Korean J. Chem. Eng., 40(2), 276-285 (2023)
DOI: 10.1007/s11814-023-1377-3
pISSN: 0256-1115
eISSN: 1975-7220

INVITED REVIEW PAPER

Recent development of machine learning models for the prediction of drug-drug interactions

Eujin Hong*, Junhyeok Jeon*, and Hyun Uk Kim*,,†**

*Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Korea
**BioProcess Engineering Research Center and BioInformatics Research Center, KAIST, Daejeon 34141, Korea
(*Received 11 September 2022 • Revised 9 December 2022 • Accepted 16 December 2022*)

Abstract−Polypharmacy, the co-administration of multiple drugs, has become an area of concern as the elderly population grows and an unexpected infection, such as COVID-19 pandemic, keeps emerging. However, it is very costly and time-consuming to experimentally examine the pharmacological effects of polypharmacy. To address this challenge, machine learning models that predict drug-drug interactions (DDIs) have actively been developed in recent years. In particular, the growing volume of drug datasets and the advances in machine learning have facilitated the model development. In this regard, this review discusses the DDI-predicting machine learning models that have been developed since 2018. Our discussion focuses on dataset sources used to develop the models, featurization approaches of molecular structures and biological information, and types of DDI prediction outcomes from the models. Finally, we make suggestions for research opportunities in this field.

Keywords: Polypharmacy, Drug-Drug Interaction, Adverse Drug Reaction, Machine Learning, Featurization

INTRODUCTION

Polypharmacy, defined to be the use of multiple drugs, may trigger pharmacological effects that would not be expected if the drugs were administered individually. Some of these unexpected pharmacological effects are associated with negative health effects, in particular adverse drug reactions (ADRs) [1]. Therefore, understanding drug-drug interaction (DDI) has received much attention to maximize the drug efficacy and to minimize unexpected ADRs. DDIs have become increasingly more important in some countries because they have reached the aging or aged societies, which potentially indicates the use of a greater number of multiple drugs at once for a patient. According to a recent study by Health Insurance Review and Assessment Service (HIRA) on the Korean National Health Insurance (NHI) claims data, 41.8% of the outpatients aged 65 or older were taking five or more medications, and 14.4% of the population were taking ten or more medications in 2019 [2]. Also, in recent years, many countries have suffered from the COVID-19 pandemic where the infected patients with an underlying medical condition (e.g., cardiovascular disease or diabetes) had to take multiple drugs [3]. Despite the importance of DDI issues, experimental examination of the DDIs is considered highly challenging due to high costs and time that are required to examine a large number of drug combinations.

To address this problem, various computational methods have been developed to predict the DDI effects, in particular by using machine learning methods in recent years. To the best of our knowledge, DeepDDI was a computational model that adopted the use of deep learning for the first time to predict DDIs for two input drugs [4]. Thereafter, a number of machine learning models were developed that used datasets from more diverse sources, and more sophisticated machine learning methods. Here, it was considered timely to review these machine learning models recently developed. Particularly, in this review, we discuss these machine learning models with a focus on dataset sources, featurization approaches of model input data (e.g., molecular structures, and graphs representing DDIs or biological knowledge), and the prediction of DDIs from machine learning models covered herein (Fig. 1). Finally, we further discuss future research opportunities in the field of DDI prediction.

PREPARATION OF A DATASET

Preparation of a dataset with sufficient volume and quality has always been a challenge when developing a machine learning model. For the issue of DDIs, DrugBank has served as a main source of a DDI dataset (Table 1); the DrugBank dataset was used by 15 out of the 18 machine learning models covered in this review, including DeepDDI [4], SSI-DDI [5], AMDE [6] and MDF-SA-DDI [7] (Table 2). DrugBank is a database that offers comprehensive information on drugs, including molecular information, mechanism of action, drug targets as well as DDIs [8]. When DeepDDI was developed by using the DrugBank dataset, the information on 192,284 DDIs from 191,878 drug pairs was used [4]. Currently, DrugBank provides information on 365,984 DDIs from 2,358 approved drugs [9]. TWOSIDES dataset [10] is another notable source used to develop DDI-predicting machine learning models (Table 1); it provides information on side effects associated with DDIs, and covers 868,221 associations between 59,220 drug pairs and 1,301 ADRs.

†To whom correspondence should be addressed.
E-mail: ehukim@kaist.ac.kr

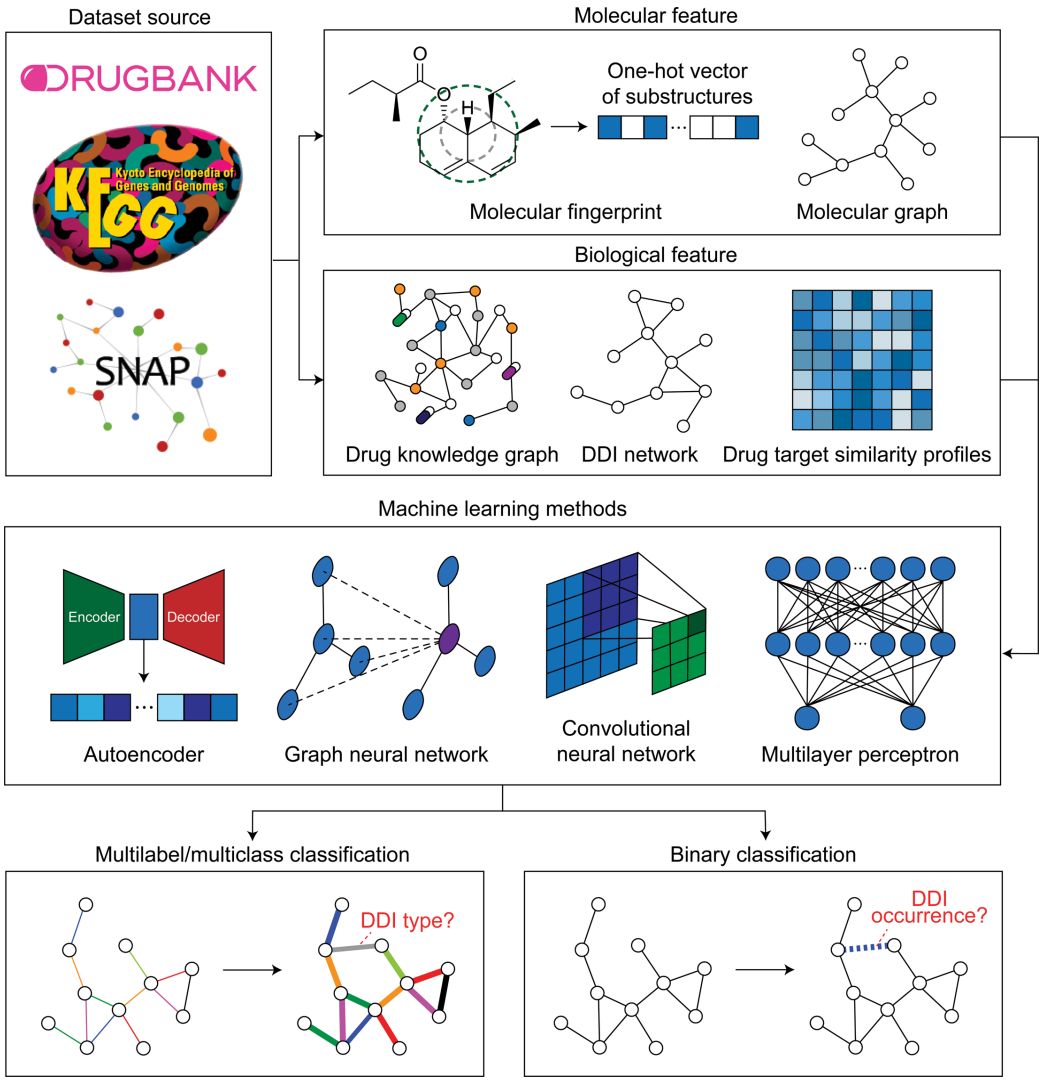

Fig. 1. Overall scheme of recently developed machine learning models that predict drug-drug interactions (DDIs). The presented machine learning methods can be used for both featurizing model input data and predicting DDIs.

Table 1. Source of datasets used to develop DDI-predicting machine learning models

Dataset source	Description	URL	Reference
Decagon	A graph convolutional network that predicts side effects for a given drug pair on the basis of a graph with heterogenous edge types	http://snap.stanford.edu/decagon/	[11]
DRKG	A biological knowledge graph showing relationships among several entities, such as drugs, diseases, biological processes and side effects, on the basis of six databases (i.e., DrugBank, Hetionet, GNBR, String, IntAct and DGIdb) as well as relevant literature	https://github.com/gnn4dr/DRKG/	[12]
DrugBank	A database containing comprehensive information on drugs and their targets, and also drug-drug interactions	https://go.drugbank.com	[8]
KEGG	A collection of databases that provide various types of biological information, including genome, metabolic pathways and biochemical reactions	https://www.genome.jp/kegg/	[13]
TWOSIDES	A dataset on significant associations between drug pairs and adverse drug events	https://tatonettilab.org/resources/nsides/	[10]

Table 2. Recent development of machine learning models that predict drug-drug interactions (DDIs)

Model	Dataset source[a]	Model input	Machine learning for featurization	Machine learning for DDI prediction	Output	Reference
DeepDDI, 2018	• DrugBank	• A pair of structural similarity profiles	• MLP[b]		• Multilabel classification for 86 DDI types	[4]
Lee et al., 2019	• DrugBank • BioGrid [33] • Gene Ontology terms [34,35]	• A pair of structural similarity profiles • A pair of Gene Ontology terms similarity profiles • A pair of target gene similarity profiles	• Autoencoder	• MLP	• Multilabel classification for 106 DDI types	[31]
DDIMDL, 2020	• DrugBank • KEGG	• A pair of structural similarity profiles • A pair of target similarity profiles • A pair of enzyme similarity profiles • A pair of pathway similarity profiles	• MLP		• Multiclass classification of 65 types of DDI-associated events	[21]
DPDDI, 2020	• DrugBank • Zhang et al. [40]	• DDI network	• GCN	• MLP	• Binary classification (the occurrence of DDIs)	[22]
GCN-BMP, 2020	• ModStore [41] • DeepDDI	• A pair of molecular graphs	• GCN	• MLP	• Binary classification (the occurrence of DDIs) • Multilabel classification for DDI types	[42]
MUFFIN, 2021	• DrugBank • TWOSIDES • DRKG	• Molecular graphs • Biological knowledge graph	• GNN • TransE • CNN	• MLP	• Binary classification (the occurrence of DDIs) • Multiclass classification for DDI types • Multilabel classification for DDI types	[28]
SSI-DDI, 2021	• DrugBank	• A pair of molecular graphs	• GAT	• Co-attention mechanism	• Multilabel classification for DDI types	[5]

Decagon [11], DRKG [12] and KEGG [13] are additional sources, from which datasets are prepared to develop the machine learning models. Decagon and DRKG are considered suitable when developing graph neural networks (GNNs) and biological/drug knowledge graphs, respectively. Meanwhile, DrugBank and KEGG are used for a wide range of machine learning methods, and they continue to be updated regularly.

Use of the well-curated datasets from the representative databases raises several considerations when developing new DDI-predicting machine learning models. First, as these databases are heavily dependent on the reported studies, it will be interesting to see how the performance of the machine learning model will improve in accordance with an increasing volume of the relevant data. Timely update of the existing machine learning model will be able to answer this question. Second, preparation of an in-house DDI dataset, which is likely distinct from the publicly available datasets mentioned above

Table 2. Continued

Model	Dataset source[a]	Model input	Machine learning for featurization	Machine learning for DDI prediction	Output	Reference
SumGNN, 2021	• DrugBank • TWOSIDES • HetioNet [43]	• Biological knowledge graph • Morgan fingerprints of drug molecules	• TransE	• GNN	• Multiclass classification for DDI types • Multilabel classification for DDI types	[29]
AMDE, 2022	• DrugBank	• Molecular graphs • Molecular substructure sequences	• GNN • Transformer	• MLP	• Binary classification (the occurrence of DDIs)	[6]
BioDKG-DDI, 2022	• DrugBank • Yue et al. [45]	• Morgan fingerprint • Drug knowledge graph • Drug-related protein receptor similarity profiles	• Mol2Context-vec • ComplEx-DURA	• MLP	• Binary classification	[30]
DeepMDDI, 2022	• DrugBank	• DDI network calculated from similarity profiles of ECFPs, MACCSkeys fingerprints and one-hot vectors of drug binding proteins	• Relation GCN		• Multilabel classification for DDI types	[27]
DeSIDE-DDI, 2022	• LINCS L1000 [46] • TWOSIDES	• Gene expression signatures	• Gated linear units (GLUs) • TransE		• Multilabel classification for side effect types	[32]
GMPNN-CS, 2022	• DrugBank • TWOSIDES	• A pair of molecular graphs	• GNN	• MLP	• Multilabel classification for DDI types	[47]
He et al., 2022	• Zhang et al. [40] • ChCh-Miner (http://snap.stanford.edu/biodata) • DrugBank	• Molecular graphs • SMILES • DDI network	• GNN • BiGRU • GCN	• MLP	• Binary classification (the occurrence of DDIs)	[48]
3WDDI, 2023	• ZhangDDI [44] • DRKG	• A pair of structural similarity profiles • A biological knowledge graph	• ComplEx • CNN	• CNN	• Binary classification (the occurrence of DDIs)	[36]

(Table 1), can also be considered, depending on a project objective (e.g., DDI prediction for a specific ethnicity). In this case, using a database that provides summarized information on DDIs (e.g., Drugs.com at https://www.drugs.com/) can facilitate the time-consuming manual curation of the reported studies on DDIs. Another more efficient approach would be the use of a text mining method, although the resulting DDI information collected is likely less accurate than the manual curation. An example would be the use of SciBERT for the prediction of DDIs from the literature [14]. Third, various biological information, such as target proteins, related diseases and pathways, can also be used in addition to the DDI information. A previous relevant example is the network constructed

Table 2. Continued

Model	Dataset source[a]	Model input	Machine learning for featurization	Machine learning for DDI prediction	Output	Reference
MDF-SA-DDI, 2022	• DrugBank	• A pair of structural similarity profiles • A pair of target similarity profiles • A pair of enzyme similarity profiles	• CNN • Autoencoder • Siamese network	• MLP	• Multiclass classification for DDI types	[7]
MS-ADR, 2022	• Decagon • PubChem • DrugBank • SIDER [49,50]	• Adverse drug reaction signed drug network • Similarity matrices of enzyme, indication, side effect and transporter	• GCN		• Binary classification (the occurrence of adverse drug reaction)	[51]
RANEDDI, 2022	• DrugBank	• DDI network • Multirelational network of DDIs	• RotatE		• Binary classification (the occurrence of DDIs)	[52]
			• RotatE	• MLP	• Multilabel classification for DDI types	

[a]The datasets presented herein were used for the model training, and not external validation.
[b]A merged cell for the two columns 'Machine learning for featurization' and 'Machine learning for DDI prediction' indicates that machine learning (e.g., GNN) was not used for the corresponding model when generating a feature vector.

by Lee et al. [15] that contains a range of biological information to analyze DDIs, and there have been advances in using such diverse information for the DDI prediction, particularly in the form of graphs as elaborated below (Figs. 1 and 2). Finally, there are databases that are highly specific to a disease, for example, 'Liverpool COVID-19 Drug Interactions' (https://www.covid19-druginteractions.org/), a database that provides experimental on evidence associated with COVID-19. This type of a dataset can be useful for the infected patients who have been taking drugs for other underlying medical conditions (e.g., hypertension and diabetes). Because such a dataset usually has a small volume, it can be useful when testing the already trained DDI prediction model, or it can be subjected to a new machine learning method that can handle a dataset with a small volume, for example, Bayesian neural network (BNN) [16,17].

FEATURIZATION OF MODEL INPUT DATA

1. Molecular Structures

The molecular structure of a drug is often considered as a main input for the DDI-predicting machine learning models (Fig. 2 and Table 2). For a molecular structure, simplified molecular-input line-entry system (SMILES) is the preferred format of input data; SMILES describes the molecular structure by using ASCII strings [18], and is the most widely used line notation in computational biology and chemistry. Once SMILES strings are prepared for input drugs, they are usually converted to molecular descriptors that correspond to numerical representations describing physicochemical properties and substructures of a molecule [19]. Among several types of molecular descriptors available, extended-connectivity fingerprint (ECFP) [20] has been most widely used to describe a drug molecule in the form of an one-dimensional (1D) binary vector. PubChem fingerprint [21,22] has also been used to develop a DDI-predicting model. Several software programs are available that generate molecular descriptors, such as RDKit (http://www.rkdit.org), Mordred [23], and ChemoPy [24]. The resulting molecular descriptors are subsequently further processed; for example, they are used to generate structural similarity profiles for two input drugs by using a similarity measure (e.g., Tanimoto coefficient) (Fig. 2).

In recent years, machine learning has been increasingly used to extract a range of important, yet non-intuitive, features that can contribute to further improving the DDI prediction. The most frequently used machine learning for the feature extraction is GNN according to the models covered in this review (Table 2). GNN has been considered suitable for handling DDI information because of its capacity to handle molecular information in an efficient and intuitive manner. GNN is trained by updating each node information in accordance with the neighbor nodes, and the most popular GNNs are graph convolutional network (GCN) [25] and graph attention network (GAT) [26]. One way of using a GNN is to extract molecular features (e.g., importance of substructures in the context of DDIs) from a molecular graph that presents atoms and their bonds of a

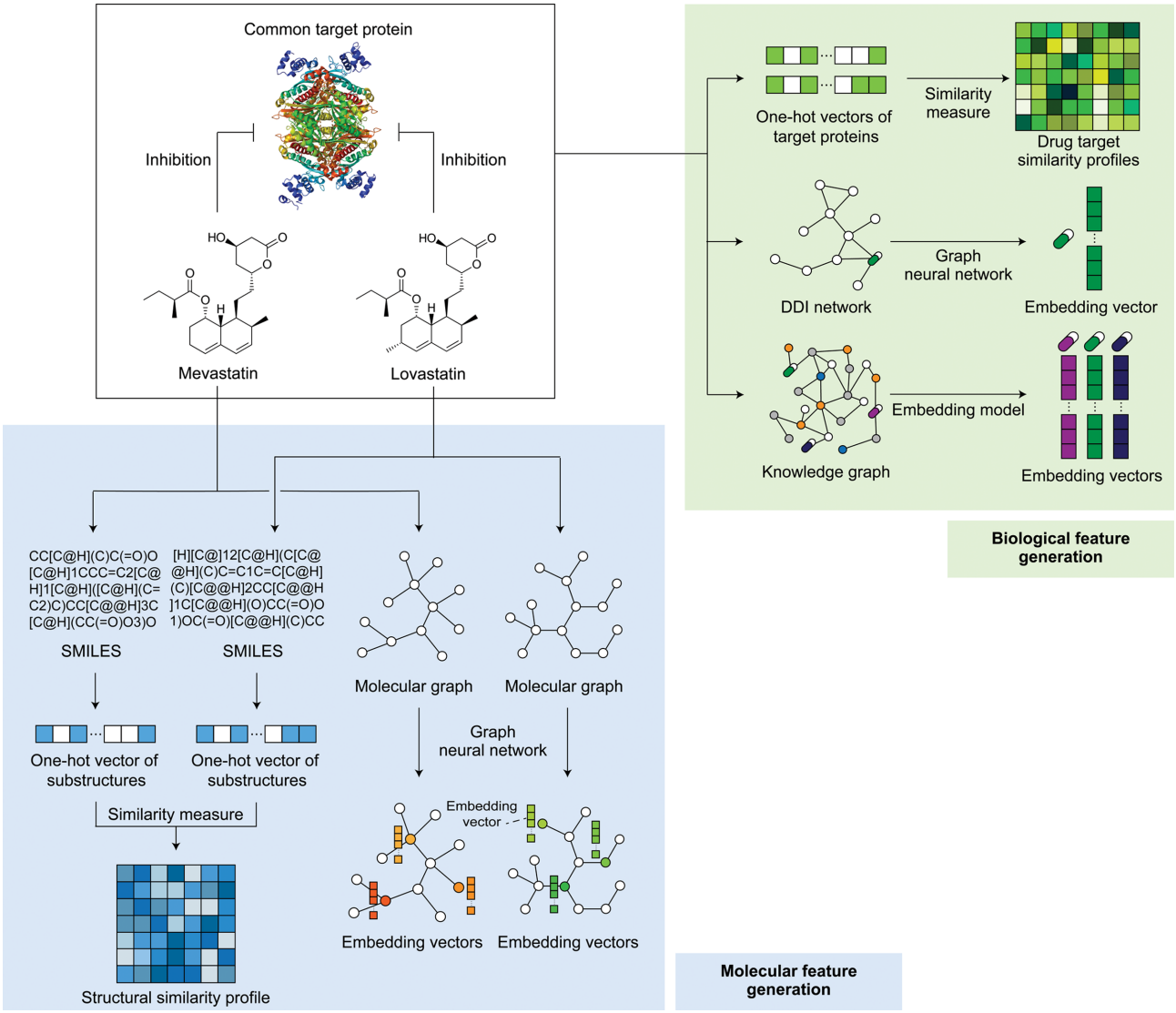

Fig. 2. Generation procedure of molecular and biological features for the prediction of DDIs using machine learning. Biological features include biological information, such as target proteins, a DDI network and a knowledge graph.

molecule (e.g., drug) by using nodes and edges, respectively (Fig. 2). Molecular graphs, which can be readily generated by RDKit, are further subjected to another machine learning method to predict DDIs. SSI-DDI [5] is a representative model that predicts DDIs by considering substructures of two input drugs, and for this, these substructures were initially extracted from molecular graphs of two input drugs by using GAT (Fig. 3(a)). SSI-DDI was developed based on previous studies reporting that only several substructures of drugs are responsible for DDIs. During the feature extraction, nodes in the GAT initially represent atoms of a drug, but eventually generate substructures as a result of updating by aggregating the surrounding nodes. These substructures are subsequently subjected to co-attention mechanism to predict DDIs.

2. Biological Information

Use of a feature vector that reflects the molecular structure of a drug has been successful for predicting DDI at a reasonably high accuracy, but there is room for incorporating other 'biological information' for further improved prediction outcome, and for enhanced explainability of the prediction. For example, the following biological information has been considered in the DDI-predicting models in addition to the molecular features (Fig. 2 and Table 2): drug target proteins for DDIMDL [21] and MDF-SA-DDI [7]; DDI networks for DPDDI [22] and deepMDDI [27]; biological/drug knowledge graphs for MUFFIN [28], SumGNN [29], and BioDKG-DDI [30]; gene ontology terms for the model by Lee et al. [31]; and gene expression signatures for DeSIDE-DDI [32]. Proteins, such as cytochromes P450 or protein drug targets, are representative examples that can significantly affect the occurrence of a DDI. Drug target information for a drug is expressed as a one-hot vector where '1' indicates that the corresponding protein in the vector is a drug target, and '0' is for a non-target protein. Subsequent comparison of the two one-hot vectors would provide a clue for possible interactions between two input drugs. If the two vectors appear to be similar, there is a high chance of a DDI because the two drugs inhibit

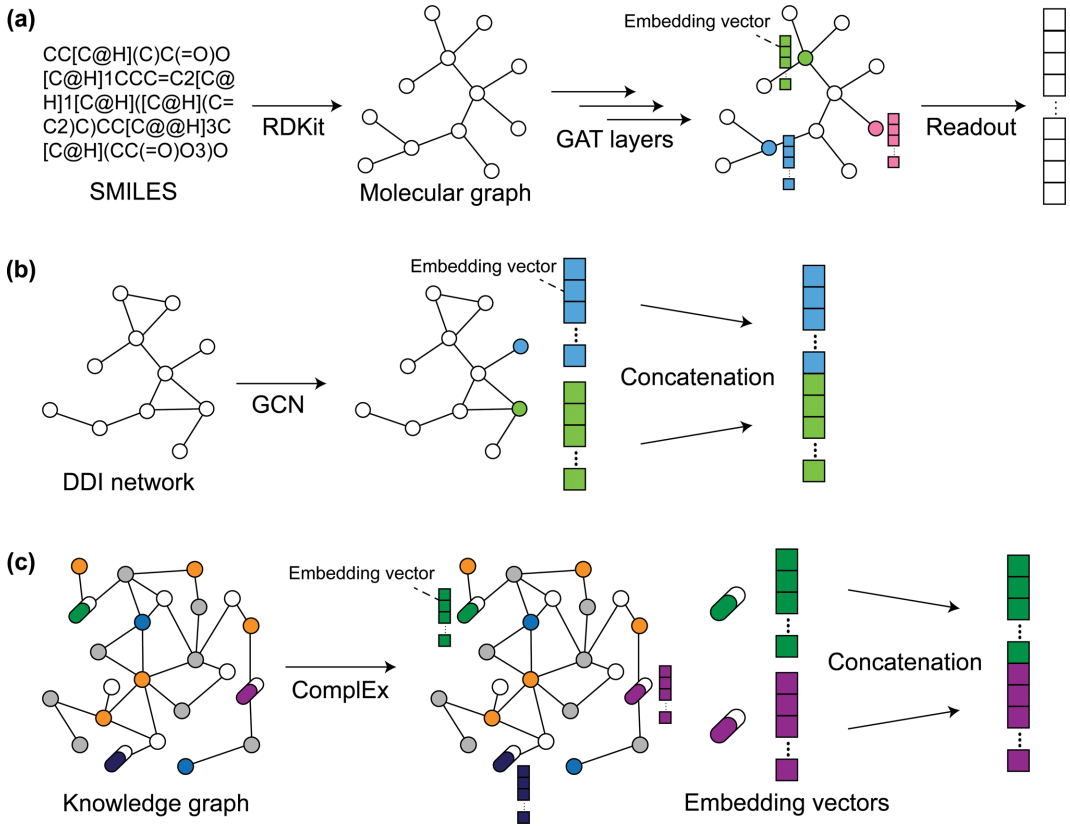

Fig. 3. Generation of molecular and biological features for the three DDI-predicting machine learning models: (a) SSI-DDI [5] by using graph attention network (GAT), (b) DPDDI [22] by using graph convolutional network (GCN), and (c) 3WDDI [36] by using knowledge graph embedding.

the same target; one drug-target interaction would affect the interaction of the other drug with the same protein target (Fig. 2).

In order to extract important biological features for the DDI prediction, GNNs have also been used in a way distinct from representing a molecular structure discussed above (i.e., using nodes and edges of a graph to indicate atoms and their bonds of a drug). In the context of biological information, nodes and edges of a GNN are designed to indicate drugs and their interactions, respectively, and thus, the GNN structure itself becomes a DDI network. DPDDI [22] is a representative example where GCN is used to capture the information on each drug (node) by aggregating and transforming information from its neighboring drugs that are connected through edges (Fig. 3(b)). Application of a GCN to a series of drugs generates a latent feature vector for each node (drug), which, in turn, is used as an input data for a machine learning model that finally predicts DDI. Additional biological information, such as genetic, protein and/or chemical interactions [33] and gene ontology [34,35], is expected to capture even more diverse effects of a DDI.

The combination of molecular and biological features indeed showed a notable improvement on the DDI prediction performance in comparison with using molecular structures alone. DDIMDL [21] is a DDI-predicting model that utilizes four different types of drug features, including drug targets, relevant enzymes (mostly, cytochromes P450), pathways (e.g., pathways having the drug targets) as well as molecular structures. To evaluate the influence of each drug feature on DDI prediction, the authors built different versions of the prediction model, each using a different combination of the four drug features. The model using only the molecular structure outperformed those using drug targets, enzymes, or pathways, individually, achieving reasonably high accuracy. However, when target and enzyme features were additionally considered together with molecular features, the model showed further improved prediction accuracy. The beneficial effects of using biological features were also demonstrated by Lee et al. [31]. In their model, the combination of three similarity profiles for two drugs, including those associated with molecular structures, target genes and gene ontology terms, increased the classification accuracy, compared to the models that used structural similarity profiles alone.

In addition to the GNNs, other knowledge graph embedding models were also utilized to extract necessary semantic features from knowledge graphs. A knowledge graph is composed of nodes in the form of triplets (h, r, t) where h and t represent head and tail entities, respectively, and r represents the relationship between them; relationship types can be diverse within one knowledge graph. For example, DRKG is a representative knowledge graph often used to develop models for the DDI prediction as it provides information on relationships among several entities, such as drugs, diseases, biological processes and side effects. To integrate all the available information (or knowledge) associated with drugs for the DDI pre-

diction, all the entities and their multi-type of semantic relationships in DRKG first need to be mapped into a low-dimensional space by using appropriate knowledge graph embedding models. 3WDDI [36] is such a DDI-predicting model that adopts the use of embedding vectors extracted from DRKG by using ComplEx [37], a semantic similarity-based knowledge graph embedding model (Fig. 3(c)). The resulting embedding vectors are used as input of the subsequent machine learning model that predicts the occurrence of DDI.

PREDICTION OF DRUG-DRUG INTERACTION

Upon generation of molecular and/or biological features for a given pair of drugs, often from a machine learning model (e.g., GNN) in recent years, the features are used as input for another machine learning model that predicts the occurrence or types of DDIs (i.e., binary or multilabel/multiclass classification, respectively). Two predominantly used machine learning methods for the DDI prediction are multilayer perceptron (MLP) and GNN (Table 2). MLP can easily process several different types of input data (e.g., molecular structures, drug targets, and enzymes), while GNN is particularly suitable for intuitively handling the DDI network (e.g., drug as a node and interaction as an edge) as discussed above. Currently, machine learning methods are mostly implemented by using PyTorch [38], TensorFlow [39] and Keras (https://github.com/fchollet/keras).

Binary classification of DDIs simply indicates whether there is an interaction between two drugs of interest. If GNNs are used, this binary classification problem corresponds to the 'link prediction' that predicts the presence or absence of an edge between two nodes in the graph. Nine out of 18 DDI prediction models perform binary classification for DDIs, including DPDDI [22], AMDE [6] and DeSIDE-DDI [32]. Multilabel and multiclass classifications have also been performed for the DDI prediction, which predicts 'DDI types'; for example, 'the increased risk or severity of adverse effects', 'increased hepatotoxic activity' or 'the decreased metabolism of a drug' can be predicted as an outcome for DeepDDI [4]. For multilabel and multiclass classification, it is common to use the DrugBank DDI dataset, but the number of DDI types to be predicted can vary, depending on the dataset preprocessing, use of other dataset (e.g., TWOSIDES) in addition to DrugBank DDI dataset, and the output format. Representative models that perform multilabel or multiclass classification include DeepDDI [4], a model by Lee et al. [31], DeepMDDI [27] and MDF-SA-DDI [7], which predict 86, 106, 11 and 65 DDI types, respectively.

Recent efforts to develop DDI-predicting machine learning models have focused on improving the prediction accuracy. Partly as a result of these efforts, 14 out of 16 DDI-predicting models, which were covered in this review (Table 2) and evaluated using an area under the receiver operating characteristic curve (AUROC), achieved greater than 0.95, and four of them even reached greater than 0.99. However, in order for the DDI-predicting models to be more practical, additional clinical and biochemical factors need to be further considered. First, pharmacological effects of DDIs may vary, depending on the individual. Thus, consideration of personal information (e.g., age, gender, medical condition, and genetic information such as presence or absence of specific genes or mutations) would be able to make the DDI prediction more person-specific; however, this additional consideration may bring about another challenge of data privacy. Another important factor is the concentration of a drug given to a patient, which would affect its therapeutic effects. To realize these additional considerations, of course, relevant datasets will be necessary, in particular, clinical information. For this reason, collaboration with medical practitioners will be essential to make a step change in developing DDI-predicting machine learning models.

CONCLUSION

In this review, we discussed the DDI-predicting machine learning models that have been developed since 2018. From this review, it is clear that an increasing number of DDI-predicting models have been developed over the years, likely due to the more public recognition of DDIs and the increased availability of relevant datasets and machine learning methods (Table 2). Among these models developed, molecular and biological features have been increasingly generated by using machine learning, for example, GNN and knowledge graph embeddings. Meanwhile, MLP and GNN have served as the main machine learning methods to predict DDIs. Despite active development of the DDI-predicting models, sophisticated clinical and biochemical factors need to be further taken into account for more practical use of these models for the DDI prediction. With active adoption of new DDI information as well as novel modeling methods, the DDI-predicting models will make a greater contribution to addressing the issues (e.g., ADRs) arising from polypharmacy.

ACKNOWLEDGEMENTS

This research was supported by the KAIST Key Research Institute (Interdisciplinary Research Group) Project. This work was also supported by internal fund/grant of Electronics and Telecommunications Research Institute (ETRI) [22RB1100, Exploratory and Strategic Research of ETRI-KAIST ICT Future Technology].

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. E. A. Davies and M. S. O'Mahony, *Br. J. Clin. Pharmacol.*, **80**, 796 (2015).
2. H. J. Cho, J. Chae, S. H. Yoon and D. S. Kim, *Front. Pharmacol.*, **13**, 866318 (2022).
3. S. Iloanusi, O. Mgbere and E. J. Essien, *J. Am. Pharm. Assoc. (2003)*, **61**, e14 (2021).
4. J. Y. Ryu, H. U. Kim and S. Y. Lee, *Proc. Natl. Acad. Sci. U. S. A.*, **115**, E4304 (2018).
5. A. K. Nyamabo, H. Yu and J. Y. Shi, *Brief. Bioinform.*, **22**, 1 (2021).
6. S. Pang, Y. Zhang, T. Song, X. Zhang, X. Wang and A. Rodriguez-Paton, *Brief. Bioinform.*, **23**, 1 (2022).

7. S. Lin, Y. Wang, L. Zhang, Y. Chu, Y. Liu, Y. Fang, M. Jiang, Q. Wang, B. Zhao, Y. Xiong and D. Q. Wei, *Brief. Bioinform.*, **23**, 1 (2022).
8. D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox and M. Wilson, *Nucleic Acids Res.*, **46**, D1074 (2018).
9. D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang and J. Woolsey, *Nucleic Acids Res.*, **34**, D668 (2006).
10. N. P. Tatonetti, P. P. Ye, R. Daneshjou and R. B. Altman, *Sci. Transl. Med.*, **4**, 125ra31 (2012).
11. M. Zitnik, M. Agrawal and J. Leskovec, *Bioinformatics*, **34**, i457 (2018).
12. V. N. Ioannidis, X. Song, S. Manchanda, M. Li, X. Pan, D. Zheng, X. Ning, X. Zeng and G. Karypis, https://github.com/gnn4dr/DRKG/ (2021).
13. M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe and M. Hirakawa, *Nucleic Acids Res.*, **38**, D355 (2010).
14. M. Asada, M. Miwa and Y. Sasaki, *Bioinformatics*, **37**, 1739 (2021).
15. K. Lee, S. Lee, M. Jeon, J. Choi and J. Kang, *2012 IEEE Int. Conf. Bioinf. Biomed.*, 1 (2012).
16. S. Ryu, Y. Kwon and W. Y. Kim, *Chem. Sci.*, **10**, 8438 (2019).
17. M. Elbadawi, S. Gaisford and A. W. Basit, *Drug Discov. Today*, **26**, 769 (2021).
18. D. Weininger, *J. Chem. Inf. Comput. Sci.*, **28**, 31 (1988).
19. J. Jeon, S. Kang and H. U. Kim, *Nat. Prod. Rep.*, **38**, 1954 (2021).
20. D. Rogers and M. Hahn, *J. Chem. Inf. Model.*, **50**, 742 (2010).
21. Y. Deng, X. Xu, Y. Qiu, J. Xia, W. Zhang and S. Liu, *Bioinformatics*, **36**, 4316 (2020).
22. Y. H. Feng, S. W. Zhang and J. Y. Shi, *BMC Bioinformatics*, **21**, 419 (2020).
23. H. Moriwaki, Y. S. Tian, N. Kawashita and T. Takagi, *J. Cheminform.*, **10**, 4 (2018).
24. D. S. Cao, Q. S. Xu, Q. N. Hu and Y. Z. Liang, *Bioinformatics*, **29**, 1092 (2013).
25. T. N. Kipf and M. Welling, *arXiv preprint arXiv:1609.02907* (2016).
26. P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio and Y. Bengio, *arXiv preprint arXiv.1710.10903* (2017).
27. Y. H. Feng, S. W. Zhang, Q. Q. Zhang, C. H. Zhang and J. Y. Shi, *Anal. Biochem.*, **646**, 114631 (2022).
28. Y. Chen, T. Ma, X. Yang, J. Wang, B. Song and X. Zeng, *Bioinformatics*, **37**, 2651 (2021).
29. Y. Yu, K. Huang, C. Zhang, L. M. Glass, J. Sun and C. Xiao, *Bioinformatics*, **37**, 2988 (2021).
30. Z. H. Ren, C. Q. Yu, L. P. Li, Z. H. You, Y. J. Guan, X. F. Wang and J. Pan, *Brief. Funct. Genomics*, **21**, 216 (2022).
31. G. Lee, C. Park and J. Ahn, *BMC Bioinformatics*, **20**, 415 (2019).
32. E. Kim and H. Nam, *J. Cheminform.*, **14**, 9 (2022).
33. A. Chatr-Aryamontri, R. Oughtred, L. Boucher, J. Rust, C. Chang, N. K. Kolas, L. O'Donnell, S. Oster, C. Theesfeld, A. Sellam, C. Stark, B. J. Breitkreutz, K. Dolinski and M. Tyers, *Nucleic Acids Res.*, **45**, D369 (2017).
34. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock, *Nat. Genet.*, **25**, 25 (2000).
35. C. The Gene Ontology, *Nucleic Acids Res.*, **45**, D331 (2017).
36. X. Hao, Q. Chen, H. Pan, J. Qiu, Y. Zhang, Q. Yu, Z. Han and X. Du, *Granular Computing*, **8**, 67 (2023).
37. T. Trouillon, J. Welbl, S. Riedel, É. Gaussier and G. Bouchard, *ICML*, **48**, 2071 (2016).
38. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein and L. Antiga, *Adv. Neural Inf. Process. Syst.*, **32**, 1 (2019).
39. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean and M. Devin, *arXiv preprint arXiv:1603.04467* (2016).
40. W. Zhang, Y. Chen, F. Liu, F. Luo, G. Tian and X. Li, *BMC Bioinformatics*, **18**, 18 (2017).
41. W. Zhang, Y. Liu, L. Wang, J. Zhou, J. Du and R. S. M. Goh, *ICCCRI*, 18 (2017).
42. X. Chen, X. Liu and J. Wu, *Methods*, **179**, 47 (2020).
43. D. S. Himmelstein and S. E. Baranzini, *PLoS Comput. Biol.*, **11**, e1004259 (2015).
44. H.-R. Zhang, F. Min and B. Shi, *Inform. Sci.*, **378**, 444 (2017).
45. X. Yue, Z. Wang, J. Huang, S. Parthasarathy, S. Moosavinasab, Y. Huang, S. M. Lin, W. Zhang, P. Zhang and H. Sun, *Bioinformatics*, **36**, 1241 (2020).
46. A. Subramanian, R. Narayan, S. M. Corsello, D. D. Peck, T. E. Natoli, X. Lu, J. Gould, J. F. Davis, A. A. Tubelli, J. K. Asiedu, D. L. Lahr, J. E. Hirschman, Z. Liu, M. Donahue, B. Julian, M. Khan, D. Wadden, I. C. Smith, D. Lam, A. Liberzon, C. Toder, M. Bagul, M. Orzechowski, O. M. Enache, F. Piccioni, S. A. Johnson, N. J. Lyons, A. H. Berger, A. F. Shamji, A. N. Brooks, A. Vrcic, C. Flynn, J. Rosains, D. Y. Takeda, R. Hu, D. Davison, J. Lamb, K. Ardlie, L. Hogstrom, P. Greenside, N. S. Gray, P. A. Clemons, S. Silver, X. Wu, W. N. Zhao, W. Read-Button, X. Wu, S. J. Haggarty, L. V. Ronco, J. S. Boehm, S. L. Schreiber, J. G. Doench, J. A. Bittker, D. E. Root, B. Wong and T. R. Golub, *Cell*, **171**, 1437 (2017).
47. A. K. Nyamabo, H. Yu, Z. Liu and J. Y. Shi, *Brief. Bioinform.*, **23**, 1 (2022).
48. C. He, Y. Liu, H. Li, H. Zhang, Y. Mao, X. Qin, L. Liu and X. Zhang, *BMC Bioinformatics*, **23**, 224 (2022).
49. M. Kuhn, M. Campillos, I. Letunic, L. J. Jensen and P. Bork, *Mol. Syst. Biol.*, **6**, 343 (2010).
50. M. Kuhn, I. Letunic, L. J. Jensen and P. Bork, *Nucleic Acids Res.*, **44**, D1075 (2016).
51. L. Zhuang, H. Wang, W. Li, T. Liu, S. Han and H. Zhang, *Soft Computing*, **26**, 11795 (2022).
52. H. Yu, W. Dong and J. Shi, *Inform. Sci.*, **582**, 167 (2022).

Hyun Uk Kim is currently Kwon Oh-Hyun Associate Professor at Department of Chemical and Biomolecular Engineering, KAIST. His research field lies in systems biology, systems medicine and metabolic engineering. His current studies are focused on computational modeling of medically important microorganisms and selected human cells, and analysis of the biological effects of drugs. He earned his B.S. at Yonsei University, and M.S. and Ph.D. at KAIST. He joined KAIST as Assistant Professor in 2018, and has published more than 80 scientific papers.