Original Article

Healthc Inform Res. 2020 January;26(1):42-49.
https://doi.org/10.4258/hir.2020.26.1.42
pISSN 2093-3681 • eISSN 2093-369X

HIR
Healthcare Informatics Research

Prediction of Drug–Drug Interactions by Using Profile Fingerprint Vectors and Protein Similarities

Selma Dere[1], **Serkan Ayvaz**[2]

[1]Department of Computer Engineering, Bahcesehir University, Besiktas, Istanbul, Turkey
[2]Department of Software Engineering, Bahcesehir University, Besiktas, Istanbul, Turkey

Objectives: Drug–drug interaction (DDI) is a vital problem that threatens people's health. However, the prediction of DDIs through *in-vivo* experiments is not only extremely costly but also difficult as many serious side effects are hard to detect in *in-vivo* and *in-vitro* settings. The aim of this study was to assess the effectiveness of similarity-based *in-silico* computational DDI prediction approaches and to provide a cost effective and scalable solution to predict potential DDIs. **Methods:** In this study, widely known similarity-based computational DDI prediction methods were utilized to discover novel potential DDIs. More specifically, known interactions, drug targets, adverse effects, and protein similarities of drug pairs were used to construct drug fingerprints for the prediction of DDIs. **Results:** Using the drug interaction profile, our approach achieved an area under the curve (AUC) of 0.975 in the prediction of a potential DDI. The drug adverse effect profile and protein profile similarity-based methods resulted in AUC values of 0.685 and 0.895, respectively, in the prediction of DDIs. **Conclusions:** In this study, we developed a computational approach to the prediction of potential drug interactions. The performance of the similarity-based computational methods was comparatively evaluated using a comprehensive real-world DDI dataset. The evaluations showed that the drug interaction profile information is a better predictor of DDIs compared to drug adverse effects and protein similarities among DDI pairs.

Keywords: Drug Interactions, Knowledge Discovery, Computing Methodologies, Drug-Related Side Effects and Adverse Reactions, Pharmaceutical Databases

Submitted: September 18, 2019
Revised: 1st, November 18, 2019; 2nd, December 20, 2019; 3rd, December 24, 2019
Accepted: December 25, 2019

Corresponding Author
Serkan Ayvaz
Department of Software Engineering, Bahcesehir University, Besiktas, Istanbul 34349, Turkey. Tel: +902123810886, E-mail: serkan.ayvaz@eng.bau.edu.tr (https://orcid.org/0000-0003-2016-4443)

I. Introduction

Drug–drug interaction (DDI) can be very harmful for patients and may cause serious health problems. When two or more drugs that can interact are administered, the patient is exposed to a potential DDI. DDIs cause up to 30% of adverse drug effects (ADEs). Thus, prevention of DDI plays a vital role in the treatment of patients. Moreover, adverse drug events are one of the primary reasons that drugs fail in clinical trials [1]. Besides their negative impact on public health, DDIs also result in significant economic losses of public resources. Prescribed interacting drugs cost tax payers billions of dollars. The number of new drugs entering the market is increasing every year (51 novel drugs were approved by the

Food and Drug Administration in 2018) [2]. Thus, accurate prediction of potential DDIs is becoming even more crucial for the prevention of drug related harm. DDIs have been taken into consideration seriously by researchers. Many *in-silico* [3], *in-vitro* [4], and *in-vivo* [5] experiments have been carried out. However, since *in-vitro* and *in-vivo* experiments are extremely expensive and not always feasible, various *in-silico* research methods that provide successful results have been developed.

Ferdousi et al. [6] presented an approach, in which functional similarities of drugs in terms of protein targets are used for DDI prediction. Tatonetti et al. [7] created an adaptive data-driven approach to predict drug effects and interactions. They utilized a drug adverse effect database by correcting omissions of confounding factors, such as concomitant medications, patient demographics, and patient medical histories. Vilar et al. [8] proposed a method for the prediction of DDIs by utilizing known DDIs. They generated interaction profile fingerprint (IPF) vectors to represent drugs and used these vectors to calculate predictions. From a different perspective, Zhang et al. [9] utilized recommender methods to predict the unknown side effects of drugs. They used the integrated neighborhood-based method and restricted Boltzmann machine-based method as recommender methods. Vilar et al. [1] proposed a comprehensive method for creating drug fingerprints and DDI predictions. They created these fingerprints for known interaction profiles, adverse effects, and protein targets, as well as 2D and 3D molecular structures.

Ayvaz et al. [10] created a complete set of DDIs to obtain a comprehensive DDI list using publicly available DDI resources, such as DrugBank, KEGG, and the NDF-RT. Noor et al. [11] developed a novel pharmacovigilance inferential framework to infer mechanistic explanations for asserted DDIs and deduce potential DDIs. In another research study, Vilar et al. [12,13] utilized molecular structures, generated fingerprints, and performed DDI predictions based on molecular similarities.

Similarly, the prediction of adverse drug reaction (ADR) has also been investigated extensively. Liu et al. [14] proposed a machine learning-based approach for ADR prediction by integrating the phenotypic characteristics of a drug, including indications, other known ADRs, the drug's chemical structures and biological properties, protein targets, and pathway information. In some studies, adverse event reports have been used to predict DDIs [15–17]. Hudelson et al. [18] considered drug metabolism for DDI prediction. Percha et al. [19] applied a text mining algorithm to medicine abstracts to predict DDI. A text mining approach was also used by Tari et al. [20].

In this study, we evaluated and compared two of these approaches mentioned above [1,6]. These approaches were selected based on their DDI prediction accuracy and effectiveness. The first approach [1], further detailed in the following, gives very accurate prediction results among the described methods, whereas the second approach [6] considers similarities between known interaction pairs. Similarity calculations based on the second approach are simpler and computationally faster. Also, the similarity measure used in this approach provides more accurate results. Taking these points into consideration, these two approaches were evaluated. The first approach predicts DDIs using profile fingerprints depending on various attributes, such as interaction profiles [1,8], adverse effect profiles, and protein profiles [1]. In this approach, drugs are represented by vector fingerprints and known predictions. The similarities of all drug pairs are used to calculate the prediction matrix. On the other hand, the protein similarities between DDI drugs are calculated and used for the prediction of DDIs in the second approach [6]. The first approach achieves superior prediction accuracy in comparison to the second one. In addition, the DDI predictive power of interaction profiles, adverse effect profiles, and protein profiles were further evaluated.

II. Methods

1. Data Collection

In the study, the drug resources Merged-PDDI dataset [10] and MedDRA [21] were used. The dataset contained the DDIs from the Merged-PDDI-dataset, corresponding protein identifiers (carrier, transport, enzyme, target) from DrugBank [22] for those DDIs, and the adverse effects of the interacting drugs were retrieved from MedDRA. Drug protein identifiers collected from DrugBank [22] were used to construct a protein vector with a total length of 5,643 items. MedDRA version 16.1 was retrieved from Embl (http://sideeffects.embl.de/download/) and DrugBank version 5.1.1 was retrieved on July 3, 2018 from the DrugBank repository (https://www.drugbank.ca/releases/latest/). A total of 2,591 unique drugs with relevant DrugBank identifiers were collected. From the Merged-PDDI dataset, the total of 297,816 known drug interaction pairs was retrieved. From MedDRA, the number of collected drug adverse effects was 309,849, and there were 6,061 unique adverse effects. The MedDRA data source contains 1,430 unique drugs, but the number of drugs overlapping with the Merged-PDDI dataset source was

995. Therefore, only the adverse effects of 995 drugs were considered during the evaluations.

2. Methods Used in DDI Predictions

In the prediction of potential DDI interactions, the similarities of drugs in known interactions were used. The similarities between drugs were calculated based on factors including drug interaction profile, adverse effect, and drug target. When predicting new interactions, our approach derives DDI predictions based on the degree of similarity between the drug pairs in known DDIs and novel drugs. Our main intuition is that if drug A has a known interaction with drug B, and the similarity between drug B and novel drug C is higher than a certain threshold, then there is a possibility that drug A might have an interaction with drug C. In other words, if two drugs are known to have an interaction and there exists another drug that is similar to one of the drugs in the DDI pair, then it is possible that the third drug might result in a DDI.

For the purpose of calculating drug similarities in the prediction of drug interactions, drug-related information must be extracted and represented in a common vector structure. Fingerprints are a type of vector structure, which were used to represent drug information as drug interaction profiles, proteins, and adverse effects. Before the similarity calculations, the vectors representing the drug IPFs, adverse effect profile fingerprints, and protein profile fingerprints were generated using training data.

When drug IPFs were constructed, each drug was represented by a vector with the length of 2,591. The index "*i*" has the value of 1 if the current drug interacts with the drug with index "*i*"; otherwise, it has the value of 0. Similarly, the adverse effect profile fingerprints with 6,061 representing all unique adverse effects were produced for the drugs. For the adverse effect fingerprint of a drug, the value of the index "*i*" was set to 1 if the drug had an adverse effect with index "*n*". Otherwise, the value of the index was set to 0. For instance, the adverse effect vector drug "leucovorin" is populated by 1 at the index 2. The index 2 in the adverse effect vector represents 'abdominal pain'. This means that the drug leucovorin may have abdominal pain as an adverse effect.

The protein profile fingerprints were also constructed in a similar way. Each drug was represented as a protein vector with the length of 5,643. This vector represents the combination of carrier, target, enzyme, and transport proteins. For each protein, an index is given, and the value of the corresponding entry has the value of 1 if the drug has a relation with this protein. Otherwise, the value is set to 0.

1) Similarity matrix generation

In this section, the similarity matrix calculations to be used for predictions are discussed in detail. To perform predictions, the interaction matrix M_1 was constructed based on the fingerprint vectors of drugs. For instance, M_1 for the drug IPFs has the size of 2,591 × 2,591. The value of $M_1(i, j)$ in the matrix is 1 if the drug with index "*i*" and the drug with index "*j*" are known to have a DDI with each other. Otherwise, the value of $M_1(i, j)$ is 0.

The similarity matrix M_2 is populated based on the similarity between the fingerprint vectors for the drug pairs represented in the interaction matrix M_1. The similarity matrix M_2 is the same size as M_1. The value of $M_2(i, j)$ represents the similarity value between a drug with index "*i*", and a drug with index "*j*" and takes a value between 0 and 1 calculated based on similarity measurement. Tanimoto coefficient measurement was used to calculate the similarities in the IPF vectors.

The Tanimoto coefficient, also known as the Jaccard index is a widely used similarity measure that gauges the similarity between finite sample sets [23]. It is calculated as the ratio of the intersection of two sets to the size of their union. It can take values in the range between 0 (minimum similarity) and 1 (maximum similarity). The Tanimoto coefficient is a suitable candidate for measuring similarity between drugs due to its simplicity, reliability, and the interpretability of its range. The similarities are calculated by taking the dot product of interaction profile vectors and applying normalization. In other words, the results are divided by the number of total 1's minus the number of common 1's in these two vectors, and the value is set to $M_2(i, j)$. The inner product and Tanimoto coefficient equations are formulated as follows:

$$S = x^t y = \sum_{i=1}^{d} S_i \,, \tag{1}$$

$$\text{where } S_i = \begin{cases} 1 & \text{if } x_i = y_i = 1, \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$S_{Tanimoto} = \frac{x^t y}{x^t x + y^t y - x^t y}. \tag{3}$$

The prediction matrix M_3 is calculated by a special cross product of interaction matrix M_1 and similarity matrix M_2. It is calculated in such a way that the result of product $M_1(i, j)$ and $M_2(j, k)$ is not summed with the result of product $M_1(i, j+1)$ and $M_2(j+1, k)$ but saved. In the end, the largest product value is set to $M_3(i, k)$. This calculation method is illustrated in Figure 1.

https://doi.org/10.4258/hir.2020.26.1.42

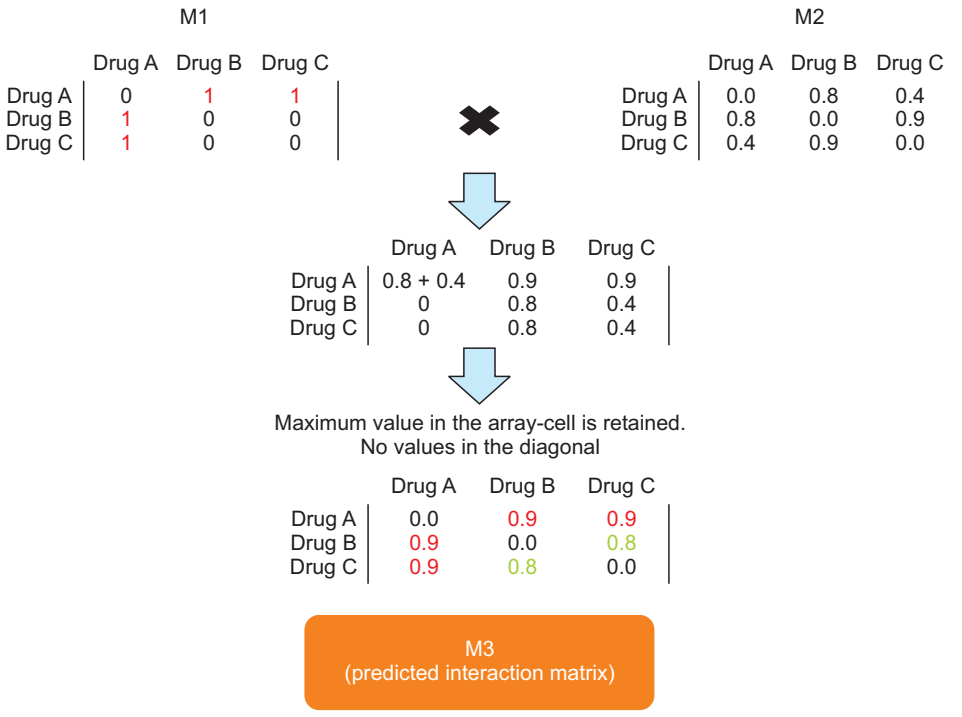

Figure 1. Prediction matrix M_3 calculation.

The prediction matrix M_3 is used to predict the likelihood of a drug with index "*i*" interacting with a drug with index "*j*". If $M_3(i, j)$ has a value greater than 0, then it is considered that the relevant drugs may interact with each other. In the evaluations, M_3 was calculated for these three different similarity matrices M_2.

2) Calculating Similarities between Protein Vectors

The impact of protein similarities between drugs in DDI was also examined. The degree of protein similarity among drugs with known interactions was explored as a predictor of potential drug interactions. Our intuition is that if drug A and drug B interact with each other, then it is possible that these two drugs might have similar protein fingerprints [6]. That is, when two drugs have high protein similarity, there is a chance that these drugs might have a potential drug interaction.

To evaluate this intuition, the same vector representations used in protein profile fingerprints among drugs with known interactions as well as interaction matrix M_1 and similarity matrix M_2 were utilized. Not only "approved" protein data but also the protein data under the "all" title in DrugBank were used in evaluations.

Each drug was represented by a carrier target enzyme transporter (CTET) protein vector. A CTET vector is a compound vector that represents a total of 5,643 features with index values of 0 or 1 derived from CTET protein vectors retrieved from DrugBank [22]. For protein identifiers, there were 83 carriers, 375 enzymes, 4,985 targets, and 200 transporters. The value 1 for an index "*i*" in a CTET protein vector means that the drug has a relation with the protein represented in the corresponding index. An example showing how drugs are represented as protein vectors is shown in Figure 2. The top panel in Figure 2 demonstrates the detailed breakdown of a CTET vector into individual protein vectors, and the bottom panel shows the combined CTET vector representation.

The accuracy of the drug interaction prediction method was evaluated by using known interactions. Then, the drug protein similarities were calculated. The similarity matrix M_2 generated for protein profile fingerprints and prediction matrix M_3 were used to assess the effectiveness of the approach.

For protein similarity calculations, the Russel Rao similarity measure was utilized. Russell-Rao similarity is a dot-product-based similarity measure in a range between 0 (minimum similarity) and 1 (maximum similarity) [24]. Similar to the Tanimoto coefficient, Russell-Rao represents the normalization of positive matches between vectors. Russell-Rao was chosen for measuring the similarity between drug interactions as it is a straightforward and suitable similarity measure with a 0 to 1 similarity range. That being said, other dot-product-based similarity measures with a 0 to 1 range can also be suitable for vector-based drug-interaction similarities. The Russel Rao similarity is calculated as

$$S_{\text{Russel - Rao}} = \frac{x^t y}{d}. \quad (4)$$

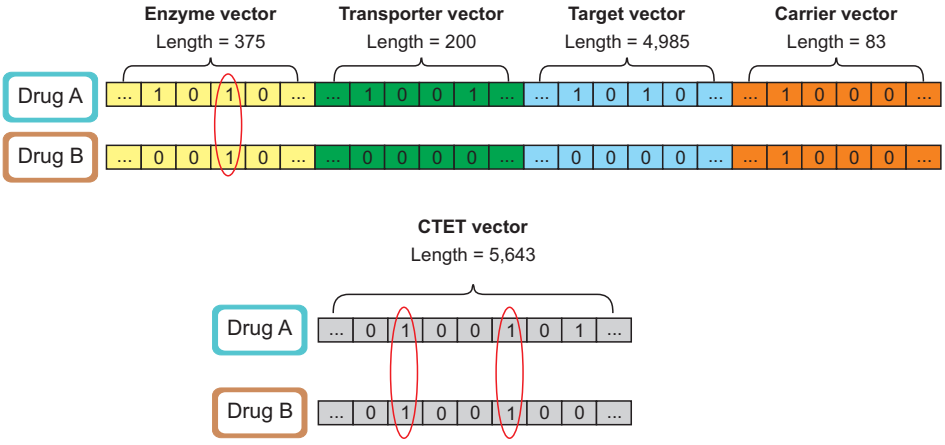

Figure 2. Carrier target enzyme transporter (CTET) vector representation of drugs. Adapted from Ferdousi R, et al. J Biomed Inform 2017;70:54-64 [6].

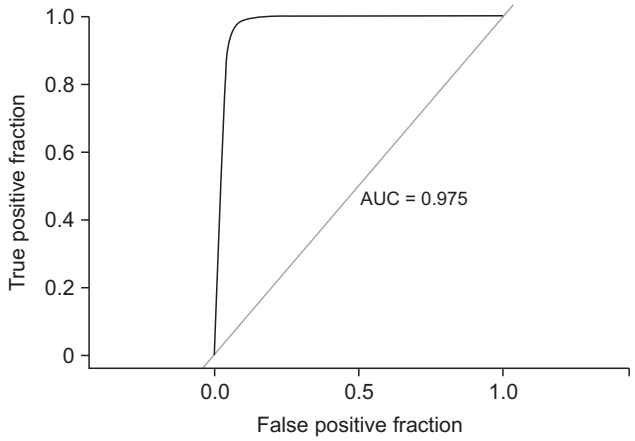

Figure 3. Receiver operating characteristic (ROC) curve for interaction profile fingerprint test data. AUC: area under the curve.

In Equation (4), the terms x, y denote binary feature vectors, and d is the vector size. The term $x^t y$ represents the positive matches between vectors.

3) Evaluation methods

In this study, we performed several evaluations on the collected dataset to assess the performance of the proposed prediction algorithm for different settings. Three different profile fingerprints were used to predict the DDIs in the evaluations, namely, interaction profile fingerprints, adverse effect profile fingerprints, and protein profile fingerprints. For the evaluations, R programming language version 3.5.0 and RStudio version 1.1.419 were used. The receiver operating characteristic (ROC) curves were used to measure the performance of the methods.

III. Results

1. DDI Predictions Based on Profile Fingerprints

1) Interaction profile fingerprint

First, the performance of our DDI prediction approach based on IPFs was evaluated. Out of 297,816 drug interaction pairs, approximately 85.02% of the dataset consisting of 253,201 interaction pairs was selected as training data, and the remaining 44,614 interaction pairs were considered as test data. The values in the prediction matrix were sorted based on the prediction values. Among the top 100 DDI predictions that were not in the training dataset, there were 49 true positive (TP) values.

When the threshold value was set to 0.7, there were 58 TP results among 100 random selections that were not in the training dataset. Similarly, when the threshold value was set to 0.4, the number of TP results was 8. For the test data, the ROC curve was generated as shown in Figure 3. It should be noted that the ROC curves and TP results presented in this paper were in line with those obtained in previous studies [1,8]. In our evaluations using IPFs, we obtained an ROC value of 0.975 for the 15% test data. It was similar to the ROC of 0.967 found in [8] and 0.98 in [1].

The results showed that the number of TPs was 44,603 among the 44,614 test data. This means nearly all the test DDIs were correctly predicted.

Despite these high TP results, the number of false positive (FP) values was quite large when the threshold value was considered as 0. A total of 2,880,989 FP results were detected among 3,355,345 possible pairs. The total number of possible pairs was calculated by 2591 × 2590 / 2. However, the performance shown in ROC curve indicated that if a certain threshold value higher than 0 was considered, the number of FP results would decrease significantly.

 https://doi.org/10.4258/hir.2020.26.1.42

2) Adverse effect profile fingerprint

An adverse effect profile fingerprint vector with 6,061 indexes was used in the similarity calculations because the evaluation dataset consisted of 6,061 unique adverse effects. According to the similarity calculations results, there were 1,675 drug pairs that had similar adverse effect values. This means that there were 1,675 × 2 = 3,350 index values that were greater than 0 in the similarity matrix M_2. Based on the prediction matrix M_3 results of the test data, we observed that there were 19,644 TP predictions out of 44,614 known test DDI pairs. The number of FPs was 403,765 among all possible 3,355,345 DDI pairs. The ROC curve for the adverse effect profile fingerprint test data is shown in Figure 4.

3) Protein profile fingerprint

Similarly, DDI predictions based on protein profile fingerprints were assessed using the testing dataset. The evaluation results showed that there were 43,430 TP predictions out of 44,614 pairs in the test data. This means that the vast majority of DDI pairs were correctly predicted as potential DDIs. On the other hand, there were 1,965,664 FP predictions among 3,355,345 possible pairs. The ROC curve based on the test data is shown in Figure 5. It should be noted that the ROC curves were similar to those reported in [1].

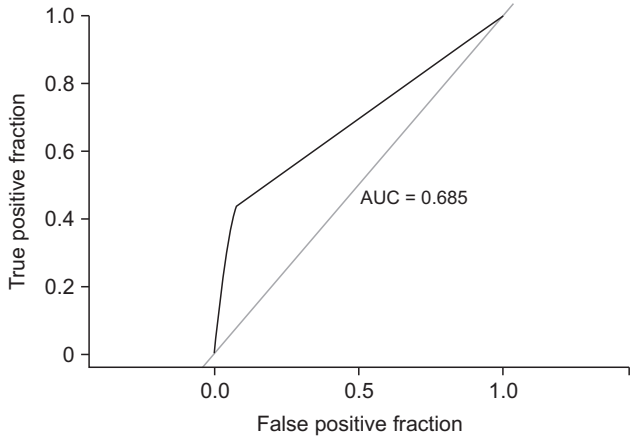

Figure 4. Receiver operating characteristic (ROC) curve for adverse effect profile fingerprint test data. AUC: area under the curve.

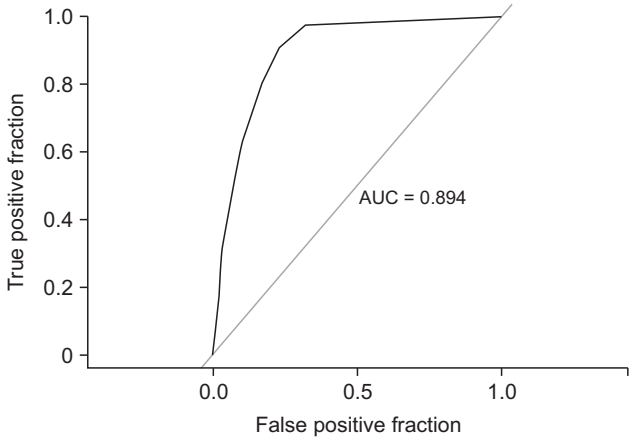

Figure 5. Receiver operating characteristic (ROC) curve for protein test data. AUC: area under the curve.

2. Evaluation of Protein Vector Similarity among Drugs with Known Interactions

The impact of protein similarity in the prediction of potential DDIs were also investigated. We observed that 125,404 pairs out of 297,816 known DDI pairs had protein similarities. In other words, the ratio of TPs was found to be 42%. It is significantly smaller than the TP ratio that was reported in [6], in which it was 72% among drugs with known interactions. The distribution of similarity measurements based on Russel-Rao calculations are shown in Table 1.

On the other hand, the protein similarity distribution ratios were quite similar to those in [6]. As stated in [6], "the higher similarity does not necessarily directly refer to a higher severity of adverse reaction". More specifically, two drugs with a few common proteins might have a significant DDI, whereas another pair of drugs sharing more common biological elements might not have an interaction. Further, the total number of drug pairs with protein similarities was 224,392 among 3,355,345 all possible pairs. If we consider that 125,404 pairs already had a potential DDI, there were 98,988 pairs that had protein similarities, but their DDI relation is unknown. If we assume that the ratio above 42% is applicable for these unknown pairs, then there might be approximately 41,500 drug pairs with potential DDIs. As seen in Table 2, the protein similarity distribution ratios were very similar to what was shown for known DDI pairs in Table 1.

IV. Discussion

In this study, a DDI prediction approach based on similari-

Table 1. Protein similarity distribution of known drug–drug interaction pairs

Degree	Range	Number of similar pairs
Low	0.0 < SRussell-Rao ≤ 0.001	116,519
Medium	0.001 < SRussell-Rao ≤ 0.0042	8,814
High	0.0042 < SRussell-Rao ≤ 0.0062	62
Very high	0.0062 < SRussell-Rao	9

Table 2. Protein similarity distribution of all possible drug pairs

Degree	Range	Number of similar pairs
Low	0.0 < SRussell-Rao ≤ 0.001	221,681
Medium	0.001 < SRussell-Rao ≤ 0.0042	2,708
High	0.0042 < SRussell-Rao ≤ 0.0062	3
Very high	0.0062 < SRussell-Rao	0

ties of profile fingerprints was developed. The performance of the approach was experimentally evaluated in detail. The evaluation results indicate that the profile fingerprint approach provides better prediction results compared to the approach of using direct protein similarity among DDI pairs based on TP results. In the profile fingerprint approach, nearly all the test data (15% of known DDI pairs) were successfully predicted with a few exceptions, whereas for the direct protein similarity approach, nearly half of the known interaction pairs were missed. Moreover, the prediction values were positively correlated with the probability of drug interaction for the first approach. On the other hand, the results of the second approach showed that higher protein similarity does not necessarily imply a higher probability of DDI. A drawback of the first approach is that it results in many more FP predictions. When certain thresholds based on ROC curves were considered, the number of FP estimations could be reduced significantly.

For profile fingerprint extraction and prediction, three different types of profiles were evaluated, namely, the drug interaction, adverse effect, and protein fingerprint profiles. Comparisons among those three methods indicated that drug interaction profile method achieved the best results. It was followed by the protein profile method, and then the adverse effect profile method. The similarity-based DDI prediction method using interaction profile vectors successfully predicted nearly all test data with an AUC of 0.975; however, there were still some errors. We think that it is possible to improve the prediction of DDIs by combining prediction methods. In future work, we plan to investigate the feasibility of developing of an improved hybrid prediction approach.

When considering the advantages and disadvantages of the methods, it should be noted that all three methods are based on drug interaction similarities and have the same computational approach; the main difference between the methods is the construction of fingerprint vectors for similarity calculations. The DDI prediction techniques based on drug interaction profiles and protein profiles achieved superior performance, but they produced more FP predictions. However, if certain thresholds based on ROC curves were considered, the number of FP estimations could be reduced significantly.

Furthermore, our evaluations verified that the IPF is a scalable and reliable source of DDI prediction calculation. This means that as the number of known DDI pairs increases and novel drugs emerge, this approach is likely to remain as a reliable DDI prediction model in the future. In comparison to the other two methods, the DDI prediction based on ADEs showed relatively poor performance. Our manual explorations showed that this was mainly due to missing ADEs. Furthermore, the lack of complete overlap between MedDRA and the Merged-PDDI dataset might have had an impact on the performance of the adverse effect profile method because MedDRA does not contain all drugs that are included in the Merged-PDDI dataset.

In conclusion, the results of this study demonstrated that using IPF vector extraction in prediction matrix calculations can achieve good performance in *in-silico* approaches for DDI prediction.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

ORCID

Selma Dere (http://orcid.org/0000-0002-9519-8165)
Serkan Ayvaz (http://orcid.org/0000-0003-2016-4443)

References

1. Vilar S, Uriarte E, Santana L, Lorberbaum T, Hripcsak G, Friedman C, et al. Similarity-based modeling in large-scale prediction of drug-drug interactions. Nat Protoc 2014;9(9):2147-63.
2. US Food and Drug Administration [Internet]. White Oak (MD): Food and Drug Administration; c2019 [cited at 2020 Jan 19]. Available from: http://www.fda.gov/.
3. 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.
4. Fowler S, Zhang H. In vitro evaluation of reversible and irreversible cytochrome P450 inhibition: current status on methodologies and their utility for predicting drug-drug interactions. AAPS J 2008;10(2):410-24.
5. Jonker DM, Visser SA, van der Graaf PH, Voskuyl

 https://doi.org/10.4258/hir.2020.26.1.42

RA, Danhof M. Towards a mechanism-based analysis of pharmacodynamic drug-drug interactions in vivo. Pharmacol Ther 2005;106(1):1-18.
6. Ferdousi R, Safdari R, Omidi Y. Computational prediction of drug-drug interactions based on drugs functional similarities. J Biomed Inform 2017;70:54-64.
7. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. Sci Transl Med 2012;4(125):125ra31.
8. 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.
9. Zhang W, Zou H, Luo L, Liu Q, Wu W, Xiao W. Predicting potential side effects of drugs by recommender methods and ensemble learning. Neurocomputing 2016;173:979-87.
10. Ayvaz S, Horn J, Hassanzadeh O, Zhu Q, Stan J, Tatonetti NP, et al. Toward a complete dataset of drug-drug interaction information from publicly available sources. J Biomed Inform 2015;55:206-17.
11. Noor A, Assiri A, Ayvaz S, Clark C, Dumontier M. Drug-drug interaction discovery and demystification using Semantic Web technologies. J Am Med Inform Assoc 2017;24(3):556-64.
12. Vilar S, Harpaz R, Uriarte E, Santana L, Rabadan R, Friedman C. Drug-drug interaction through molecular structure similarity analysis. J Am Med Inform Assoc 2012;19(6):1066-74.
13. Vilar S, Uriarte E, Santana L, Friedman C, Tatonetti NP. State of the art and development of a drug-drug interaction large scale predictor based on 3D pharmacophoric similarity. Curr Drug Metab 2014;15(5):490-501.
14. Liu M, Wu Y, Chen Y, Sun J, Zhao Z, Chen XW, et al. Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. J Am Med Inform Assoc 2012;19(e1):e28-35.
15. Tatonetti NP, Fernald GH, Altman RB. A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports. J Am Med Inform Assoc 2012;19(1):79-85.
16. Harpaz R, Haerian K, Chase HS, Friedman C. Statistical mining of potential drug interaction adverse effects in FDA's spontaneous reporting system. AMIA Annu Symp Proc 2010;2010:281-5.
17. Ibrahim H, Saad A, Abdo A, Sharaf Eldin A. Mining association patterns of drug-interactions using post marketing FDA's spontaneous reporting data. J Biomed Inform 2016;60:294-308.
18. Hudelson MG, Ketkar NS, Holder LB, Carlson TJ, Peng CC, Waldher BJ, et al. High confidence predictions of drug-drug interactions: predicting affinities for cytochrome P450 2C9 with multiple computational methods. J Med Chem 2008;51(3):648-54.
19. Percha B, Garten Y, Altman RB. Discovery and explanation of drug-drug interactions via text mining. In: Altman RB, Dunker AK, Hunter L, Murrey TA, Klein TE, editors. Biocomputing 2012. Singapore: World Scientific Publishing; 2011. p. 410-21.
20. Tari L, Anwar S, Liang S, Cai J, Baral C. Discovering drug-drug interactions: a text-mining and reasoning approach based on properties of drug metabolism. Bioinformatics 2010;26(18):i547-53.
21. Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). Drug Saf 1999;20(2):109-17.
22. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res 2006;34(Database issue):D668-72.
23. Fligner MA, Verducci JS, Blower PE. A modification of the Jaccard–Tanimoto similarity index for diverse selection of chemical compounds using binary strings. Technometrics 2002;44(2):110-9.
24. Willett P. Similarity-based approaches to virtual screening. Biochem Soc Trans 2003;31(Pt 3):603-6.