



On the road to explainable AI in drug-drug interactions prediction: A systematic review



Thanh Hoa Vo ^{a,1}, Ngan Thi Kim Nguyen ^{b,1}, Quang Hien Kha ^c, Nguyen Quoc Khanh Le ^{d,e,f,*}

^a Master Program in Clinical Genomics and Proteomics, College of Pharmacy, Taipei Medical University, Taipei 110, Taiwan

^b School of Nutrition and Health Sciences, College of Nutrition, Taipei Medical University, Taipei 11031, Taiwan

^c International Master/Ph.D. Program in Medicine, College of Medicine, Taipei Medical University, Taipei 110, Taiwan

^d Professional Master Program in Artificial Intelligence in Medicine, College of Medicine, Taipei Medical University, Taipei 106, Taiwan

^e Research Center for Artificial Intelligence in Medicine, Taipei Medical University, Taipei 106, Taiwan

^f Translational Imaging Research Center, Taipei Medical University Hospital, Taipei 110, Taiwan

ARTICLE INFO

Article history:

Received 12 March 2022

Received in revised form 15 April 2022

Accepted 15 April 2022

Available online 19 April 2022

Keywords:

Explainable artificial intelligence

Drug-drug interaction

Machine learning

Deep learning

Chemical structures

Natural language processing

ABSTRACT

Over the past decade, polypharmacy instances have been common in multi-diseases treatment. However, unwanted drug-drug interactions (DDIs) that might cause unexpected adverse drug events (ADEs) in multiple regimens therapy remain a significant issue. Since artificial intelligence (AI) is ubiquitous today, many AI prediction models have been developed to predict DDIs to support clinicians in pharmacotherapy-related decisions. However, even though DDI prediction models have great potential for assisting physicians in polypharmacy decisions, there are still concerns regarding the reliability of AI models due to their black-box nature. Building AI models with explainable mechanisms can augment their transparency to address the above issue. Explainable AI (XAI) promotes safety and clarity by showing how decisions are made in AI models, especially in critical tasks like DDI predictions. In this review, a comprehensive overview of AI-based DDI prediction, including the publicly available source for AI-DDIs studies, the methods used in data manipulation and feature preprocessing, the XAI mechanisms to promote trust of AI, especially for critical tasks as DDIs prediction, the modeling methods, is provided. Limitations and the future directions of XAI in DDIs are also discussed.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Contents

| | |
|------------------------------------------------------------------------------|------|
| 1. Introduction | 2113 |
| 2. Study selection | 2113 |
| 3. Dataset, input data, and features for AI-DDIs studies | 2113 |
| 3.1. DDIs information retrieved from text-based sources | 2113 |
| 3.2. Molecule-based input data and feature preprocessing for DDIs prediction | 2114 |
| 4. Conventional ML-based prediction models of DDIs | 2116 |
| 4.1. Single ML algorithm-based predictive model | 2116 |
| 4.2. Ensemble learning predictive model | 2117 |
| 5. Deep learning-based prediction model of DDIs | 2118 |
| 5.1. Artificial neural network (ANN) | 2118 |
| 5.2. Convolutional neural network (CNN) | 2118 |
| 5.2.1. Conventional CNN | 2118 |
| 5.2.2. Dependency-based CNN | 2118 |
| 5.2.3. Deep CNN | 2118 |
| 5.3. Graph convolutional neural network (GCNN) | 2119 |

* Corresponding author at: Professional Master Program in Artificial Intelligence in Medicine, College of Medicine, Taipei Medical University, Taipei 106, Taiwan.

E-mail addresses: khanhlee87@gmail.com, khanhlee@tmu.edu.tw (N.Q.K. Le).

¹ These authors contributed equally to this work.

5.4. Recurrent neural network 2119

6. Interpretability methods in XAI and XAI in DDIs prediction 2119

7. Challenges and opportunities 2120

8. Conclusion 2120

CRediT authorship contribution statement 2120

Declaration of Competing Interest 2120

Acknowledgments 2120

References 2120

1. Introduction

Drug-drug interactions (DDIs) usually happen in polypharmacy instances when the effects of a drug alter that of others in a combined regimen. In treatment, preferably, synergistic action and therapeutic benefit are expected. However, in multi-diseases treatment, adverse drug events (ADEs) that cause toxicity or reduced treatment effect may also inevitably happen. These can eventually lead to increased morbidity and mortality in patients [1-3]. In addition, an increased number of recently frequent launches and approval of new drugs and indications in marketed medicines introduces more possible DDIs occurrences [4,5]. However, wet-lab experiments for verifying DDIs can drain researchers' time and resources and make it difficult for numerous and regular adoptions. Therefore, artificial intelligence (AI) models have been applied to predict DDIs [6-9]. These models have been continuously studied and improved along with the expansion and completeness of drug-database resources to support clinical decisions.

However, since the introduction of AI-models in DDIs recognition, many efforts have been applied to boost the predictive power of algorithms by putting forward more complex systems, turning these models into those called "black-box AI" that hinder the ability of users to explain how these models work [10]. Specifically, higher performance models are associated with more sophisticated systems, but lower performance tools with simple approaches are easier to comprehend [11]. Despite various benefits given by widespread industrial adoption of machine learning (ML) models, a critical domain as healthcare should be taken more seriously due to its immense value to humans. Additionally, from a human-oriented research angle, the ambiguity of complicated models in making predictive decisions hamper its successful adoption in medical settings as unable-to-interpreted systems are difficult to be trusted. Since the fundamental application of AI in drug treatment must first do with DDIs, explainable DDIs-AI models are pivotal for clinicians and patients to understand and trust their prediction. In response, the ignition of the field explainable artificial intelligence (XAI), which concentrates on methods to interpret ML models, has revived over recent years. XAI can facilitate clinical applications of DDIs prediction models regarding their requirement of robust yet human-understandable systems to provide clear justifications and promote safety, reliability, and transparency.

This review focuses on the advances of recently developed DDIs prediction models regarding their data manipulation technique, feature selection process, modeling approach, XAI method, and the challenge of assuring explainability and transparency of DDIs-prediction models without compromising the predictive power of these systems.

2. Study selection

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline was referenced when conducting literature reviewing [12]. We searched five electronic databases up to December 2021: Cochrane Library, PubMed, EMBASE,

IEEE, and Scopus. The search strategy combined the Medical Subject Headings terms and free terms "drug drug interaction" or "drug-drug interaction", in combination with "artificial intelligence" or "machine learning" or "deep learning" or "neural network" and "prediction model".

The eligibility criteria consisted of DDI predictive models that were built up using ML - and/or DL-based algorithms. The articles were screened and selected independently by two reviewers (N.T. K.N and H.T.V.), and disagreements were resolved by the third reviewer (N.Q.K.L.). All the retrieved publications were entered into reference-manager software (EndNote X9, Excel 2018).

We identified 643 records through Cochrane Library, IEEE, PubMed, EMBASE, Scopus database, and two records from reference lists of review paper. After removing 215 duplicates, 116 records were excluded according to the screening of titles and abstracts. Of 314 remaining research studies, 220 studies were removed after evaluating the selection criteria: (1) related to DDIs, (2) related to predictive model, (3) focused on ML or/and DL. As a result, we had 94 different research studies. Fig. 1 shows the flow diagram of the systematic search. Table 1 shows the detailed information of 94 selected studies.

The flowchart of AI-based DDI prediction model is illustrated in Fig. 2. From the whole flowchart, we would like to conduct our review based on two main aspects: input data (DDIs extraction and feature preprocessing) and AI algorithms (traditional machine learning and deep learning). The evolution of DDI prediction models separated by these two aspects is also shown in Fig. 3.

3. Dataset, input data, and features for AI-DDIs studies

In response to the growing number of pharmaceutical drugs entering the market over the past decades, many drug-related information databases have been updating and expanding to facilitate DDIs prediction [13-15]. Generally, most DDIs studies referred to datasets from DDIExtraction 2011 [16,17], DDIExtraction 2013 [18] and DrugBank database [19]. These public sources provide various types of drugs' characteristics and DDIs events to leverage AI approaches for DDIs discovery. The quantitative information about the DDIs is a necessary part of creating the described system. The data record format usually has binary characters encoded as 1 if there is an interaction between two drugs and 0 if there is a lack of known interaction.

Depending on the DDIs features-based view of different approaches, appropriate data extraction and feature preprocessing methods for DDIs prediction tasks can be applied.

3.1. DDIs information retrieved from text-based sources

This method involves extracting DDIs information in the form of biomedical text, especially in scientific literature since these sources represent valuable information for the retrieval of knowledge about the interaction between drugs. The amount of biomedical literature, which holds a vast amount of DDIs, has been growing over the past years and facilitating many DDIs extracting

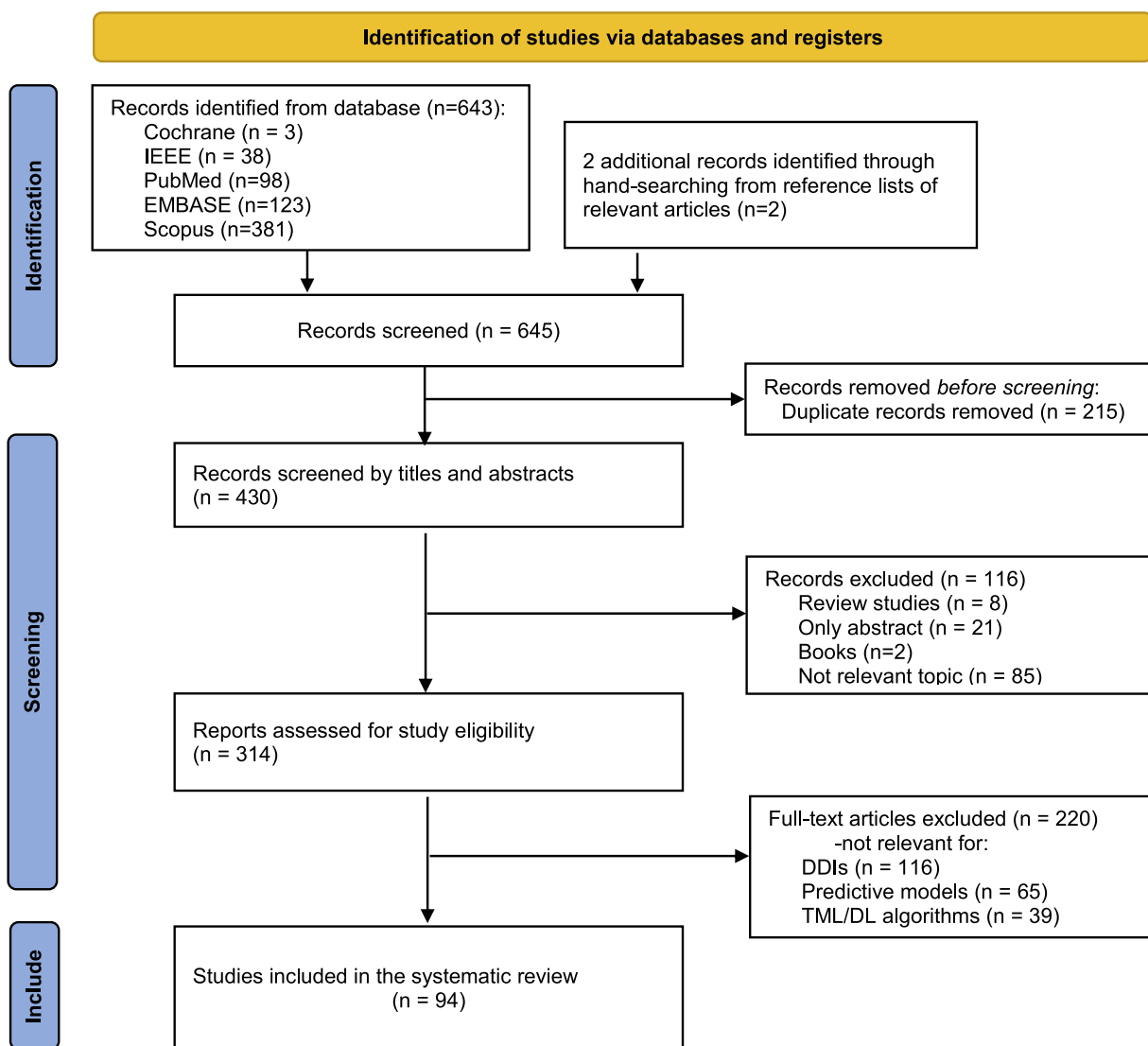


Fig. 1. PRISMA diagram showing our literature strategy search.

studies [20–22]. Aside from studies using public available DDI corpora [23,24], some studies have also used additional user-generated content to compensate for the limits of delayed updates of the medical database [25,26]. In addition, multi-information sources DDI corpora have been constructed based on useful information from FDA adverse event reports [27,28], electronic health records (EHRs) [29,30], or by following specific annotation guidelines [31] to construct corpus for DDIs extracting.

In these DDIs extraction approaches, feature preprocessing is essential. In detail, tokenization and lower casing are the first vital steps in reducing the sparsity of feature space. Also, many dimensionally reduction text preprocessing techniques have been used for DDIs extraction. Some compression techniques such as sentence pruning [32] and anaphora resolution have been applied [33]; Zhao used syntax word embedding strategy [34] instead of the common word embedding technique, some used Bidirectional Encoder Representations from Transformers (BERT) that relies on attention mechanism to capture high-quality contextual information [35,36]. The domain-specific ontologies approach attempted to use ancestors' sequences in the ontology to represent each entity [37]. Bokharaeian *et al.* [31] proposed clause dependency features to improve the relation extraction performance. Also, Ben Abacha *et al.* [38] used the CRF-based algorithm trained by a

set of linguistic and semantic features for the drug name recognition. Later, the DDIs extraction task was built on a hybrid method of both feature-based and kernel-based machine learning approaches. Moreover, the imbalanced class distribution problem has also been considered in many articles since this issue can diminish the power of classification [39,40]. Liu *et al.* used several rules to filter negative instances [41]; others added random negative sampling as part of the active learning algorithm to deal with the imbalanced issue [42] or use focal loss function to mitigate against this problem [43].

3.2. Molecule-based input data and feature preprocessing for DDIs prediction

Usually, DDIs studies utilize chemical, molecular, and pharmacological properties information to elucidate drug interactions insights. In detail, the chemical properties of drugs are typically described via the simplified molecular-input line-entry system (SMILES). This flexible chemical notation allows the generation of computer-feedable input [44]. These SMILES structural representations of drugs are post-processed to capture features of drug pairs associated with DDIs events [45]. Moreover, pharmacological properties such as targets [8,46], enzymes, transporters, genes and pro-

Table 1
Input data type of all papers reviewed in this study.

| No. | Method | Authors | Year | Input data | Algorithm | Performance |
|-----|--------|------------------------------|------|------------|-----------------------------------------------------------|----------------------------|
| 1 | TML | Cheng et al. [6] | 2014 | structure | SVM | AUC ~ 0.565 to 0.666 |
| 2 | | Hunta et al. [54] | 2017 | structure | SVM | AUC = 0.901 |
| 3 | | Deepika et al. [81] | 2018 | structure | meta classifier | F1-score = 0.909 |
| 4 | | Dhami et al. [51] | 2018 | structure | kernel learning | Accuracy > 0.7 |
| 5 | | Mahadevan et al. [48] | 2019 | structure | ensemble learning | Accuracy > 0.9 |
| 6 | | Zhang et al. [70] | 2019 | structure | ensemble learning | AUC = 0.9951 |
| 7 | | Song et al. [84] | 2019 | structure | SVM | AUC > 0.97 |
| 8 | | Qian et al. [60] | 2019 | structure | gradient boosting | AUC = 0.689 |
| 9 | | Wang et al. [85] | 2020 | structure | SVM | AUC = 0.985 |
| 10 | | Rohani et al. [79] | 2020 | structure | integrated similarity-constrained matrix factorization | F1-score = 0.885 |
| 11 | | Zhan et al. [92] | 2020 | structure | Bayesian networks coupled with level-wise algorithm | Precision = 0.5445 |
| 12 | | Huang et al. [141] | 2020 | structure | Chemical Sequential Pattern Mining | AUC = 0.91 |
| 13 | | Hung et al. [94] | 2021 | structure | ensemble learning | Accuracy = 0.7 |
| 14 | | Dang et al. [49] | 2021 | structure | XGBoost | F1-score = 0.65 |
| 15 | | Patrick et al. [72] | 2021 | structure | ensemble learning | AUC > 0.9 |
| 16 | | Dewulf et al. [142] | 2021 | structure | combined multi-regression | AUC = 0.843 |
| 17 | | Mei et al. [83] | 2021 | structure | L2-regularized logistic regression | AUC = 0.9884 |
| 18 | | Thomas et al. [17] | 2011 | text | ensemble learning | F1-score = 0.657 |
| 19 | | Minard et al. [143] | 2011 | text | SVM | F1-score = 0.5965 |
| 20 | | Garcia-Blasco et al. [16] | 2011 | text | RF | F1-score = 0.6341 |
| 21 | | Boyce et al. [87] | 2012 | text | SVM | F1-score = 0.859 |
| 22 | | Zhang et al. [89] | 2012 | text | single kernel | AUC = 0.924 |
| 23 | | Hailu et al. [19] | 2013 | text | SVM | F1-score = 0.5 |
| 24 | | Bjorne et al. [18] | 2013 | text | Turku Event Extraction System | F1-score = 0.59 |
| 25 | | Bobic et al. [95] | 2013 | text | LibLINEAR, perceptron Naïve Bayes | F1-score = 0.704 |
| 26 | | Yan et al. [73] | 2013 | text | Drug-Entity-Topic | AUC = 0.96 |
| 27 | | Zhang et al. [90] | 2015 | text | Label Propagation | AUC = 0.864 |
| 28 | | Ben Abacha A et al. [38] | 2015 | text | Hybrid CRF based | F1-score = 0.6398 |
| 29 | | Bokharaeian et al. [31] | 2016 | text | bag of word kernel | sign test p-value < 0.0001 |
| 30 | | Mahendran et al. [144] | 2016 | text | bag of word | F1-score = 0.769 |
| 31 | | Zhang et al. [28] | 2017 | text | ensemble learning | - |
| 32 | | Celebi et al. [75] | 2019 | text | RF | AUC = 0.91 |
| 33 | | Javed et al. [82] | 2021 | text | RF | Accuracy = 0.954 |
| 34 | | Xie et al. [42] | 2021 | text | LR | Precision = 0.9 |
| 35 | DL | Polak et al. [59] | 2005 | structure | ANN | AUC = 0.82 |
| 36 | | Herrero-Zazo et al. [53] | 2016 | structure | ANN | F1-score = 0.64 |
| 37 | | Ryu et al. [7] | 2018 | structure | DNN | Accuracy = 0.924 |
| 38 | | Lee et al. [55] | 2018 | structure | RWR coupled with KNN | AUC = 0.67 |
| 39 | | Karim et al. [145] | 2019 | structure | Graph Auto-Encoders | AUC = 0.98 |
| 40 | | Rohani et al. [77] | 2019 | structure | ANN | AUC from 0.954 to 0.994 |
| 41 | | Lee et al. [80] | 2019 | structure | auto-encoder coupled with a deep feed-forward network | Accuracy > 0.95 |
| 42 | | Hou et al. [45] | 2019 | structure | DNN | AUC = 0.942 |
| 43 | | Liu et al. [146] | 2019 | structure | multilayer bidirectional LSTM | F1-score = 0.7243 |
| 44 | | Karim et al. [66] | 2019 | structure | Convolutional-LSTM network | F1-score = 0.92 |
| 45 | | Shukla et al. [97] | 2019 | structure | convolutional mixture density RNN | Accuracy = 0.982 |
| 46 | | Deng et al. [50] | 2020 | structure | Multi DNN | F1-score = 0.7585 |
| 47 | | Lin et al. [68] | 2020 | structure | Knowledge Graph Neural Network | AUC = 0.9912 |
| 48 | | Zhang et al. [62] | 2020 | structure | multi-modal deep auto-encoders | F1-score = 0.8498 |
| 49 | | Feng et al. [52] | 2020 | structure | GCN-DNN | F1-score = 0.84 |
| 50 | | Shankar et al. [71] | 2020 | structure | ANN | AUC = 0.69 |
| 51 | | Masumshah et al. [102] | 2021 | structure | ANN | F1-score = 0.936 |
| 52 | | Zitnik et al. [74] | 2021 | structure | spectral convolution | AUC = 0.928 |
| 53 | | Lin et al. [56] | 2021 | structure | CNNs, auto-encoders with Siamese network | F1-score = 0.9117 |
| 54 | | Schwarz et al. [61] | 2021 | structure | multi-modal neural network | AUPRC from 0.77 to 0.92 |
| 55 | | Luo et al. [57] | 2021 | structure | graph convolutional auto-encoder network | - |
| 56 | | Nyamabo et al. [65] | 2021 | structure | graph neural network | AUC = 0.9838 |
| 57 | | Chen et al. [107] | 2021 | structure | integrated modules neural network | AUC = 0.9994 |
| 58 | | Pathak et al. [29] | 2013 | text | Linked Data | - |
| 59 | | Zhao et al. [34] | 2016 | text | Syntax CNN | F1-score = 0.686 |
| 60 | | Liu et al. [41] | 2016 | text | CNN | F1-score = 0.6975 |
| 61 | | Quan et al. [109] | 2016 | text | multichannel CNN | F1-score = 0.702 |
| 62 | | Zhang et al. [24] | 2016 | text | SVM | F1-score = 0.8497 |
| 63 | | Suárez-Paniagua et al. [105] | 2017 | text | CNN | F1-score = 0.6198 |
| 64 | | Zheng et al. [130] | 2017 | text | RNN with LSTM units | F1-score = 0.773 |
| 65 | | Kavuluru et al. [123] | 2017 | text | character-level RNNs | F1-score = 0.7081 |
| 66 | | Wang et al. [147] | 2017 | text | RNN with LSTM and an attention mechanism | F1-score = 0.715 |
| 67 | | Yi et al. [129] | 2017 | text | RNN | F1-score = 0.722 |
| 68 | | Jiang et al. [127] | 2017 | text | skeleton-LSTM | F1-score = 0.714 |
| 69 | | Li et al. [96] | 2017 | text | relation classification framework based on topic modeling | F1-score = 0.48 |
| 70 | | Wang et al. [120] | 2017 | text | LSTM | F1-score = 0.72 |
| 71 | | Zhang et al. [33] | 2017 | text | hierarchical RNN | F1-score = 0.729 |
| 72 | | Xu et al. [26] | 2018 | text | bidirectional LSTM network | F1-score = 0.7115 |
| 73 | | Sun et al. [112] | 2018 | text | Deep CNN | F1-score = 0.845 |
| 74 | | Lim et al. [21] | 2018 | text | recursive neural network | F1-score = 0.838 |

(continued on next page)

Table 1 (continued)

| No. | Method | Authors | Year | Input data | Algorithm | Performance |
|-----|--------|------------------------|------|------------|----------------------------------------------|--------------------|
| 75 | | Zhou et al. [126] | 2018 | text | BiLSTM | F1-score = 0.7299 |
| 76 | | Zhang et al. [20] | 2018 | text | RNN-CNN | F1-score = 0.648 |
| 77 | | Zitnik et al. [113] | 2018 | text | spectral convolution | AUC = 0.928 |
| 78 | | Paniagua et al. [104] | 2018 | text | CNN | F1-score = 0.6456 |
| 79 | | Hou et al. [100] | 2018 | text | LSTM- DNN | F1-score = 0.875 |
| 80 | | Sahu et al. [119] | 2018 | text | LSTM | F1-score = 0.6939 |
| 81 | | Zhang et al. [93] | 2019 | text | variational autoencoder | F1-score = 0.579 |
| 82 | | Xiong et al. [114] | 2019 | text | combined GCNN and BiLSTM | F1-score = 0.77 |
| 83 | | Liu et al. [146] | 2019 | text | non-linear unsupervised neural network + RF | F1-score = 0.8498 |
| 84 | | Sun et al. [43] | 2019 | text | recurrent hybrid CNN | F1-score = 0.7548 |
| 85 | | Shtar et al. [101] | 2019 | text | ensemble-based classifier | AUC 0.807 to 0.990 |
| 86 | | Xu et al. [25] | 2019 | text | full-attention network | F1-score = 0.712 |
| 87 | | Wu et al. [108] | 2020 | text | stacked bidirectional GRU + CNN | F1-score = 0.75 |
| 88 | | Zhu et al. [36] | 2020 | text | bidirectional transformer + BiGRU | F1-score = 0.809 |
| 89 | | Liu et al. [27] | 2020 | text | stacked autoencoders + weighted SVM | – |
| 90 | | Park et al. [32] | 2020 | text | Attention-based Graph Convolutional Networks | F1-score = 0.7686 |
| 91 | | Zaikis et al. [128] | 2020 | text | stacked Bi-LSTM + CNN | – |
| 92 | | Allahgholi et al. [23] | 2020 | text | ANN | Accuracy = 0.954 |
| 93 | | Warikoo et al. [35] | 2020 | text | Lexically-aware Transformer-based BERT | F1-score = 0.645 |
| 94 | | Fatehifar et al. [40] | 2021 | text | LSTM | F1-score = 0.783 |

TML: traditional machine learning, DL: deep learning, '-'the information was not reported in the original paper.

teins [6,47], interaction pathways like enzymes and transporters [48–61] can also be manipulated to represent drugs features through a set of descriptors. Network interaction mining [62–64] and molecular graph representations have also been used to describe substructures of drugs that come in distinctive shapes and sizes or the structural relations between entities [65–68]. Additionally, to overcome the lack of data overlap between chemical content and biological characteristics, the combined structure-based input that includes both chemical and biological data by hybridizing cheminformatics and bioinformatics techniques to link all chemical information and biological effects have also been applied to serve as a meaningful method for DDIs discovery in many studies [69–71].

Many techniques have also been applied to cover multi pharmacological facets of DDI by admitting heterogeneous characterizations from various data sources that represent different drug characteristics and physiological effects [72–74]. The knowledge graphs (KGs)-based features integrated from multiple sources such as DrugBank, PharmGKB, and KEGG drugs [75] were used to overcome the limited information issue in single-source methods. Along with this, some efforts have been made to address the problem of increased noise in the integrated similarity. The similarity selection heuristic process ranks matrices based on the entropy calculated in each matrix and calculates their pair-wise distance for the final selection based on redundancy minimization [76,77].

The classification feature constructing step usually requires the similarity analysis of paired drugs. In most studies, the chemical structural similarity was measured using the structures of the compound of drugs on DrugBank represented by their SMILES [6]. Structural representation of the drugs can be constructed using different molecular fingerprints generation techniques. The principle of this technique is to represent a molecule as a bit vector that codes the attendance or non-attendance of specifically assigned bit position structural features. Similarity measurements between molecular fingerprints are calculated using different methods; one commonly applied technique uses the Tanimoto coefficient [8,48,78]. Besides, many studies combine various drug-drug similarity measures representing relations between chemical, molecular physiological, or target pathways of drugs for the DDIs prediction task to gain more helpful information about DDIs [79,80]. On the other hand, the network-based features processing method exploits the topological properties of the DDI network.

Node2vec for Feature Network (FN) construction was used in [81] to present drug features as low-dimensional feature vectors.

4. Conventional ML-based prediction models of DDIs

Given the advanced computer science development and growing network pharmacology approaches, the development of a traditional ML-based model using multi-dimensional drug properties has been widely applied as a promising strategy to predict unknown DDIs [82,83].

4.1. Single ML algorithm-based predictive model

Support vector machine (SVM) was a common algorithm used to predict DDIs due to its high performance with a broad range AUC value of 0.565 – 0.985 [6,19,54,84–87]. Indeed, the number of recruiting features has a certain role in the predictive model, e.g., a study applied the features reducing method and achieved an increase of 0.02 in the F-measure score (0.5786 vs 0.5965) of the predictive model [86]. Kernel machines are a class of algorithms for pattern analysis whose best-known member is the SVM. Kernel classifiers were used for classifying the drug pairs, including all-paths graph (APG), k-band shortest path spectrum (kBSPS), and the shallow linguistic (SL) kernel [17,31,88,89]. Noteworthy, Thomas et al. [17] showed that SL and APG outperformed other methods, such as case-based reasoning and ensemble learning based on F1-score (0.606 vs. 0.416 and 0.583, respectively). Also, Zhang et al. [90] used the label propagation algorithms to work with the scenario where only a small portion of nodes in the undirected weighted network being labeled. In the meantime, logistic regression (LR) algorithm has been less used to establish DDIs prediction model. Xie et al. [91] integrated active learning, random negative sampling, and uncertainty sampling in clinical safety DDI information retrieval (DDI-IR) analysis using SVM and LR. In addition, Drug-Entity-Topic (DET) model following Bayes-rules was an example in leveraging augmented text-mining features to improve prediction performance in terms of discrimination and calibration [73]. Due to the growing demand for adverse DDIs (ADDIs) signal detection, Bayesian network framework and domain knowledge were combined to identify direct associations between a combination of medicines and the target ADEs [92]. Fur-

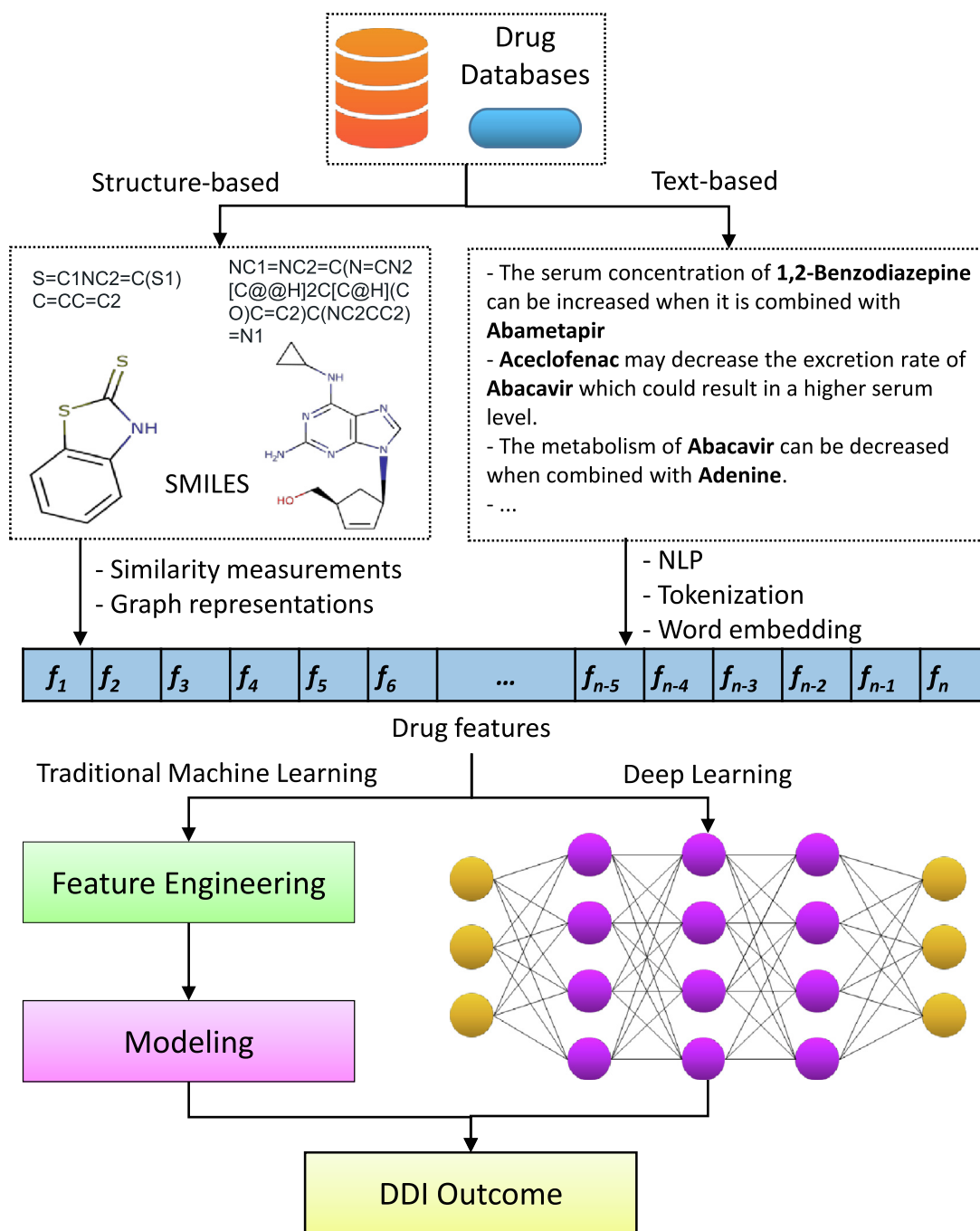


Fig. 2. Overall workflow of traditional ML and DL for DDIs prediction.

thermore, gradient boosting-based algorithm XGBoost was employed to achieve robust DDI prediction even for drugs whose interaction profiles were completely unseen during training [60]. XGBoost performed better or comparable to other algorithms, such as SVM, random forest, and the standard gradient boosting in terms of predictive performance and speed in DDIs prediction [49,60].

4.2. Ensemble learning predictive model

Ensemble methods use multiple learning algorithms to obtain better predictive performance than separate models in DDIs prediction [17,33,48,72,93,94]. Combined ML algorithms using LIBLINEAR, which consists of linear SVM, Naïve Bayes, and Voting

Perceptron classifiers, outperformed the original (unbalanced) train corpora model based on F-score (70.4% vs. 69.0%)[95]. Similarly, a heterogeneous network-assisted inference (HNAI) framework consisting of five different ML algorithms, including Naive Bayes (NB), decision tree (DT), k-nearest neighbors (k-NN), LR, and SVM, was proposed to detect the unknown DDIs with AUC of 0.67, higher than that of separated algorithms (NB:0.66, DT:0.565, k-NN:0.6, LR:0.655, and SVM:0.666) [6]. Other ensemble methods including genetic algorithm and LR in classifier ensemble rule for DDIs prediction could obtain AUC value up to 1 and accuracy >90%, regardless of approved and unproved drug pairs being selected [48]. One of the significant concerns for developing a high-accuracy DDIs prediction model is integrating heterogeneous drug features. Thus, Zhang et al. [62] proposed a multi-modal deep

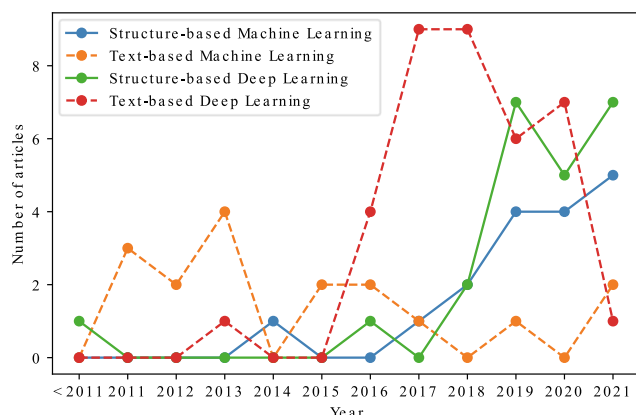


Fig. 3. Evolution of DDI prediction models separated by different input data and algorithms.

auto-encoders based drug representation learning method (DDI-MDAE) to predict DDIs from large-scale, noisy and sparse data. DDI-MDAE encompasses RF classifier in the positive-unlabeled learning setting. Another computational experiment established a sparse feature learning ensemble method with linear neighborhood regularization (SFLN) to predict DDIs, even unknown DDIs. Although SFLN presented high accuracy and outperformed benchmark methods, it costs a reasonable amount of running time [70].

5. Deep learning-based prediction model of DDIs

As many as the number of drugs have entered the market over the past decades, the deep and complex interactions between drugs can go far beyond the capacity of simple traditional ML algorithms [96]. Therefore, DL, with multiple processing layers-concepts, is applied in DDIs prediction due to its ability to deal with complex relations [97]. Inspired by the architecture of human brains [98], the superior performance of DL in classification tasks over conventional methods leverages its growing application in DDIs prediction. Unlike the traditional ML method, which depends on hand-crafted features engineering, DL performed the data representation and prediction in a joint task. In a complex, ill-defined, and highly nonlinear problem as DDIs prediction, DL emerges as a suitable approach for solving these stochastic issues. DL can be seen as representation learning, in which the machine, which involves multiple sequential layers, can develop its feature representations [99]. We devoted this section to describing all leading DL frameworks in the DDIs extraction and prediction tasks since DL entered the field.

5.1. Artificial neural network (ANN)

ANN is a data-driven algorithm that seeks hidden functional relations from the dataset. In ANN, many neurons are connected in complex interconnections to solve linear or nonlinear problems. Previous studies have successfully manipulated ANN models for DDIs prediction tasks [100,101]. The two layers ANN model has been used in the study of Rohani *et al.* [77] to work on a feature set of different similarity matrices collected from five different data sources. Masumshah *et al.* [102] used a feed-forward neural network with fully connected layers and the ReLU activation function was used between layers of the model as a sigmoid activation function for the output layer. Additionally, Shtar *et al.* [101] applied the ANN and propagation method over DDI graph nodes represented by an adjacency matrix. They used an XGBoost classifier for the DDIs classification, which output a binary value repre-

senting whether there is an interaction between the drug pairs or not.

5.2. Convolutional neural network (CNN)

CNN, which was inspired by the pattern of the animal Visual Cortex [103], has been introduced as an effective approach to deal with data with a grid pattern. The main goal of CNN is to transform the input into an easy-to-process form without compromising the prediction power. This characteristic makes CNN a potential candidate for the DDIs extraction task [104,105] that requires valuable feature learning aspects and massive datasets scalability. The central concept of CNNs utilizes hidden convolution and pooling layers to identify spatially localized features via a set of receptive fields in kernel form. Usually, a CNN architecture consists of convolution, pooling, and fully connected layers. According to the task, it is also essential to have a suitable activation function. For example, a sigmoid function is often used in binary classification, while the softmax function is often applied in multiclass classification [106]. Different forms of CNN have been proposed for DDI prediction as follows.

5.2.1. Conventional CNN

Chen *et al.* [107] used the CNN in the feature fusion module of their model, which was designed using a bi-level strategy with cross-and-scalar-level units. The CNN was used to learn the local and global features in the cross-level unit. The element-wise product was used in the scalar-level unit to get the fine-grained interactive feature between two features. These features will be concatenated to predict DDIs in the classifier module. The method proposed by Wu *et al.* [108] adopted two CNNs and the maximum pooling operation to extract features in the two location features from the word features preprocessed by the attention mechanism with a recurrent neural network (RNN). These features were then before fed into a softmax function to get the normalized probability score for each class. The model of Quan *et al.* [109] takes a DDIs instance represented by the word embedding and feeds them into the convolutional layer to get the filtered features. Then, the max-pooling layer extracts the essential local features; this layer also helps reduce the complexity of the model by reducing the feature dimension. Finally, in this model, a softmax layer is used for classifying DDIs types.

5.2.2. Dependency-based CNN

The process of feeding local information into convolution operation in traditional CNN is not practical considering the case of long-distance relationships between words in candidate DDIs instances. Attempts to enlarge the window can lead to the data sparsity problem. Therefore, the dependency-based convolutional model (Dep-CNN) has been applied to capture long-distance dependencies between words of a sentence and extract DDIs from candidate instances. Dep-CNN performs convolution operation on adjacent words in word sentences and dependency parsing trees of candidate DDIs instances. In the model proposed by Liu *et al.* [110], they first generate a dependency parsing tree where each node corresponds to a word in the instance and syntactic dependency between two words denoted by the directed edge. Their Dep-CNN model is a four-layer neural network, consisting of a look-up table layer, a convolutional layer, a max-pooling layer, and a softmax progressing layer to feed the feature vector to a fully connected neural network for classification.

5.2.3. Deep CNN

Considering various properties in texts, the successful application of Deep CNN (DCNN) in identifying complex patterns of image and video in computer vision [111] suggested its application in

DDIs extraction task. Sun *et al.* [112] proposed a DCNN model which utilized a small convolution architecture to operate directly at the word level of the raw biomedical text input to get the embedding-based convolutional features. Then, the softmax classifier will be used to operate these features and extract DDIs from biomedical literature.

5.3. Graph convolutional neural network (GCNN)

In many DDIs prediction approaches, the molecular structure of drugs has been extensively exploited to extract the characteristics of the drug that link to the DDIs events. In non-Euclidean domains, where complex relationships and interdependencies between molecular structure representation of drugs or interactions between drug targets betokened as graphs [113], the application of GCNN in DDIs prediction was introduced. The most fundamental part of a GCNN is a graph, a data structure consisting of two components: nodes and edges [101]. The nodes usually represent the drug and edges are associated with interactions between nodes [114]. The first graph convolutional network was proposed by Bruna *et al.* [115] for applying neural networks to graph-structured data. Also, a model called SC-DDIS was introduced by Liu *et al.* [74] can learn the final embedding of drugs via a graph spectral CNN. Besides, it deals with the multiple complex structured entities that consist of two graph types: local graph for structured entities and global graph to capture structured entities' interactions. Wang *et al.* [85] proposed a graph to GCNN model called GoGNN to extract features in both graphs in a hierarchical fashion to leverage the DDIs prediction performance.

5.4. Recurrent neural network

RNN is highly manipulated in NLP [116,117] and it mainly deals with sequential data. What makes RNNs differ from CNNs is their memory mechanism that gets information for the prior inputs to influence the current input and output. The DDIs extraction task is considered a relation extraction task in NLP. Many have utilized the long short term memory (LSTM) network to extract DDIs from literature [118–120]. Even though Char-RNNs are more common for modeling morphologically richer languages [121] and were introduced for text classification [122]. Kavuluru *et al.* [123] has also considered the role of character-level embedding in DDIs extraction, and they used an LSTM on the character embedding to extract the word vectors.

Luo *et al.* [57] presented a model that used an LSTM model for DDIs prediction in diabetes using the embedded drug-induced transcriptome data. The LSTM is a typical RNN architecture introduced by Hochreiter and Schmidhuber [124] to deal with the problem of long-term dependencies. In LSTM, cells in the hidden layers contain an input gate, an output gate, and a forget gate to control the flow of information required for the Prediction. Also, the gated recurrent units (GRU) was introduced to address the short-term memory problem of the RNNs model [125]. However, unlike the LSTM, GRUs use hidden states and two gates: reset and update gate to control the information to retain for the prediction.

For the DDIs extraction task, a hierarchical RNN was introduced by Zhang *et al.* [33]. This model framework considers the shortest dependency path (SDP) between two entities and uses the RNN to learn the feature representation of sentence sequence and SDP for extracting DDIs. Zhou *et al.* [126] introduced an attention-based BiLSTM model to encode biomedical text sentences.

Besides, considering the difference between DDIs instance and typical sentence, Jiang *et al.* [127] used a skeleton structure to represent the DDIs instances and the LSTM model to work with the structure (skeleton-LSTM). In their framework, a sentence is first tokenized into token units followed by a corresponding skeleton

unit, distance to the first drug, and distance to the second drug. These units are input to the embedding layer of the skeleton-LSTM.

However, traditional Encoder-Decoder architecture using RNN or LSTM remained several drawbacks as it can cause the information loss problem, especially in the case of long sentences. Attention mechanism has been applied to deal with the problem mentioned above [128]. The model proposed by Yi *et al.* [129] used a bidirectional RNN layer to generate a sentence matrix as the word's semantic representation. Then, the attention layer is applied to create the final representation by combining several relevant sentences of the same drug pairs. The softmax classifier was used to classify specific DDIs. Zheng *et al.* [130] also introduced a model to classify DDIs from texts using a combined attention mechanism and an RNN with LSTM units.

6. Interpretability methods in XAI and XAI in DDIs prediction

The surge in the predictive performance of AI tools is achieved by increasing model complexity. This turns these models into black-box systems and causes uncertainty regarding their operation mechanism. This ambiguity hinders the wide adaptation of AI models in critical domains like healthcare. As a result, explainable Artificial Intelligence (XAI) focuses on understanding behind the prediction of AI models to accommodate the demand for transparency in AI tools. Interpretability methods of AI models can be classified based on the type of algorithms, the interpretation scale, and the data type [131]. Additionally, based on the purposes of interpretability, approaches can be categorized as white-box models creation, black-box models explanation, enhancement of model fairness and predictive sensitivity testing [132].

In terms of methods to explain DL models, the gradient-based attribution method [133] attempts to explain the prediction by attributing them to the network's input features. This method is often applied when predictions are made from a DNN system and therefore, can be potential approach for some black-box DNN models in DDIs prediction like [110,112]. Moreover, the DeepLIFT is a popular algorithm applied on top of DNN models that showed considerable advantages compared to gradient-based methods [134]. On the other hand, Guided BackPropagation method can be applied to network structures [135]. Under this, a convolutional layer with improved stride can replace max-pooling in CNN to deal with accuracy loss. This approach suggests a potential application in some CNN-based DDIs prediction such as [111]. On top of this, the [136] was proposed in NLP-based neural networks. This method used rationales (small pieces of input text) and tried to produce the same prediction as the full-text input type. Under this method, the architecture consists of two components, generator and encoder, to look for text subsets highly related to the prediction result. Since the DDIs extraction task is conducted via NLP-based models [109,114], the above methods should be considered for application to promote the clarity of these models.

Apart from this, methods to create white-box models such as linear, decision tree, rule-based models, or sophisticated yet transparent models have also been proposed in XAI. However, due to the limited predictive power, especially in the NLP-based domain as in the DDIs extraction task, these approaches are given less interest. Additionally, various methods have been proposed to tackle fairness in AI. Nevertheless, a minimal number of these scientific pieces of literature considered fairness in non-tabular data such as text-based information for DDIs extraction. While many DDIs studies applied the word embedding method [62,109], it was revealed that vectorized representing of text data could carry strong bias [137]. Therefore, methods to assure fairness should be taken into more consideration in DDIs studies. Furthermore,

some methods aim to analyze the sensitivity of AI models to ensure the reliability of those tools. In the Adversarial Example-based Sensitivity Analysis, Zugner *et al.* [138] used this approach to study the graph-structured data. This method considers modifying node connections or node features to attack node classification models. Since graph-based methods are widely applied in DDIs studies [67,68], approaches as in the above research suggest potential application in DDIs prediction model. Also, using perturbations to the word embeddings [139] in RNN should also be considered. Significantly, the input reduction method in the study of Feng *et al.* [140] to reveal oversensitivity in NLP models can be a possible approach in DDIs extracting studies. Literature regarding the explosion of the weakness of DL models in NLP-tasks is complete; however, applications in DDIs- NLP models are still limited.

In the DDIs study of Schwarz *et al.* [61], an attempt has been made to offer their model interpretability using the Attention scores computed at all layers of modeling. Using these scores, the contribution of the similarity matrices to the drug representation vectors is determined and the drug characteristics that lead to better encoding are selected. This approach leverages information that passes through all layers of the network.

7. Challenges and opportunities

Though traditional ML performed effectively in extracting DDIs, even from the unstructured package insert (aka drug product label) [87], conventional ML-based methods still have several drawbacks. ML-based models are learned from positive and negative data, making it difficult in real-world domains due to the lack of true negative DDIs or a “gold standard” non-DDI. Therefore, it is necessary to identify positive data from many unlabeled data containing positive and negative samples and avoid biased sampling by random negative sampling and validation set updating. Additionally, it is unknown whether there is DDI between two drugs in a negative class dataset because some new DDIs drug pairs may not be reported yet. Another issue is different types of DDI data, such as clinical drug safety and pharmacokinetic data with different targeted samples and proportions in DDI-relevant databases or articles. Also, it is more time-consuming to accomplish the annotated corpora and determine optimal parameters in traditional ML-based methods. Hence, DNN models, including CNN and sequential neural networks such as RNN, have been referred to as an optimal resolution for feature selection and DDIs extraction without complicated feature engineering [120]. However, we assumed that several paths should be investigated in future work. First, drug-related textual data sources, such as patent information, are essential. Second, it is unknown how to use drug domain knowledge or semi-structured drugs, such as paragraph that describes the pharmacodynamics or mechanism of action, protein binding, or experimental properties of a drug in building up predictive models.

In addition, DL with superior performance and capability to automatically generate hierarchical input for the classification tasks has gained huge research attention in DDIs prediction domain. Still, these DL methods are neither easily explainable nor commonly trusted by medical staff because of their explainability deficiency. In the DDIs prediction field, only a few studies have considered the explainable aspect of their models, which leaves plenty of room to improve, innovate, and ensure predictive performance and model interpretability in ML-based DDIs prediction models. We, therefore, think that either approaches to explain black-box models, methods to create high-accuracy white-box models, strategies to ensure models fairness, or strict sensitivity analyses of models in DDIs prediction should be given more consideration in the coming years to produce trust and fairness in

these models' performance and bring them closer to clinical application. Since XAI aims to explain the machine learning models, its application does not lead to less accuracy in current models. Also, further studies can show the potential of XAI in sacrificing accuracy in the field of DDIs extraction task (NLP) if text based approach is usually used for replenishment of databases and one can refine the found dependencies in the initial sources. Addressing it may open a new road in the application of XAI in DDI prediction in the future, especially for DDI extraction task using NLP.

8. Conclusion

The management of DDIs, which can cause ADEs and affect patients' health, plays a crucial role in pharmacovigilance and medical practice. The main contribution of this study is the establishment of detailed taxonomy of existing models for predicting DDIs. Given remarkable breakthroughs in DDIs prediction over the past years, weakness in terms of model interpretability exposed considerable limits. We, therefore, believe that XAI in DDIs prediction still holds many potential aspects to unlock in future studies.

CRedit authorship contribution statement

Thanh Hoa Vo: Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Ngan Thi Kim Nguyen:** Methodology, Formal analysis, Validation, Writing – original draft, Writing – review & editing, Visualization. **Quang Hien Kha:** Validation, Data curation. **Nguyen Quoc Khanh Le:** Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Ministry of Science and Technology, Taiwan [grant number MOST110-2221-E-038-001-MY2].

References

- [1] Askari M *et al.* Frequency and nature of drug-drug interactions in the intensive care unit. *Pharmacoepidemiol Drug Saf* 2013;22(4):430–7.
- [2] Raschetti R *et al.* Suspected adverse drug events requiring emergency department visits or hospital admissions. *Eur J Clin Pharmacol* 1999;54(12):959–63.
- [3] Budnitz DS *et al.* National surveillance of emergency department visits for outpatient adverse drug events. *JAMA* 2006;296(15):1858–66.
- [4] Reis AM, Cassiani SH. Evaluation of three brands of drug interaction software for use in intensive care units. *Pharm World Sci* 2010;32(6):822–8.
- [5] Vonbach P *et al.* Evaluation of frequently used drug interaction screening programs. *Pharm World Sci* 2008;30(4):367–74.
- [6] Cheng F, Zhao Z. Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. *J Am Med Inform Assoc* 2014;21(e2):e278–86.
- [7] Ryu JY, Kim HU, Lee SY. Deep learning improves prediction of drug–drug and drug–food interactions. *Proc Natl Acad Sci U S A* 2018;115(18):E4304.
- [8] Vilar S *et al.* Similarity-based modeling in large-scale prediction of drug–drug interactions. *Nat Protoc* 2014;9(9):2147–63.
- [9] Vilar S, Uriarte E, Santana L, Tatonetti NP, Friedman C. Detection of drug–drug interactions by modeling interaction profile fingerprints. *PLoS ONE* 2013;8(3):e58321.
- [10] Gunning, D., *et al.*, *XAI—Explainable artificial intelligence*. *Science Robotics*, 2019. 4(37): p. eaay7120.

- [11] David Gunning, M.S., Jaesik Choi, Timothy Miller, Simone Stumpf and Guang-Zhong Yang, *XAI—Explainable artificial intelligence*. Sci. Robotics, 2019. **eaay7120**.
- [12] Page MJ et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2020;2021:372.
- [13] Wishart DS et al. Drugbank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* 2006;1(34) (D668-72. 16381955.).
- [14] Whirl-Carrillo M et al. An evidence-based framework for evaluating pharmacogenomics knowledge for personalized medicine. *Clin Pharmacol Ther* 2021.
- [15] Kanehisa M et al. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 2015;44(D1):D457–62.
- [16] García Blasco, S., et al. *Automatic drug-drug interaction detection: A machine learning approach with maximal frequent sequence extraction*. in *CEUR Workshop Proceedings*. 2011. CEUR Workshop Proceedings.
- [17] Thomas, P., et al., *Relation extraction for drug–drug interactions using ensemble learning*. 1st Challenge task on Drug-Drug Interaction Extraction (DDIExtraction 2011), 2011: p. 11–18.
- [18] Björne, J., S. Kaewphan, and T. Salakoski. *UTurku: drug named entity recognition and drug–drug interaction extraction using SVM classification and domain knowledge*. in *Second Joint Conference on Lexical and Computational Semantics (* SEM)*, Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013). 2013.
- [19] Hailu, N., L. Hunter, and K.B. Cohen. *UColorado_SOM: extraction of drug–drug interactions from biomedical text using knowledge-rich and knowledge-poor features*. in *Second Joint Conference on Lexical and Computational Semantics (* SEM)*, Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013). 2013.
- [20] Zhang Y et al. A hybrid model based on neural networks for biomedical relation extraction. *J Biomed Inform* 2018;81:83–92.
- [21] Lim S, Lee K, Kang J. Drug drug interaction extraction from the literature using a recursive neural network. *PLoS ONE* 2018;13(1):e0190926.
- [22] Liu J et al. Drug-Drug Interaction Extraction Based on Transfer Weight Matrix and Memory Network. *IEEE Access* 2019;7:101260–8.
- [23] Allahgholi M et al. ADDI: Recommending alternatives for drug–drug interactions with negative health effects. *Comput Biol Med* 2020;125:103969.
- [24] Zhang, Y., et al., *Extracting drug–enzyme relation from literature as evidence for drug drug interaction*. *J Biomed Semantics*, 2016. **7**: p. 11–11.
- [25] Xu B et al. Incorporating User Generated Content for Drug Drug Interaction Extraction Based on Full Attention Mechanism. *IEEE Trans Nanobioscience* 2019;18(3):360–7.
- [26] Xu B et al. Full-attention Based Drug Drug Interaction Extraction Exploiting User-generated Content. 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2018.
- [27] Liu N, Chen CB, Kumara S. Semi-Supervised Learning Algorithm for Identifying High-Priority Drug–Drug Interactions Through Adverse Event Reports. *IEEE J Biomed Health Inform* 2020;24(1):57–68.
- [28] Zhang W et al. Predicting potential drug–drug interactions by integrating chemical, biological, phenotypic and network data. *BMC Bioinf* 2017;18(1):18.
- [29] Pathak J, Kiefer RC, Chute CG. Using linked data for mining drug–drug interactions in electronic health records. *Stud Health Technol Inform* 2013;192:682–6.
- [30] Duke JD et al. Literature based drug interaction prediction with clinical assessment using electronic medical records: novel myopathy associated drug interactions. *PLoS Comput Biol* 2012;8(8):e1002614.
- [31] Bokharaeian B, Diaz A, Chitsaz H. Enhancing extraction of drug–drug interaction from literature using neutral candidates, negation, and clause dependency. *PLoS ONE* 2016;11(10):e0163480.
- [32] Park C, Park J, Park S. AGCN: Attention-based graph convolutional networks for drug–drug interaction extraction. *Expert Syst Appl* 2020;159:113538.
- [33] Zhang Y et al. Drug–drug interaction extraction via hierarchical RNNs on sequence and shortest dependency paths. *Bioinformatics* 2017;34(5):828–35.
- [34] Zhao Z et al. Drug drug interaction extraction from biomedical literature using syntax convolutional neural network. *Bioinformatics* 2016;32(22):3444–53.
- [35] Warikoo N, Chang Y-C, Hsu W-L. LBERT: Lexically aware Transformer-based Bidirectional Encoder Representation model for learning universal bio-entity relations. *Bioinformatics* 2020;37(3):404–12.
- [36] Zhu Y et al. Extracting drug–drug interactions from texts with BioBERT and multiple entity-aware attentions. *J Biomed Inform* 2020;106:103451.
- [37] Lamurias A et al. BO-LSTM: classifying relations via long short-term memory networks along biomedical ontologies. *BMC Bioinf* 2019;20(1):10.
- [38] Abacha AB et al. Text mining for pharmacovigilance: Using machine learning for drug name recognition and drug–drug interaction extraction and classification. *J Biomed Inform* 2015;58:122–32.
- [39] Chowdhury MFM, Lavelli A. Impact of less skewed distributions on efficiency and effectiveness of biomedical relation extraction. *Proceedings of COLING 2012: Posters*, 2012.
- [40] Fatehifar M, Karshenas H. Drug–Drug interaction extraction using a position and similarity fusion-based attention mechanism. *J Biomed Inform* 2021;115:103707.
- [41] Liu S et al. Drug-Drug Interaction Extraction via Convolutional Neural Networks. *Comput Math Methods Med* 2016;2016:6918381.
- [42] Xie W et al. Integrated Random Negative Sampling and Uncertainty Sampling in Active Learning Improve Clinical Drug Safety Drug-Drug Interaction Information Retrieval. *Front Pharmacol* 2021;11(2225).
- [43] Sun X et al. Drug-Drug Interaction Extraction via Recurrent Hybrid Convolutional Neural Networks with an Improved Focal Loss. *Entropy (Basel)* 2019;21(1).
- [44] Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci* 1988;28(1):31–6.
- [45] Hou, X.a.Y., Jiaying and Hu, Pingzhao, Predicting Drug-Drug Interactions Using Deep Neural Network, in *Proceedings of the 2019 11th International Conference on Machine Learning and Computing*, 2019, Association for Computing Machinery: New York, NY, USA, p. 168–172.
- [46] Zhao XM et al. Prediction of drug combinations by integrating molecular and pharmacological data. *PLoS Comput Biol* 2011;7(12):e1002323.
- [47] Luo, H., et al., *DDI-CPI, a server that predicts drug–drug interactions through implementing the chemical–protein interactome*. *Nucleic Acids Res*, 2014. **42** (Web Server issue): p. W46–52.
- [48] Mahadevan AA et al. A Predictive Model for Drug-Drug Interaction Using a Similarity Measure. 2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). IEEE; 2019.
- [49] Dang LH et al. Machine Learning-Based Prediction of Drug-Drug Interactions for Histamine Antagonist Using Hybrid Chemical Features. *Cells* 2021;10(11):3092.
- [50] Deng Y et al. A multimodal deep learning framework for predicting drug–drug interaction events. *Bioinformatics* 2020;36(15):4316–22.
- [51] Dhami DS et al. Drug-Drug Interaction Discovery: Kernel Learning from Heterogeneous Similarities. *Smart Health* 2018;9–10:88–100.
- [52] Feng Y-H, Zhang S-W, Shi J-Y. DPDDI: a deep predictor for drug–drug interactions. *BMC Bioinf* 2020;21(1):419.
- [53] Herrero-Zazo M, Lille M, Barlow DJ. Application of Machine Learning in Knowledge Discovery for Pharmaceutical Drug–drug Interactions. *KDWeb* 2016.
- [54] Hunta S, Aunsri N, Yooyatvong T. Integrated action crossing method for Drug-Drug Interactions prediction in noncommunicable diseases based on neural networks. 2017 International Conference on Digital Arts, Media and Technology (ICDAMT), 2017.
- [55] Lee I, Nam H. Identification of drug–target interaction by a random walk with restart method on an interactome network. *BMC Bioinf* 2018;19(8):208.
- [56] Lin S et al. MDF-SA-DDI: predicting drug–drug interaction events based on multi-source drug fusion, multi-source feature fusion and transformer self-attention mechanism. *Brief Bioinform* 2021.
- [57] Luo Q et al. Novel deep learning-based transcriptome data analysis for drug–drug interaction prediction with an application in diabetes. *BMC Bioinf* 2021;22(1):318.
- [58] Olha Marushchak RK. Designing of Information Model for Prediction of Drug–Drug Interactions based on Calculation of Target and Therapeutic Similarity. 3rd International Conference on Informatics & Data-Driven Medicine. 2020. Sweden: CEUR Workshop Proceedings, 2020.
- [59] Polak, S.a.B., J. and Mendyk, A, *Neural System for in silico Drug-Drug Interaction Screening, in Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*. 2005. p. 75–80.
- [60] Qian S, Liang S, Yu H. Leveraging genetic interactions for adverse drug–drug interaction prediction. *PLoS Comput Biol* 2019;15(5):e1007068.
- [61] Schwarz K et al. AttentionDDI: Siamese attention-based deep learning method for drug–drug interaction predictions. *BMC Bioinf* 2021;22(1):412.
- [62] Zhang Y et al. Predicting drug–drug interactions using multi-modal deep auto-encoders based network embedding and positive-unlabeled learning. *Methods* 2020;179:37–46.
- [63] Udrescu ML, Udrescu A. Drug Repurposing Method Based on Drug-Drug Interaction Networks and Using Energy Model Layouts. In: Vanhaelen Q, editor. *Computational Methods for Drug Repurposing*. New York, NY: Springer New York; 2019. p. 185–201.
- [64] Takarab M et al. Network-based analysis and characterization of adverse drug–drug interactions. *J Chem Inf Model* 2011;51(11):2977–85.
- [65] Nyamabo AK, Yu H, Shi J-Y. SSI-DDI: substructure–substructure interactions for drug–drug interaction prediction. *Brief Bioinform* 2021;22(6).
- [66] Decker, M.R.K.a.M.C.a.J.B.J.a.M.U.a.O.B.a.S., *Drug-Drug Interaction Prediction Based on Knowledge Graph Embeddings and Convolutional-LSTM Network*, in *10th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, ACM-BCB*. 2019, Association for Computing Machinery, Inc: Niagara Falls, United States.
- [67] Sun M, Wang F, Elemento O, Zhou J. Structure-Based Drug-Drug Interaction Detection via Expressive Graph Convolutional Networks and Deep Sets. *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI Press; 2020.
- [68] Xuan Lin ZQ, Wang Z-J, Ma T, Xiangxiang Zeng KGNN. Knowledge Graph Neural Network for Drug-Drug Interaction Prediction. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 2020.
- [69] Bo Peng, X.N. *Deep Learning for High-Order Drug-Drug Interaction Prediction*. in *In 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (ACM-BCB '19)*. 2019, NY,USA.
- [70] Zhang W et al. SFLN: A sparse feature learning ensemble method with linear neighborhood regularization for predicting drug–drug interactions. *Inf Sci* 2019;497:189–201.

- [71] Shankar S et al. Predicting adverse drug reactions of two-drug combinations using structural and transcriptomic drug representations to train an artificial neural network. *Chem Biol Drug Des* 2021;97(3):665–73.
- [72] Patrick MT et al. Advancement in predicting interactions between drugs used to treat psoriasis and its comorbidities by integrating molecular and clinical resources. *J Am Med Inform Assoc* 2021;28(6):1159–67.
- [73] Yan S, Jiang X, Chen Y. Text Mining Driven Drug-Drug Interaction Detection. In: *Proceedings. IEEE International Conference on Bioinformatics and Biomedicine*. p. 349–55.
- [74] Liu T et al. Modeling polypharmacy effects with heterogeneous signed graph convolutional networks. *Appl Intell* 2021;51.
- [75] Celebi R et al. Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings. *BMC Bioinf* 2019;20(1):726.
- [76] Olayan RS, Ashoor H, Bajic VB. DDR: efficient computational method to predict drug-target interactions using graph mining and machine learning approaches. *Bioinformatics* 2018;34(7):1164–73.
- [77] Rohani N, Eslahchi C. Drug-Drug Interaction Predicting by Neural Network Using Integrated Similarity. *Sci Rep* 2019;9(1):13645.
- [78] Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J Cheminform* 2015;7(1):20.
- [79] Rohani N, Eslahchi C, Katanforoush A. ISCMF: Integrated similarity-constrained matrix factorization for drug-drug interaction prediction. *Network Modeling Analysis in Health Informatics and Bioinformatics* 2020;9(1):11.
- [80] Lee G, Park C, Ahn J. Novel deep learning model for more accurate prediction of drug-drug interaction effects. *BMC Bioinf* 2019;20(1):415.
- [81] Deepika SS, Geetha TV. A meta-learning framework using representation learning to predict drug-drug interaction. *J Biomed Inform* 2018;84:136–47.
- [82] Javed R et al. An Efficient Pattern Recognition Based Method for Drug-Drug Interaction Diagnosis. 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA), 2021.
- [83] Mei S, Zhang K. A machine learning framework for predicting drug-drug interactions. *Sci Rep* 2021;11(1):17619.
- [84] Song D et al. Similarity-based machine learning support vector machine predictor of drug-drug interactions with improved accuracies. *J Clin Pharm Ther* 2019;44(2):268–75.
- [85] Wang, H., et al. GoGNN: graph of graphs neural network for predicting structured entity interactions, in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. 2021: Yokohama, Yokohama, Japan. p. Article 183.
- [86] Minard A-L et al. Feature selection for drug-drug interaction detection using machine-learning based approaches. *Challenge Task on Drug-Drug Interaction Extraction (DDI) SEPLN*, 2011.
- [87] Boyce R, Gardner G, Harkema H. Using natural language processing to identify pharmacokinetic drug-drug interactions described in drug package inserts. In: *Proceedings of the 2012 Workshop on Biomedical Natural Language Processing*. Montreal, Canada: Association for Computational Linguistics; 2012. p. 206–13.
- [88] Dhami DS et al. Drug-drug interaction discovery: kernel learning from heterogeneous similarities. *Smart Health* 2018;9:88–100.
- [89] Zhang Y et al. A Single Kernel-Based Approach to Extract Drug-Drug Interactions from Biomedical Literature. *PLoS ONE* 2012;7(11):e48901.
- [90] Zhang P, Wang F, Hu J, Sorrentino R. Label Propagation Prediction of Drug-Drug Interactions Based on Clinical Side Effects. *Sci Rep* 2015;5:12339.
- [91] Xie W et al. Integrated Random Negative Sampling and Uncertainty Sampling in Active Learning Improve Clinical Drug Safety Drug-Drug Interaction Information Retrieval. *Front Pharmacol* 2021;11:2225.
- [92] Zhan C et al. Detecting high-quality signals of adverse drug-drug interactions from spontaneous reporting data. *J Biomed Inform* 2020;112:103603.
- [93] Zhang Y, Lu Z. Exploring Semi-supervised Variational Autoencoders for Biomedical Relation Extraction. *Methods* 2019;166.
- [94] Hung TNK et al. An AI-based Prediction Model for Drug-drug Interactions in Osteoporosis and Paget's Diseases from SMILES. *Mol Inform* 2022;2100264.
- [95] Bobić, T., J. Fluck, and M. Hofmann. SCAL: Extracting drug-drug interactions using a rich feature vector. in *Second Joint Conference on Lexical and Computational Semantics (* SEM)*, Volume 2: *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*. 2013.
- [96] Li D et al. A Topic-modeling Based Framework for Drug-drug Interaction Classification from Biomedical Text. *AMIA Annu Symp Proc* 2016;2016:789–98.
- [97] Kumar Shukla P et al. Efficient prediction of drug-drug interaction using deep learning models. *IET Syst Biol* 2020;14(4):211–6.
- [98] Sejnowski TJ. The unreasonable effectiveness of deep learning in artificial intelligence. *Proc Natl Acad Sci* 2020;117(48):30033–8.
- [99] Esteva A et al. A guide to deep learning in healthcare. *Nat Med* 2019;25(1):24–9.
- [100] Hou WJ, Ceasay B. Extraction of drug-drug interaction using neural embedding. *J Bioinform Comput Biol* 2018;16(6):1840027.
- [101] Shtar G, Rokach L, Shapira B. Detecting drug-drug interactions using artificial neural networks and classic graph similarity measures. *PLoS ONE* 2019;14(8):e0219796.
- [102] Masumshah R, Aghdam R, Eslahchi C. A neural network-based method for polypharmacy side effects prediction. *BMC Bioinf* 2021;22(1):385.
- [103] Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern* 1980;36(4):193–202.
- [104] Suárez-Paniagua V, Segura-Bedmar I. Evaluation of pooling operations in convolutional architectures for drug-drug interaction extraction. *BMC Bioinf* 2018;19(8):209.
- [105] Suárez-Paniagua V, Segura-Bedmar I, Martínez P. Exploring convolutional neural networks for drug-drug interaction extraction. *Database (Oxford)* 2017.
- [106] Lapin M et al. Analysis and Optimization of Loss Functions for Multiclass, Top-k, and Multilabel Classification. *IEEE Trans Pattern Anal Mach Intell* 2018;40(7):1533–54.
- [107] Chen Y et al. MUFFIN: multi-scale feature fusion for drug-drug interaction prediction. *Bioinformatics* 2021;37(17):2651–8.
- [108] Wu H et al. Drug-drug interaction extraction via hybrid neural networks on biomedical literature. *J Biomed Inform* 2020;106:103432.
- [109] Quan C et al. Multichannel Convolutional Neural Network for Biological Relation Extraction. *Biomed Res Int* 2016;2016:1850404.
- [110] Liu S et al. Dependency-based convolutional neural network for drug-drug interaction extraction. 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2016.
- [111] Zeng T et al. Deep convolutional neural networks for annotating gene expression patterns in the mouse brain. *BMC Bioinf* 2015;16(1):147.
- [112] Sun X et al. Deep Convolution Neural Networks for Drug-Drug Interaction Extraction. 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2018.
- [113] Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* 2018;34(13):i457–66.
- [114] Xiong W et al. Extracting Drug-drug Interactions with a Dependency-based Graph Convolution Neural Network. 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2019.
- [115] Bruna J et al. Spectral networks and deep locally connected networks on graphs. 2nd International Conference on Learning Representations, 2014.
- [116] Collobert R et al. Natural Language Processing (Almost) from Scratch. *J Mach Learn Res* 2011;12:2493–537.
- [117] Sutskever, I., O. Vinyals, and Q.V. Le. *Sequence learning with neural networks*, in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*. 2014, MIT Press: Montreal, Canada. p. 3104–3112.
- [118] Zhang, S., et al. *Bidirectional Long Short-Term Memory Networks for Relation Classification*. in *PACLIC*. 2015.
- [119] Sahu SK, Anand A. Drug-drug interaction extraction from biomedical texts using long short-term memory network. *J Biomed Inform* 2018;86:15–24.
- [120] Wang W et al. Dependency-based long short term memory network for drug-drug interaction extraction. *BMC Bioinf* 2017;18(16):578.
- [121] Kim Y et al. Character-aware neural language models. In: *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. Phoenix, Arizona: AAAI Press; 2016. p. 2741–9.
- [122] Zhang, X., J. Zhao, and Y. LeCun. *Character-level convolutional networks for text classification*, in *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*. 2015, MIT Press: Montreal, Canada. p. 649–657.
- [123] Kavuluru R, Rios A, Tran T. Extracting Drug-Drug Interactions with Word and Character-Level Recurrent Neural Networks. In: *IEEE International Conference on Healthcare Informatics*. IEEE International Conference on Healthcare Informatics. p. 5–12.
- [124] Hochreiter S, Schmidhuber J. Long Short-Term Memory. *Neural Comput* 1997;9(8):1735–80.
- [125] Cho, K., et al. *On the Properties of Neural Machine Translation: Encoder-Decoder Approaches*. in *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*. 2014.
- [126] Zhou D, Miao L, He Y. Position-aware deep multi-task learning for drug-drug interaction extraction. *Artif Intell Med* 2018;87:1–8.
- [127] Jiang Z, Gu L, Jiang Q. Drug drug interaction extraction from literature using a skeleton long short term memory neural network. 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2017.
- [128] Zaikis D, Vlahavas I. Drug-Drug Interaction Classification Using Attention Based Neural Networks. In: *11th Hellenic Conference on Artificial Intelligence*. Athens, Greece: Association for Computing Machinery; 2020. p. 34–40.
- [129] Yi Z et al. Drug-drug interaction extraction via recurrent neural network with multiple attention layers. *International Conference on Advanced Data Mining and Applications*. Springer; 2017.
- [130] Zheng W et al. An attention-based effective neural model for drug-drug interactions extraction. *BMC Bioinf* 2017;18(1):445.
- [131] Adadi A, Berrada M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 2018;2169–3536.
- [132] Guidotti, R., et al. *A Survey of Methods for Explaining Black Box Models*. *ACM Comput. Surv.*, 2018. **51**(5): p. Article 93.
- [133] Simonyan, K., A. Vedaldi, and A. Zisserman. *Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps*. *CoRR*, 2014. abs/1312.6034.
- [134] Shrikumar, A., P. Greenside, and A. Kundaje. *Learning important features through propagating activation differences*, in *Proceedings of the 34th International Conference on Machine Learning - Volume 70*. 2017, JMLR.org: Sydney, NSW, Australia. p. 3145–3153.

- [135] Springenberg, J.T., et al., *Striving for Simplicity: The All Convolutional Net*. CoRR, 2015. abs/1412.6806.
- [136] Tao Lei RB, Jaakkola T. Rationalizing Neural Predictions. 2016 Conference on Empirical Methods in Natural Language Processing. Austin, Texas: Association for Computational Linguistics; 2016.
- [137] Bolukbasi T et al. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Adv Neural Inform Processing Syst* 2016;29.
- [138] Zügner, D., A. Akbarnejad, and S. Günnemann, *Adversarial Attacks on Neural Networks for Graph Data*, in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2018, Association for Computing Machinery: London, United Kingdom. p. 2847–2856.
- [139] Miyato, T., A.M. Dai, and I. Goodfellow, *Adversarial training methods for semi-supervised text classification*. arXiv preprint arXiv:1605.07725, 2016.
- [140] Feng, S., et al., *Pathologies of neural models make interpretations difficult*. arXiv preprint arXiv:1804.07781, 2018.
- [141] Huang K et al. CASTER: Predicting Drug Interactions with Chemical Substructure Representation. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. p. 702–9.
- [142] Dewulf P, Stock M, De Baets B. Cold-Start Problems in Data-Driven Prediction of Drug-Drug Interaction Effects. *Pharmaceuticals (Basel)* 2021;14(5).
- [143] Minard A-L et al. Feature selection for drug-drug interaction detection using machine-learning based approaches. In *Challenge Task on Drug-Drug Interaction Extraction (DDI)*, SEPLN, 2011.
- [144] Mahendran D, Nawarathna RD. An automated method to extract information in the biomedical literature about interactions between drugs. 2016 Sixteenth International Conference on Advances in ICT for Emerging Regions (ICTer), 2016.
- [145] Karim, M.R., et al., *Drug-Drug Interaction Prediction Based on Knowledge Graph Embeddings and Convolutional-LSTM Network*, in *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*. 2019, Association for Computing Machinery: Niagara Falls, NY, USA. p. 113–123.
- [146] Liu, S.a.H., Ziyang and Qju, Yang and Chen, Yi-Ping Phoebe and Zhang, Wen, Structural Network Embedding using Multi-modal Deep Auto-encoders for Predicting Drug-drug Interactions, in 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2019, IEEE.
- [147] Wang, Y., et al., *Dependency and AMR Embeddings for Drug-Drug Interaction Extraction from Biomedical Literature*, in *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. 2017, Association for Computing Machinery: Boston, Massachusetts, USA. p. 36–43.