covers a broad range of information including treatment, clinical conditions, patient history Literature is accurate as it is based on the experience and knowledge of healthcare professionals	The information can be bias from the local documentation Information is limited to their experiences and knowledge Information is also limited to local populations
Major secondary data sources:			

TABLE 1 (*Continued*)

Data sources	Justifications	Advantages	Limitations
SIDER 4 [Kuhn *et al.*, 2016]	Contains data on 1430 drugs, 5880 ADRs and 140 064 drug–ADR pairs of approved drugs. A database of the drugs, adverse events and indications mined from the FDA drug labels and Canada's MedEffect resource	Data on drugs, targets and side effects are combined into a more complete picture of the therapeutic mechanism of actions of drugs and the ways in which they cause adverse reactions. Able to reduce the rate of false positives by identifying medical terms that do not correspond to ADRs	Accuracy and number of side-effects depend on management of processing of drug labels Not all ADEs that occur are captured by spontaneous reporting systems. The drug effects need to be observed, recognized, attributed to a drug and then reported.
OFF-SIDES [Tatonetti *et al.*, 2012]	Contains off-label side-effects of drugs generated by using FAERS that collects reports from doctors, patients and drug companies	Contains information complementary to that found in SIDER and improves the prediction of protein targets and drug indications.	Therefore, differential reporting and covariate biases prohibit a straightforward interpretation of the reports
TWOSIDES [Tatonetti *et al.*, 2012]	It is a resource of polypharmacy side-effects for pairs of drugs, generated from adverse event reporting systems	It adapts to specific drug–event pairs, does not require data to be split across strata and can implicitly correct for unmeasured covariates	Same as limitations of OFF-SIDES data source The side-effect frequencies of ADEs and placebo administration were not recruited
MetaADEDB [Cheng *et al.*, 2013]	A comprehensive database of adverse drug events with all compounds and ADEs constructed based on the data downloaded from the comparative toxicogenomics dataset (CTD), SIDER 2 and OFFSIDES. The integrated data from these three databases were annotated with concepts defined in medical subject heading (MeSH)	It serves as valuable resources for chemical biology, drug discovery and pharmacoepidemiology studies The MetaADEDB data source and the network topological properties were much larger than those of SIDER and OFFSIDES. It contained 527 216 drug–ADE associations connecting 3059 compounds and 13 200 ADE items, the largest database developed so far	A small fraction of the drug–ADE associations might be false positive

[a] FAERS indicates FDA adverse event reporting system.

Internet archive to estimate when evidence of an ADR first appeared in the public domain and adjusted the index date in a backdated analysis. They reported that it was possible to detect ADRs in advance of public knowledge using a log-based signal-detection methodology. However, they did not have ground truth on the intentions behind people's observed activity, hence they faced difficulty ascertaining whether the people involved were truly the group who experienced ADRs. They also concluded that the symptoms experienced might be caused by a variety of medical conditions and only some were related to ADRs. Table 2 shows the methods used for processing the data collected from the spontaneous reporting system (SRS) and observational healcare data; Table 3 shows the methods used for processing data from social media. Most of the studies above (or in general pharmacovigiance) focus on using single structured and coded data. For the majority of studies, only one data mining algorithm is used for developing drug safety surveillance, hence the performance of these systems is largely affected by bias owing to the single data source and the shortcoming of adopting one data mining algorithm. By contrast, for studies on 'big data', a large amount of manual annotation is required for conducting large-scale drug safety surveillance across different large data sources. The huge annotation task involved costs a lot of money and it is very time-consuming. For more-accurate and -effective ADR detection, there is a need to use data across different data sources and adopt a more effective model that can reduce the huge volume of manual annotation work [43].

Causation of ADRs

In addition to medication errors, equipment errors and administration errors, ADRs can also be caused by abnormal pharmacokinetics as a result of genetic factors when a drug systematically interacts with human proteins [30,34]. From the biological perspective, drugs are considered as molecules that induce perturbation to biological systems, which involves various molecular interactions such as protein–protein interactions, signaling pathways and pathways of drug action and metabolism [13,31]. This results in systemic ADRs to the drug consumed.

Many patients are on multiple prescriptions and over-the-counter medications in real-life settings [46]. Liu *et al.* [29] defined a synergistic drug–drug interaction (DDI)-induced ADR as arising from simultaneous interactions between coadministered drugs and their protein targets, which caused a new or enhanced ADR beyond what either drug could trigger on its own. Figure 1 schematically illustrates the concept defined by Liu *et al.* [29], showing two drugs that individually interact in the body, each causing therapeutic effects and adverse reactions or enhanced adverse effects when administered together. These effects are rarely investigated at post-drug-marketing stages. Although this is rarely investigated, DDIs are crucially important in knowledge discovery related to DDI-induced ADRs.

Prediction of ADRs using machine learning

The worldwide explosion of big data characterized by its huge and messy volume, high speed and complexity has posed many challenges to healthcare providers. The traditional analytic methods have been found to be unsuitable for dealing with the sheer volume of data and their characteristics. Seeing that machine learning methods have been used extensively in the consumer

TABLE 2

Summary of methods and framework used for processing data collected from multi resources

Study	Methods	Major Findings
Signal detection[a] methods applied to SRS data[b]		
Holle and Bauchau [19]	147 015 data extracted from corporate SRS, dated from 1987 to 2000 Methods: multi-item gamma poisson shrinker (MGPS) vs a time-to-onset (TTO) algorithm	When data are of low quality, both methods are suggested to benefit from the greater ability of TTO to detect true positive (TP)[c] signals, while avoiding signals being missed (or delayed)
Harpaz *et al.* [18]	4 784 337 data extracted from public FDA adverse event reporting system (FAERS) reports from 1968 to 2011 Methods MGPS, proportional reporting ratio (PRR), reporting odds ratios (ROR), logistic regression (LR), extended logistic regression (ELR)	Multivariate modeling methods are more superior than DP-based (disproportionality-based) methods but DP-based methods are simpler and faster to compute Not all events are equally detectable
Ibrahim *et al.* [21]	632 722 data extracted from FAERS reports from 4th quarter 2012 till 2nd quarter 2013. Methods 'hybrid Apriori algorithm'. It extracted 2933 interacting DIAE. Used Apriori algorithm proof to detect severe life-threatening FD-labeled effectively besides having the ability to detect rare ADRs	Apriori algorithm. Precision of 85%, sensitivity of 81%
Signal detection methods applied to observational healthcare data		
Ryan *et al.* [41]	Data ten observational databases elaborated observational medical outcomes partnership (OMOP) (over 130 millionvrecords) Methods eight methods from the OMOP library	Many false positive (FP)[d] associations obtained from all methods. Result is dependent on the desired tradeoff between Sn and Sp with no clear optimal algorithm
Schuemie et al. [45]	Data extracted from seven databases across three European countries, dated from 1997 to 2000 Four DP-based methods (BCPNN, Bayesian confidence propagation neural network; GPS, gamma poisson shrinker; PPR, proportional reporting ratio; ROR, reporting odds ratio), three cohort methods (BHM, Bayesian hierarchical model; IRR, incidence rate ratio; LGPS, longitudinal gamma poisson shrinker), two case-based methods [matched CC (case control), SCCS (the self controlled case series)], one method for eliminating protopathic bias employed in combination with previous methods (LEOPARD, longitudinal evaluation of observational profiles of adverse events related to drugs)	LEOPARD had a positive impact on the overall performance of all methods DP-based methods had lower but not statistically significant performance Some ADRs were not detected by all methods
Peddi *et al.* [40]	Data from EHR database Methods baseline regularization (BR) and multiple self-controlled case series (MSCCS)	This MSCCS model has improved the performance in detecting ADRs
Semantically-Enriched, Integrated Signal Detection Method		
Koutkias and Jaulent [23]	Multiple heterogeneous data sources explored by using multiple methods	Systematic frameworks that will enable pharmacovigilance stakeholders to seamlessly share, access, and different data sources and computational methods for signal detection

[a] Signal that used to detect the occurrence of ADRs.
[b] Data collected from spontaneous reporting system.
[c] TP means true positive which is a positive result.
[d] FP means false positive which is an error in data reporting in which a test result improperly indcates presence of a condition (such as a disease) when in reality it is not present.

retail sector, researchers started to explore this alternative to replace traditional statistical methods. The term 'machine learning' refers to techniques that give the computer the capability of acting without being explosively programmed, and algorithms are constructed to enable computers to learn from data and make a prediction [7]. Commonly used machine learning methods for developing classification models are as follows [49]:

i. Linear discriminant analysis (LDA);
ii. k nearest neighbor (kNN) and kNN regression (kNNR);
iii. Artificial neural network (ANN);
iv. Probabilistic neural network (PNN);
v. Support vector machine (SVM) and support vector regression (SVR);
vi. C4.5 decision tree (C4.5DT);
vii. Recursive partitioning (RP) classifiers;
viii. Random Forest (RF);
ix. Naive Bayesian classifiers;
x. Multiple linear regression (MLR);
xi. Partial least squares regression (PLSR) and logistic PLS;
xii. Combined classifiers approach.

During recent years, machine learning methods have been used to extract information on drugs and predict side-effects of drugs during the drug discovery process. Supervised machine learning had been used to extract drug information and predict ADR relations. Nikfarjam and Gonzalez [37] applied the association rule to mine annotated comments for identifying language patterns frequently used for ADR mentions in social media. Patki *et al.* [38] used two supervised machine learning algorithms: multinomial naive Bayes and SVM, for mining health-related forums. Applying a two-step approach, they classified user comments into two classes: one with mentions of ADR and another one without mentions of ADR. Then, they inferred whether the combined

TABLE 3

Summary of methods used for processing data collected from social media

Study.	Concept	Methods	Major findings
Leaman *et al.* [25]	Concept/relation extraction	Data trained on conditional random field (CRF). Lexicon-based (450 comments for system development)	Quantitative. Against manually annotated data. [a] Precision: 78.3%; [b]recall: 69.9%; [c]F-measure: 73.9%
Nikfarjam and Gonzalez [37]	Association rule mining on a set of annotated comments	Lexical pattern-matching (2400 comments for pattern building). Association rule mining to identify patterns	Quantitative. Against manually annotated data. Precision: 70%; recall: 66.3%; F-measure: 68.0%
Chee *et al.* [12]	Drug classification	Ensemble classification using support vector machine (SVM) and naive Bayes	Mixed. Classification results are combined to generate drug scores for three drugs, which are compared against scores for drugs (12) with known adverse effects, able to identify risky drugs for FDA scrutiny
Benton *et al.* [6]	Concept/relation extraction	Lexicon-based. association rule mining to identify drug–reaction pairs	Quantitative. Adverse reactions associated with drugs obtained from product labels and compared against system reported adverse events
Hadzi-Puric and Grmusa [17]	Concept/relation extraction	Lexicon-based approach for ADR detection. Statistical scoring for identifying drug–relation associations	Mixed. Qualitative analysis of identified ADRs against known ADRs. Precision: 75.3%; recall: 64.7% F-score: 69.599%
Bian *et al.* [11]	ADR classification	Classification of tweets using SVM classifiers. Two classifiers built: one to predict whether a user has used a drug (based on the tweets) and the second to classify if a post contains an adverse effect	Mixed. Evaluation and training is performed on the same data. Only classification accuracies reported. Precision: 70%; recall: 69%
Liu *et al.* [31]	Concept/relation extraction	Lexicon-based approach for ADR and drug detection. Shortest dependency-path-based machine learning algorithm for relation extraction	Quantitative. Separate evaluations for entity extraction, ADR detection and classification of patient experiences using 200 manually annotated comments

[a] Precise measures the number of actual positive cases out of those predicted positive cases.
[b] Recall calculates how many of the actual positives the model captures through labeling it as positive.
[c] F1 score is used to seek a balance between precision and recall owing to uneven data distribution.

comments of the drug indicated an inordinate incidence of adverse reactions. However, the supervised machine learning approach is laborious. All user comments must be annotated by domain experts or under the guidance of a pharmacology expert to create the training data. This might not be feasible when researching many drugs and ADRs with only a very small amount of drug-related ADRs posted. Moreover, results are highly dependent on the richness and representativeness of the training data. This is very challenging because patients tend to misspell and use their own terms instead of medical terms to discuss health issues on social media.

The machine learning methods were also extended to studies that took into account the chemical and biological features of drugs such as chemical structures, target protein, protein–protein interaction, gene ontology annotations and others at the pre-marketing stage. Huang *et al.* [20] used drug targets, protein–protein interactions and gene ontology annotations as features and adopted SVM and logistic regression to construct predictors; whereas Pauwels *et al.* [39] used four machine methods, namely kNN, SVM, ordinary canonical correlation analysis and sparse canonical correlation analysis to train the prediction model based on chemical structure. Mizutani *et al.* [33] adopted the sparse canonical correlation to build prediction models for the combined chemical structure and target proteins. By contrast, Liu *et al.* [31] integrated a wide variety of drug-related features and adopted five machine learning classifiers, namely logistic regression, naive Bayes, kNN, Random Forest and SVM, to build the prediction models. At the post-marketing stage, Cheng and Zhao [13] applied five machine-learning-based predictive models: naive Bayes, decision tree, kNN, logistic regression and SVM, on a large DDI dataset from the DrugBank database to predict unknown DDIs based on drug phenotypic, therapeutic, structural and genomic similarities. They showed that the machine-learning-based integration of drug phenotypic, therapeutic, structural and genomic similarities using a systems pharmacology approach was a simple yet efficient strategy to predict unknown DDIs. Zhang *et al.* [59] designed two recommender methods [i.e., the integrated neighborhood-based method and the restricted Boltzmann machine (RBM)-based method] to make prediction on potential side-effects of drugs. They concluded that the proposed RBM method, integrated neighborhood-based method and the ensemble method were promising for predicting the potential side-effects of approved drugs. Dimitri and Lio [16] presented a new machine learning tool known as DrugClust as well as K-seeds, and they adopted the Bayesian score to build a valid instrument for predicting drug side-effects. This framework could be easily modified and adapted to different kinds of drug features and clustering methods. DrugClust showed itself to be a promising and easy-to-use tool for predicting drug side-effects. A study by Dimitri and Lio [16] also proved that machine learning techniques such as DrugClust as well as the cluster algorithms and K-seeds have shown accuracy in predicting ADRs. They used the results from Zhang *et al.* [58], Liu *et al.* [31], and Mizutani *et al.* [33] as the benchmark references for their results. Their study showed that the results produced from DrugClust are consistent with the three benchmark references.

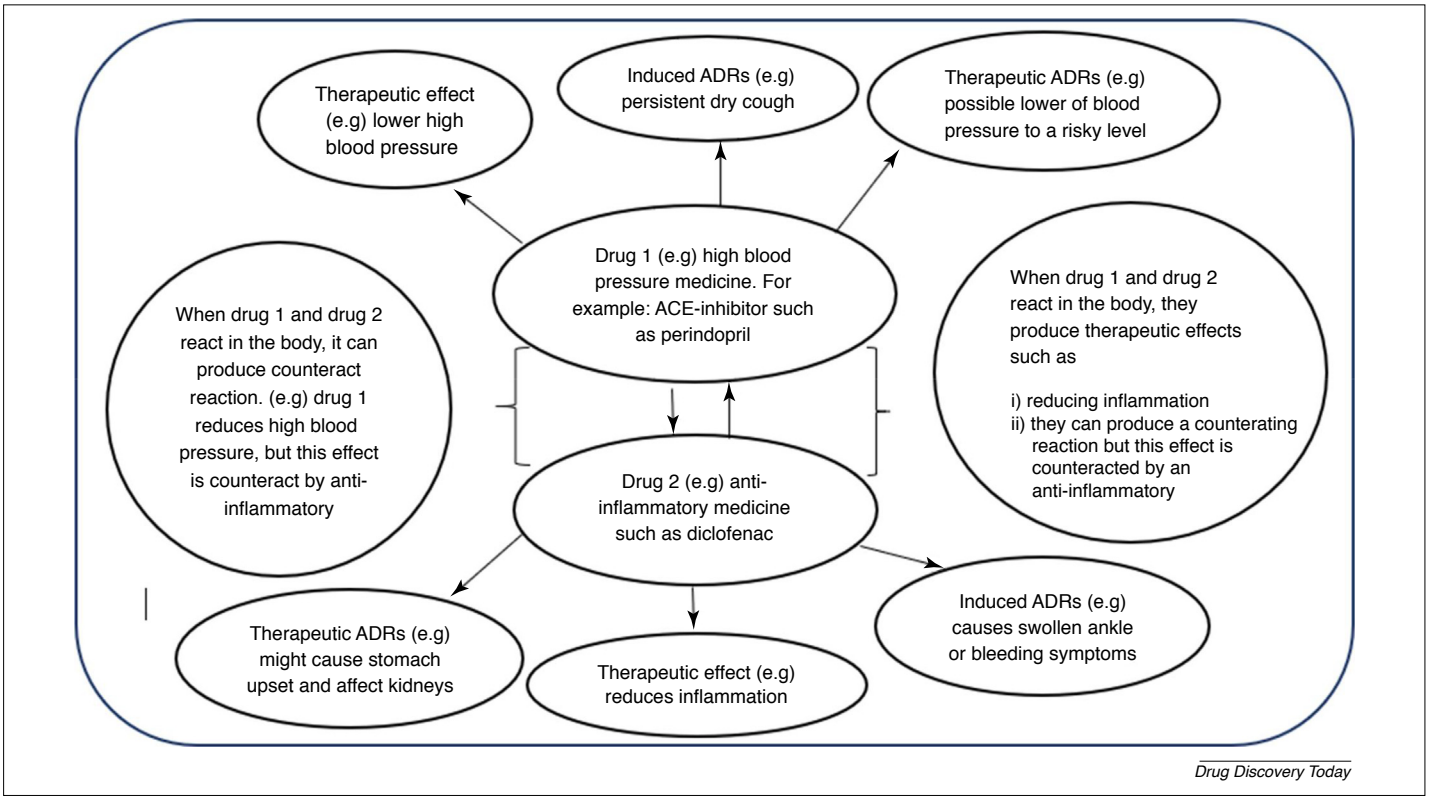

FIGURE 1

Schematic illustration of drug reactions in the body. Co-administration of Drug 1 and Drug 2 may result in therapeutic effect or may induce other adverse effects or may enhance the existing ADRs caused by individual drugs.

In the study conducted by Xiong *et al*. [55], the authors used machine learning approaches to study cytokine release syndrome (CRS). Many ADRs of drugs that were not discovered during animal testing were discovered when machine learning approaches were used. Xiong *et al*. [55] used machine learning techniques such as hierarchical cluster analysis (HCA), principal component anaylysis (PCA) and decision tree classification (DTC) to study CRS on donors. Their results showed that these three approaches reveal different information on the response produced in the body from the drugs [55].

Combined classifier models, which were ensembles of different independent machine learning classifiers including SVM, decision tree, kNN and naive Bayes, fused by a back-propagation artificial neural network (BP-ANN), were developed to predict inhibition of five major CYP isoforms, namely CYP1A2, 2C9, 2C19, 2D6 and 3A, based on a large dataset of >24 700 unique compounds (Cheng *et al*. [13]). All developed models were validated by fivefold cross-validation and a diverse validation set composed of ~9000 diverse unique compounds. The overall performance of the combined classifiers fused by BP-ANN was found to be superior to that of three classic fusion techniques (mean, maximum and multiply). Another combined classifier framework was developed through integration of five machine learning algorithms: logistic regression, Random Forest, kNN, SVM and neural network (NN) (Cheng *et al*., 2018 [13]). A NN algorithm was used to integrate the best four single classifiers for each cardiovascular outcome (hypertension, heart block, arrhythmia, cardiac failure and myocardial infarction). The combined classifiers were recorded to have higher performance with an AUC range from 0.784 to 0.842 compared with single classifiers.

The nonliner and imbalanced nature of biological data together with the need to analyze the massive amount of big data has motivated researchers to use machine learning as an alternative to traditional statistical models; machine learning methods demonstrate good performance only when the dataset is small. Machine learning methods that are operating on big data do not correct the problem of bias [20]. The big data fed into machine learning can help to provide broader data by linking more variables such as patient medical comordities, the drugs they take and so on [41,47]. However, the data are unable to measure disease severity, disease stage and others [14]. The emergence of big data has motivated some researchers to explore a more powerful and efficient approach known as deep learning, which is a subset of machine learning designed to deal with that massive amount of data for drug discovery [56].

From machine learning to deep learning for prediction of ADRs

As described by Zhang *et al*. [57], deep learning is like deep NNs that consist of nonlinear processing units that tranform information from one level to another level. It is a special but increasingly large branch of machine learning. One area of machine learning is NNs [57]. NNs can be divided into various forms such as convolutional neural networks (CNN or Convnet) and recurrent neural networks (RNN). The networks consist of many layers of representation, starting from the low representation (raw input) to hidden layers, and then to higher levels of representations. The addition of

TABLE 4

Differences between machine learning and deep learning

Machine learning	Deep learning
Only consists of one layer without having the hidden layers [23]	Consists of many hidden layers of representation learning. It can transform the representations or features at one level (raw input) into a representation at a higher and more representative level [26]
Requires manual annotation by domain experts, very time consuming [55,57]	Does not require manual annotation, able to perform automatic feature extraction (e.g., deep learning methods can automatically select representations from raw, high dimension and heterogeneous molecular descriptors) [55,57]
Suitable to process small datasets. The data size can be increased greatly, but the machine is unable to correct the problem of bias if the data are lacking key clinical severity measures [14]	Has the capability of processing huge datasets (big data) that carry high and heterogeneous dimensions from diverse sources (e.g., can even process large chemical and biological data) [55,57]
Common classifiers used are [55,57]: - Linear discriminant analysis (LDA) - Support vector machines (SVM) - Decision trees (DTs) - Random forest (RF) - k-nearest neighbor (kNN) - Artificial neural networks (ANN)	Commonly used networks are [55,57]: - Convolutional neural networks (CNN) - Stacked autoencoders - Deepe belief networks (DBN) - Restricted Boltzmann machine (RBM)

several hidden layers makes the NN 'deep' and, hence, it is a broad and expanding category of algorithmic methods with the ability of taking in large amounts of data and processing them through the neural network, which then extracts the information and produces the results.

The most commonly used deep learning networks are CNN, stacked autoencoders, deep belief networks (DBN) and restricted Boltzmann machines [57]. Zhang *et al.* [57] provided an illustration of deep learning where the raw input (e.g., the color of the image) data were collected and then transformed in the next layer of the processing unit to produce output data. Table 4 and Fig. 2 illustrate the difference between machine learning and deep learning.

Deep learning has been used in drug discovery for new drug molecule identification, protein engineering, gene expression data analysis and pharmacodynamics modeling. According to Wang *et al.* [52], the deep learning approach was utilized in protein engineering with the aim of discovering more potential functions of protein because the results produced are of higher accuracy. Owing to its capability to take in a massive amount of data and extract these data, deep learning was also adopted in medicine development, sequence specification prediction and genomic modeling for drug repurposing [26]. Deep learning techniques are now used in precision medicine where they are used to genetically tailor to the gene of an individual to predict future diseases that could be contracted by an individual [28]. It was demonstrated that the efficacy of restricted deep learning on addressing multilabeled learning with incomplete labels was good, especially at addressing a vanishing gradient problem [53].

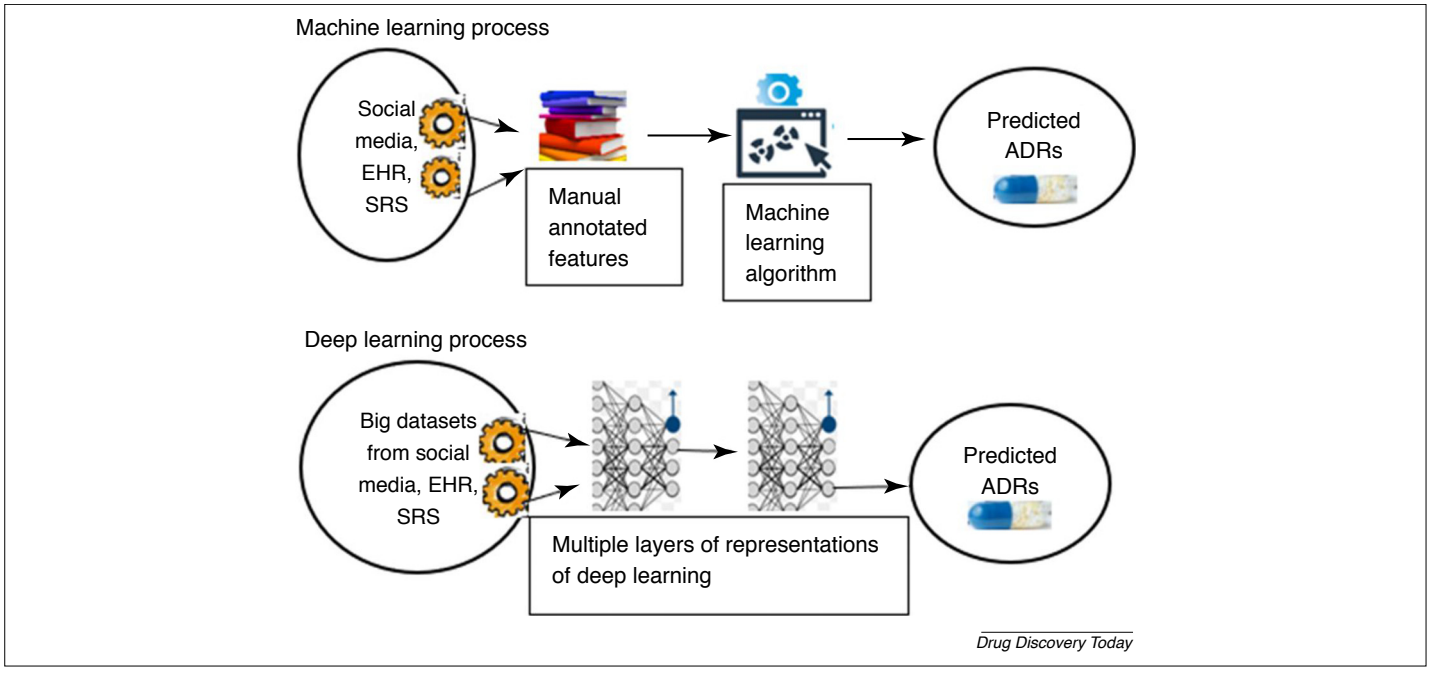

FIGURE 2

Schematic overview of steps involved in machine learning process and deep learning process for prediction of ADRs.

Although much effort has been expended on using deep learning for drug discovery, very little effort has been made to predict ADRs using deep learning methods. Based on its demonstrated potential in drug discovery, it is proposed that deep learning must be exploited for the prediction of ADRs because it has the capability of eliminating the need for the manual annotation of big data, which is very time-consuming. It can also perform automatic feature extraction, without the need for human input, from raw big data that carry high dimensionality, contain a lot of noise, are heterogeneous, sparse, incomplete and contain random errors and systematic biases.

For deep learning, when large NNs are constructed and trained with an increasing amount of data, their performance will continue to increase, which is different from machine learning techniques which can reach a plateau in performance at a certain stage of training. It can take much more time at the beginning to fine-tune the network to allow features from the higher levels of the hierarchy to be formed by the composition of the lower-level features; but the learning features at multiple levels of abstraction will allow a system to learn complex functions and map the input directly to the output data without depending on human-crafted features. If the network is very deep, the computer will continue to teach itself as it reads more input data and it will become increasingly efficient as the training continues.

The black box

The 'non-interpretability' of a deep learning model and its limited 'explainability' is a hot topic in the artificial intelligence (AI) community. It is frequently referred to as a black box with data fed to the machine as an input at one end, and decisions come out as an output at the other end, but the processes between input and output are opaque. This highly complex deep learning model, where input data can undergo complex transformation in multiple hidden layers in an unexplainable way, has led many to question the transparency of the model. Because of this, although deep learning models are giving increasingly advanced results in diverse problems, their lack of interpretability is a major problem. Compared with the network-based approach which uses network proximity measures to clearly quantify the relationship between side-effect and drug target in a human protein–protein interaction network, the outcomes obtained using the deep learning approach are still nonqualitative and unexplainable, and it only provides a predictive score for each prediction, where a higher score means a higher probability of occurrence. By contrast, the network-based approach can give a clear explanation on how incorporating network proximity together with large-scale patient longitudinal data and *in vitro* experimental assays can facilitate prediction and drug repurposing (Cheng *et al.* [13]).

For prediction of ADRs, an understanding of why, what and how ADRs are detected by algorithm learning models is important. Healthcare providers expect the machine to be able to explain why, what and how a decision is reached. This is especially important for deep learning, where input data undergo complex transformation in multiple hidden layers and the model can become very complex. For deep learning to enter the highly regulated medical field, the science of implementing deep learning must be improved and interpretable deep learning algorithms must be developed to unfold the black box. Its performance can be further enhanced by incorporating the network interaction features into the model.

Potential applications for deep learning

Deep learning has shown remarkable growth and outstanding performance in many areas ranging from image recognition and natural language processing to biomedical applications. Despite its outstanding performance and its great potential capability of performing automatic feature extraction, deep learning is still widely used as a simple classifier. In this review, we focus on developing a novel conceptual framework of an interpretable end-to-end deep learning model that can predict ADRs from hidden features of chemical structures of drugs and all the latent features from the biological, chemical, phenotypic and network data. By automatically learning from hidden features of the drug structure and other pertinent data directly, our approach can alleviate the sophisticated effort and time required for expensive feature engineering processes and it enables prediction of interpretable ADRs that might be unknown to domain experts.

Recognizing the importance of using various health datasets for the prediction of ADRs, we propose exploiting the capability of deep learning for the prediction of ADRs as shown in our two-stage deep learning framework in Fig. 3. The first stage comprises the collection of various types of data from publicly available big datasets. In this paper, to overcome the 'black box' nature of deep learning, we propose to incorporate pathway- and network-based databases such as drug substructure data, drug target data, drug enzyme data, drug transporter data, drug pathway data, drug indication data, drug side-effect data, drug off-side-effect data as well as network data as training and testing datasets. To produce interpretable ADR predictions directly from low-level representations, SMILES – a simple line text-based molecular notation format – represents the chemical structure that will be used. The information about targets, transporters, enzymes and pathways can be extracted from DrugBank. To obtain the drug indication and pharmacological information, the side-effect records of drugs in the SIDER, OFFSIDES, social media, clinical narrative and spontaneous reporting systems are also included. Deep learning has the capability of automatically learning meaningful hidden feature representations by training and transforming the multisource input data in the complex multiple layer networks. The learned network can discover high-level features, increase interpretability and provide biological insights to understand the nature of the ADR prediction. The results obtained can be compared with those of FAERs to evaluate the accuracy of prediction.

After predicting the ADRs for the drugs, we further extend the applications of deep learning to the second stage for other predictive applications, such as drug repurposing and precision medicine. For example, in the case of drug repurposing, drugs that cause ADRs of hypotension can be used for the treatment of hypertension; whereas drugs that cause hair growth as a side-effect can be repurposed for treating hair loss. Deep learning methods such as DNNs can be applied to predict pharmacologic properties of the drugs and train the networks to repurpose the drug usage (Fig. 3). This will save the drug manufacturers an enormous amount of time and effort in getting a new drug approved because detailed information on pharmacology, formulation and potential toxicity of the existing drugs is already available and, hence, these drugs

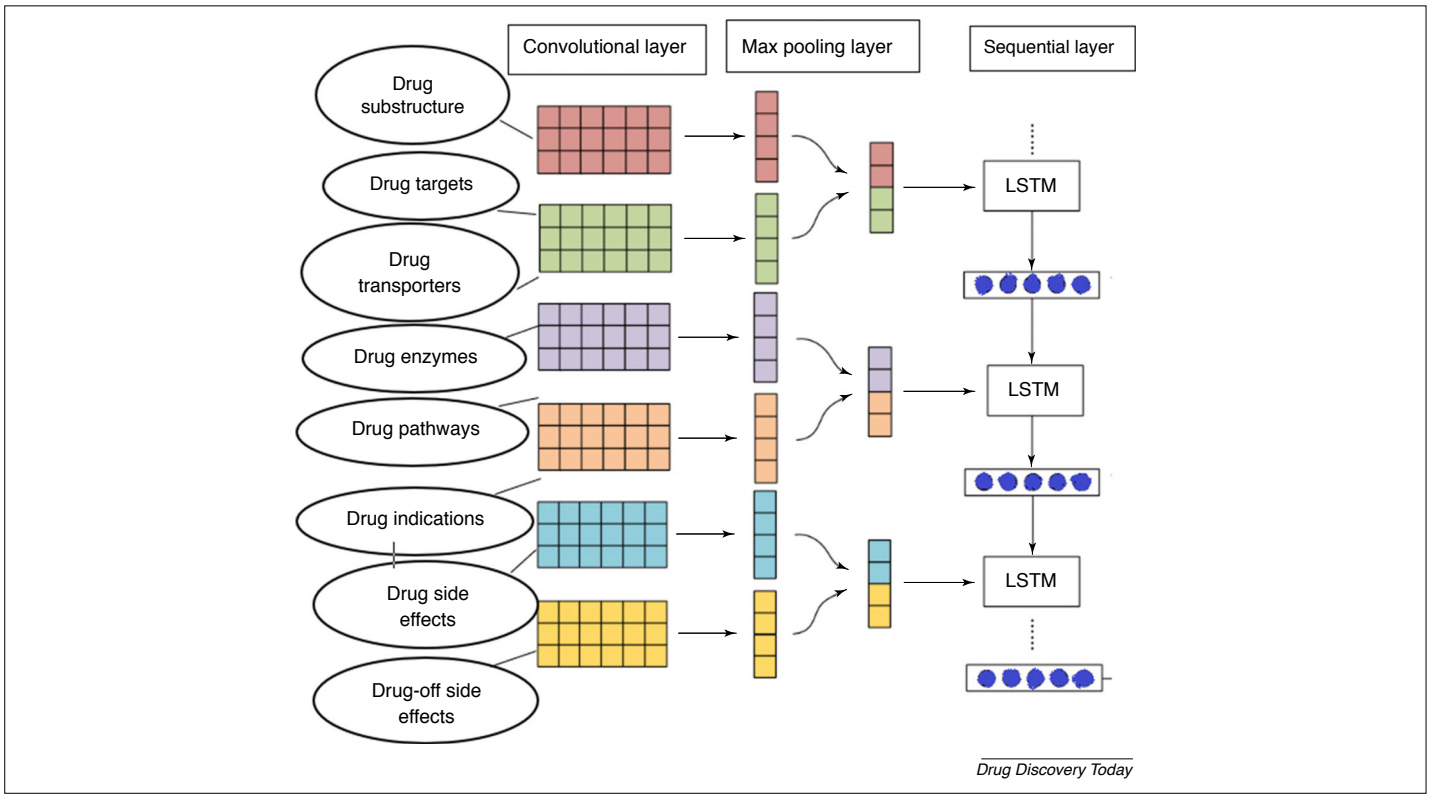

FIGURE 3

Schematic overview of our proposed two-stage deep learning conceptual framework.

can be ready for clinical trials quickly and reviewed by the FDA for immediate approval.

In pharmacovigilance, the body's response to drugs is a complex phenomenon. Every drug is different when it interacts with different human bodies. A drug that works well for one person might not work well for another. Conventional prediction methods, although they have been successful in showing the association between drug(s) and associated ADRs, are unable to deduce the causal relationships of ADRs, establish the biological properties of the drug and patient, as well as determine their degree of severity. Because the mechanism and reasons why ADRs occur are largely unknown, many times, physicians are unaware of ADRs that can occur for patients with no previous ADR experiences or physicians might fail to foresee any new ADRs that are different from previous cases that have occured for the same patient. Physicians tend to prescribe what has worked for other patients in the past, as opposed to giving precision medicine treatment to patients based on biological, chemical and phenotypic information. To improve the healthcare service, deep learning can be used to drive toward precision medicine, as shown in the second stage of our framework in Fig. 3. Patient medical profile data, such as weight, age, blood pressure, polypharmacy, previous ADRs, genomic and phenotypic data, as well as data on chemical compounds, among others, available in the large-scale electronic healthcare data can be used as input for the deep learning process. Every patient in the data warehouse can be represented using these features. We can then derive the patient representation using a multilayer NN in a deep learning architecture. Each layer of the network is trained under unsupervised learning to produce a higher-level representation of the observed patterns, based on the data it receives as input from the layer below. Every level produces a representation of the input pattern that is more abstract than the previous level because it is obtained by composing more nonlinear operations. Briefly, our proposed two-stage deep learning conceptual framework in Fig. 3 can be summarized as follows:

- Stage one of the deep learning process:
  - preprocessing stage to obtain data representations from the big data;
  - the data representations are modeled by the unsupervised deep architecture leading to a set of general and robust features for prediction of ADRs.
- Stage two of the deep learning process:
  - for precision medicine, the deep learning is applied to drugs and ADR features together with the entire extracted EHR database to derive precision medicine presentations;
  - for repurposing of drugs, deep learning can be applied to drugs and their associated ADR features together with chemical and biological properties associated with the problem to be solved to repurpose the drug usage.

With the availability of predicted ADRs for patients based on biological, chemical, phenotypic as well as network data, physicians can then use deep learning ADR prediction approaches to monitor the progress of their patients' treatment to check whether any ADRs are likely to occur before any new medicine is prescribed. The platform can also automatically detect any patient with a high probability to develop a certain ADR and the level of its seriousness. Healthcare providers can then take appropriate action to

address the issue by carefully monitoring the treatment progress of the related patient. This will result in better clinical decision making and achieve better clinical outcomes.

Concluding remarks

Here, we review and compare numerous traditional approaches and machine learning methods for pharmacovigilance. Based on demonstrated achievements of deep learning in the drug discovery stage for detection of drug side-effects, we propose a new two-stage deep learning framework for prediction of ADRs during post-marketing drug surveillance. The new framework consists of two stages with the first stage used to predict the association between drugs and their adverse reactions by using deep learning methods running on big data. The second stage focuses on precision medicine by integrating an individual's biological data into the system. Drugs that are considered to be harmful are then exploited for repurposing. In comparison with traditional statistical models and machine learning methods, the advantages of using this two-stage deep learning framework are: (i) deep learning has a powerful capability to accommodate and process useful information using big data obtainable from heterogeneous sources for prediction of ADRs; (ii) it can integrate information written using different vocabularies from multiple sources; (iii) it can perform many layers of learning to form new associations and representations. Hence, our proposed two-stage deep learning conceptual framework shows great promise to help maximize the predictive capabilities that can effectively infer associations and model new potential properties of drugs, thereby showing potential to predict ADRs as well as improving the potential for drug repurposing. With this proposed framework, it is our hope that the crucial ADR issue can be reduced or resolved as physicians can make full use of the approach to perform ADR prediction, which will enable them to monitor their patients' treatment progress and hence they can check whether an ADR is likely to occur before any new medicine is prescribed. For the pharmaceutical industry, the drug company can fully translate the side-effects of drugs into different drug usage for better healthcare purposes. With the adoption of this framework, drug manufacturers, healthcare providers, as well as patients will receive tremendous benefits in the future.

References

1 Bailey, C. *et al.* (2016) Adverse drug event reporting systems: a systematic review. *Br. J. Clin. Pharmacol.* 82, 17–29
2 Banda, J.M. *et al.* (2016) Feasibility of prioritizing drug–drug–event associations found in electronic health records. *Drug Safety* 39, 45–57
3 Banda, J.M. *et al.* (2016) Data descriptor: a curated and standardized adverse drug event resource to accelerate drug safety research. *Sci. Data* 3, 160026
4 Bender, A. *et al.* (2007) Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure. *ChemMedChem* 2, 861–873
5 Chan, M. *et al.* (2001) Adverse drug events as a cause of hospital admission in the elderly. *Intern. Med. J.* 31, 199–205
6 Benton, A. *et al.* (2011) Identifying potential adverse effects using the web: a new approach to medical hypothesis generation. *J. Biomed. Inf.* 44, 989–996
7 Liu, W. *et al.* (2017) A survey of deep neural network architectures and their applications. *Neurocomputing* 234, 11–26
8 Pirmohamed, M. *et al.* (2004) Adverse drug reactions as cause of admission to hospital: prospective anaysis of 18820 patients. *BMJ* 329, 15–19
9 Rothschild, J.M. *et al.* (2010) Medication errors recovered by emergency department pharmacists. *Ann. Emerg. Med.* 55, 513–521
10 Jordan, S. *et al.* (2016) Adverse drug reactions, nursing and policy: narrative review. *Ann. Nurs. Pract.* 3, 1050
11 Bian, J. *et al.* (2012) Towards large-scale twitter mining for drug-related adverse events. In *Proceedings of the 2012 International Workshop on Smart Health and Wellbeing* 2012, 25–32
12 Chee, B.W. *et al.* (2011) Predicting adverse drug events from personal health messages. *AMIA Annu. Symp. Proc.* 2011, 217–226
13 Cheng, F. and Zhao, Z. (2014) Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. *J. Am. Med. Inf. Assoc.* 21, e278–286
14 Crown, W.H. (2015) Potential application of machine learning in health outcomes research and some statistical cautions. *Value Health* 18, 137–140
15 Cullinan, S. *et al.* (2016) Use of a frailty index to identify potentially inappropriate prescribing and adverse drug reaction risks in older patients. *Age Aging* 45, 115–120
16 Dimitri, G.M. and Lio, P. (2017) DrugClust: a machine learning approach for drugs side effects prediction. *Comput. Biol. Chem.* 68, 204–210
17 Hadzi-Puric, J. and Grmusa, J. (2012) Automatic drug adverse reaction discovery from parenting websites using disproportionality methods. In *Proceedings Mining (ASONNAM 2012) IEEE Computer Society* . http://dx.doi.org/10.1109/ASONAM.2012.144
18 Harpaz, R. *et al.* (2013) Performance of pharmacovigilance signal-detection algorithms for the FDA adverse event reporting system. *Clin. Pharmacol. Ther.* 93, 539–546
19 Holle, V.L. and Bauchau, V. (2014) Signal detection on spontaneous reports of adverse events following immunization: a comparison of the performace of a disproportionality-based algorithm and a time-to-onset-based algorithm. *Pharmacoepidemiol. Drug Saf.* 23, 178–185
20 Huang, L.C. *et al.* (2011) Predicting adverse side effects of drugs. *BMC Genom.* 12, S11
21 Ibrahim, H. *et al.* (2016) Mining association patterns of drug-interactions using post marketing FDA's spontaneous reporting data. *J. Biomed. Inf.* 60, 294–308
22 Kane-Gill, S.L. *et al.* (2016) Clinical decision support for drug related events: moving towards better prevetion. *World J. Crit. Care Med.* 5, 204–211
23 Koutkias, V.G. and Jaulent, M.C. (2015) Computational approaches for pharmacovigilance signal detection: toward integrated and semantically-enriched frameworks. *Drug Saf.* 38, 219–232
24 Kuang, Z.B. *et al.* (2017) Parmacovigilance via baseline regularization with large-scale longitudinal observational data. *Appl. Data Sci. Pap.* 2017, 1537–1546
25 Leaman, R. *et al.* (2010) Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks. In *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing, Association for Computational Linguistics* 2010, 117–125
26 LeCun, Y. *et al.* (2015) Deep learning. *Nature* 521, 436–444
27 Lee, S. *et al.* (2017) Standard-based comprehensive detection of adverse drug reaction signals from nursing statements and laboratory results in electronic health records. *J. Am. Med. Inform. Assoc.* 24, 697–708
28 Liang, M. *et al.* (2015) Integrative data analysis of multi-platform cancer data with a multimodal deep learning approach. *Bioinformatics* 12, 928–937
29 Liu, R. *et al.* (2017) Data-driven prediction of adverse drug reactions induced by drug–drug interactions. *BMC Pharmacol. Toxicol.* 18, 44
30 Liebler, D.C. and Guengerich, F.P. (2005) Elucidating mechanisms of drug-induced toxicity. *Nat. Rev. Drug Discov.* 4, 410–420
31 Liu, M. *et al.* (2012) Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. *J. Am. Med. Inf. Assoc.* 19, e28–e35
32 Marcum, Z.A. *et al.* (2012) Prevalence of unplanned hospitalizations caused by adverse drug reactions in older veterans. *J. Am. Geriatr. Soc.* 60, 34–41
33 Mizutani, S. *et al.* (2012) Relating drug–protein interaction network with drug side effects. *Bioinformatics* 28, 1522–1528
34 Moore, C. *et al.* (2009) Predictive value of alert triggers for identification of developing adverse drug events. *J. Patient Saf.* 5, 223–228
35 Nair, N.P. *et al.* (2016) Hospitalization in older patients due to adverse drug reactions — the need for a prediction tool. *Clin. Intervent. Aging* 11, 497–505
36 Nebeker, J.R. *et al.* (2004) Clarifying adverse drug events: a clinician's guide to terminology, documentation, and reporting. *Ann. Intern. Med.* 140, 795–801
37 Nikfarjam, A. and Gonzalez, G.H. (2011) Pattern mining for extraction of mentions of adverse drug reactions from user comments. *AMIA Annu. Symp. Proc.* 2011, 1019–1026

38 Patki, A. *et al.* (2014) Mining adverse drug reaction signals from social media: going beyond extraction. In *Proceedings of the 22nd Annual International Conference on Intelligent Systems for Moleculer Biology* 2014, 9–16

39 Pauwels, E. *et al.* (2011) Predicting drug side-effects profiles: a chemical fragment-based approach. *BMC Bioinf.* 12, 169

40 Peddi, D. *et al.* (2016) Designing an adverse drug event reporting system to prevent unintentional reexposures to harmful drugs: study protocol for a multiple methods design. *JMIR Res. Protoc.* 5, 169

41 Ryan, P.B. *et al.* (2012) Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. *Stat. Med.* 31, 4401–4415

42 White, R.W. *et al.* (2016) Early identification of adverse drug reactions from search log data. *J. Biomed. Inf.* 59, 42–48

43 Salathe, M. (2016) Digital pharmacovigilance and disease surveillance: combining traditional and big-data systems for better public health. *J. Infect. Dis.* 214, s399–s403

44 Sarker, A. and Gonzalez, G. (2015) Portable automatic text classification for adverse drug reaction detection via multi-corpus training. *J. Biomed. Inf.* 53, 196–207

45 Schuemie, M.J. *et al.* (2012) Using electronic health care records for drug safety signal detection: a comparative evaluation of statistical methods. *Med. Care* 50, 890–897

46 Sonichsen, A. *et al.* (2016) Polypharmacy in chronic diseases — reduction of inappropriate medication and adverse drug events in older populations by electronic decision support (PRIMA-eDS): study protocol for a randomized controlled trial. *Trials* 17, 57

47 Song, Q. *et al.* (2017) An evolutionary deep neural network for predicting morbidity of gastrointestinal infections by food contamination. *Neurocomputing* 226, 16–22

48 Stockwell, D.C. and Kane-Gill, S.L. (2010) Developing a patient safety surveillance system to identify adverse events in the intensive care unit. *Crit. Care Med.* 38, S117–S125

49 Tao, L. *et al.* (2015) Recent progresses in the exploration of machine learning methods as in-silico ADME prediction tools. *Adv. Drug Deliv. Rev.* 86, 83–100

50 Topaz, M. *et al.* (2016) Towards improved drug allergy alerts: multidisciplinary expert recommendations. *Int. J. Med. Inf.* 2016, 353–355

51 Voss, E.A. *et al.* (2017) Accuracy of an automated knowledge base for identifying drug adverse. *J. Biomed. Inf.* 66, 72–81

52 Wang, M.D. and Hassanzadeh, H.R. (2017) DeeperBlind: enhancing prediction of sequence specificities of DNA binding proteins. *BIBM* . http://dx.doi.org/10.1109/BIBM.2016.7822515

53 Wang, Y. and Zeng, J. (2013) Predicting drug–target interactions using restricted Boltzmann machines. *Bioinformatics* 29, 126–134

54 Xiao, C. *et al.* (2017) Adverse drug reaction prediction with symbolic latent Dirichlet allocation. *Conf. Artif. Intell.* 2017, 1590–1597

55 Xiong, F. *et al.* (2014) Analysis of cytokine release assay data using machine learning approaches. *Int. Immunopharmacol.* 22, 465–479

56 Yamanishi, Y. *et al.* (2012) Drug side-effect prediction based on the integration of chemical and biological spaces. *J. Chem. Inf. Model.* 52, 3284–3292

57 Zhang, L. *et al.* (2017) From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug Discov. Today* 22, 1680–1685

58 Zhang, W. *et al.* (2016) Predicting potential side effects of drugs by recommender methods and ensemble learning. *Neurocomputing* 173, 979–987