Computerized techniques pave the way for drug-drug interaction prediction and interpretation

Reza Safdari1†, Reza Ferdousi1,2†, Kamal Azizihersi3, Sharareh R. Niakan-Kalhori1, Yadollah Omidi1†

1 Department of Health Information Management, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran
2 Research Center for Pharmaceutical Nanotechnology, Faculty of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran
3 Department of Mathematical Sciences, University of Tabriz, Tabriz, Iran

Abstract
Introduction: Health care industry also patients penalized by medical errors that are inevitable but highly preventable. Vast majority of medical errors are related to adverse drug reactions, while drug-drug interactions (DDIs) are the main cause of adverse drug reactions (ADRs). DDIs and ADRs have mainly been reported by haphazard case studies. Experimental in vivo and in vitro researches also reveals DDI pairs. Laboratory and experimental researches are valuable but also expensive and in some cases researchers may suffer from limitations.

Methods: In the current investigation, the latest published works were studied to analyze the trend and pattern of the DDI modelling and the impacts of machine learning methods. Applications of computerized techniques were also investigated for the prediction and interpretation of DDIs.

Results: Computerized data-mining in pharmaceutical sciences and related databases provide new key transformative paradigms that can revolutionize the treatment of diseases and hence medical care. Given that various aspects of drug discovery and pharmacotherapy are closely related to the clinical and molecular/biological information, the scientifically sound databases (e.g., DDIs, ADRs) can be of importance for the success of pharmacotherapy modalities.

Conclusion: A better understanding of DDIs not only provides a robust means for designing more effective medicines but also grants patient safety.

Introduction
Errors are plausible in every medical related procedures. They can occur at a rate of 5.7%, while with a good management and predictive and preventive forethoughts such errors can be reduced and prevented up to 50%. Drug-drug interactions (DDIs) are main cause of pharmaceutical errors, especially in elderly patients and patients with poly-therapy procedures, the prevalence of DDIs is about 20-40 percent. Of note, DDIs occur when function of a drug changed or influenced by presence of another drug. They can be categorized as two major classes including (a) pharmacokinetics (PK) DDIs and (b) pharmacodynamics (PD) DDIs. PK-DDIs are attributed to the interactions of drugs during absorption (e.g., allopurinol enhances the oral bioavailability of mercaptopurine by inhibiting metabolizing enzyme), distribution (e.g., interaction of warfarin with other protein-bound anticancer drugs such as paclitaxel and etoposide), metabolism (e.g., interactions of drugs that are substrate to metabolizing enzymes such as CYP3A4 and CYP2D6) and elimination (e.g., trimethoprim enhances plasma concentration of methotrexate to toxic level causing nephrotoxic impacts) processes, in which various enzymes and/or transporters could be involved. Based upon the rate of pharmaceutical errors as well as its financial, social and health costs, prevention of these errors has become a global priority. It should be pointed out that one out of 10 cases of the unexpected hospital admissions of patients with cancers are related to adverse drug reaction (ADR) problem(s). In elder people, most of ADRs are resultant form DDIs, in which about 10% of all studied ADRs are not related to DDIs. Further, usually about 20-30% of patients hospitalized have been shown to experience ADRs related to DDI. On the other hands,
one out of every five patients exposed to a potential DDI experience at least a related ADR.\textsuperscript{9}

It is deemed that the prediction of DDIs, and hence their prevention, can be of great benefits to patients and health care systems. Recent researches have shown that the number of DDIs are closely related to the number of prescribed drugs, leading to an elongated hospitalization and consequently high expenditure.\textsuperscript{10}

The main focus of the current study is to investigate the place of \textit{in silico} approaches in DDIs’ prediction and prevention. We will also provide some insights on applications and outcomes of utilization of computerized techniques in handling DDIs phenomena.

**Hurdles and promises**

One of the approaches to study DDIs is based on common methods using PK parameters that describe what body does to a designated drug molecules. PK based DDIs can be resultant from drugs interactions during the absorption, distribution, metabolism and elimination (ADME) processes of drug molecules that are analyzed and studied to understand and examine DDIs. Despite precision and usefulness of PK based methods for analysis of DDIs, this approach may fail to reveal the actual mechanism(s) of the DDIs and ADRs. The reason lies in the fact that the end point about drugs usually is a serious side effect instead of a measurable changes on drug concentration.\textsuperscript{11}

Also there are some studies that are based on clinical evidence and reveals the experienced DDIs,\textsuperscript{12-15} and these types of studies are useful to prevent from experienced DDIs. Nonetheless, they cannot be helpful in terms of other unexperienced or unreported DDIs.

It should be noted that the number of discovered drugs continues to rapidly grow and possible occurrence DDIs and ADRs can cause high costs and burdens on health systems. Therefore, DDIs prediction, prior to the clinical experiences, can offer great benefits to pharmaceutical companies in drug designing as well as physicians and patients.

**Impacts of computerization and modeling**

Considering costs and doability of \textit{in vitro} and \textit{in vivo} experiments, there is a high tendency to \textit{in silico} studies among researchers. Nowadays, a large number of researchers capitalize on the computerized modeling and simulations in order to examine hypotheses and understand mechanisms related to drug actions and side effects.\textsuperscript{16} The \textit{in silico} modeling approaches have improved the health related sciences as well as pharmaceutical industry, in large part because of optimality and cost-effectiveness of the computerized techniques. Likewise, pharmaceutical scientists have exploited the computerized modeling methods in different aspects of pharmaceutical sciences including drug design and discovery, PK and/or PD analyses. Accordingly, in addition to commercial packages, there are many open source and free tools that have been developed by universities or group of researchers to satisfy educational and scientific needs. Further, genomics, proteomics, metabolomics and molecular data related to drugs are great resources of information that are available in various molecular and biological databases, which can be used to simulate and understand drugs and their interactions. Among them, there exist growing algorithms and methods of artificial intelligence such as machine-learning and data-mining techniques that have great impacts on discovery and prediction of DDIs. Similarly, there exist many success stories in terms of biological simulations and modelings in particular for drug design.

**Successful stories**

In drug development phases,\textsuperscript{17} as shown in Fig. 1, the advantages of computer-aided drug design (CADD) are obvious. The blue color shows the stages that CADD tools are involved. In each step, various computerized techniques are used.

Target identification is considered as the foundation of the modern drug discovery process. The selection of biological targets is typically based upon some key experimental criteria including (a) disease linkage (e.g., mutation(s) related to a certain disease), (b) mechanistic aspects (e.g., protein involved in a regulatory pathway related to initiation and/or progression of a specified disease), and (c) genetic screenings using high throughput techniques such as DNA microarray\textsuperscript{18-21} and next generation RNA sequencing. High throughput screening techniques have experimentally been utilized for the generation of phage peptide/antibody libraries as well as the synthetic oligonucleotide libraries used for the selection of aptamers with promising applications in drug discovery.\textsuperscript{22-26} In addition, the target must be druggable and drug-target interaction result is meaningful impact(s). Target prediction is based on some key techniques such as bioinformatics, reverse docking and protein structure prediction. As shown in Fig. 2, the target druggability\textsuperscript{27} and tool compound design are the key tasks related to the target validation stage, lead discovery stage achievable by some approaches such as library design, docking scoring, de novo design, pharmacophore and target flexibility. Of note, interaction of a designated drug to a biological/
molecular marker (the so-called druggable target) must change the functionality of the biological target, resulting in certain therapeutic benefit(s) to a specified disease with accountability. Up to now, the quantitative structure–activity relationship (QSAR), 3D-QSAR and structure based optimization techniques for lead optimization stage and finally in silico ADME-Toxicity (ADMET) prediction and physiologically-based PK (PBPK) simulations have been employed in preclinical test stages. Of note, in silico target prediction work-flow, as portrayed in Fig. 2, shows that, in order to use computational models for the prediction of protein targets of small molecules, target class models for proteins need to be constructed. It should be pointed out that modeling of multiple thousands of proteins takes couple of hours for circular fingerprints in combination with naive Bayes classifiers as simple probabilistic classifiers. In the next step, for a chemical structure as an input (panel A) can be annotated with most probable macromolecular interaction partners (panel C). Besides, it should be noted that the use of social media among people has become extremely popular, upon which people share their personal health-related information. It is envisioned that such information would be a great resource for the pharmacovigilance researches, by which ADRs and DDIs can be detected even though the obtained information could be preliminary findings and considered as informal user-expressed concepts that are non-technical. Natural language processing (NLP) techniques can be used to extract structured and important information from the contents of social media. Fig. 3 exemplifies the power of machine-learning based NLP techniques named ADRmine.

Molecular similarity based detection of DDIs is another successful computational approaches to identify potential interaction(s). In this approaches, molecular similarity between drugs is computationally compared. For instance, if drugs A and B known as DDI pair, then, drug that is chemically similar to drug A or B can be considered as a drug that can interact with drug A or drug B. The prediction of DDIs through protein-protein interactions (PPIs) network is another successful methodology. As demonstrated in Fig. 4, drugs can be connected to the PPI network by their target protein(s). Existence of shared proteins between two drugs and number of links reveals possible interaction(s) between two drugs.

**Common computerized techniques applicable in DDIs**

Of the computerized strategies and algorithms used in pharmaceutical sciences for the prediction of DDIs, networking techniques such as network analysis or network mining have recently been shown to be emerging technologies. Further, different biological networks are now available, which can satisfy information needs for the prediction and interpretation of DDIs. Classification and clustering techniques are another category of techniques that have been being used for the supervised and unsupervised learning methods to foreseeing the biological facts. Text-mining and data-mining approaches consist of many different techniques can also be used for the pattern recognition and knowledge discovery from available big biological data bases and texts. Altogether, it can concluded that the integration of mathematic, biology, computer science and visualization techniques provides
comprehensive modeling and simulation tools in different aspects of life science at molecular, cellular and dynamic behavior levels. Common modeling and simulation used in pharmaceutical science is ADMET modeling, while there are registries and other information systems such as the FDA adverse event reporting system (FAERS) that have been used to benefit patients to some points.

Table 1 lists the available computerized techniques that are applicable in the management of DDIs. We will briefly discuss some of these approaches that are most relevant in terms of DDIs analysis.

Machines and DDIs
Currently more than 8,200 type of drug substances are available, including more than 2,300 FDA-approved drugs and over 6,000 experimental drugs. Considering such huge number of drug substances, their possible interactions between pairs of drugs can result in emergence of huge DDIs network. If we consider just reported DDI pairs network for FDA approved drugs, then we will face with a network of 1,441 nodes and 45,530 edges (Fig. 5). Intriguingly, if we consider all possible interactions for all drugs, then the network will consist of over 2,000,000 edges. Taken all, capitalizing on such approach, we can get huge amount of data, for which the problem space is big enough to employ computerized techniques for analyzing DDIs. Of the computational \textit{in silico} approaches, similarity measures are the core element of most pattern recognition, classification and clustering algorithms. In fact, the binary representation vectors favor complex matters such as DDIs to be more computer interpretable. Therefore, different binary feature vector similarity systems have been developed. While the inner-product based measures consider only positive matches in vectors (e.g., Jacquard, Dice), some other measures are based on both negative and positive matches. These measures consider equal or variable weights for the positive/negative matches. Fig. 6 shows different categories and measures for binary feature vector similarity.

Applications of machine based DDI analysis
Adding all properties of drugs biological functions to aforementioned network of DDIs make it so big, yet more informative, so that one can use it to analyze DDIs. As shown in Fig. 7, there exist some key steps for applications of computerized techniques in management of DDIs including (a) detection, (b) prediction, (c) interpretation,
(d) alarming, and (e) determination of strength and seriousness of DDI.

**The key step is DDI detection**

In most of patient’s treatment procedures, co-admiration of drugs is an inevitable decision. However, lack of knowledge about the exact functional behavior(s) of administered drugs and uncompromising side effects results in inadvertent drug interactions. To gather all necessary information, some DDIs registry systems have been established to collect information about experienced DDIs that can be used by the health care

---

**Table 1. List of the available computerized techniques applicable in the management of DDIs**

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Description</th>
<th>Examples/ Methods</th>
</tr>
</thead>
</table>
| **Network**            | Mathematical analysis based on graph theories for biological system in terms of nodes and edges beneficial for understanding and discovery in biology. | • Protein–protein interaction networks  
 • Gene regulatory networks (DNA–protein interaction networks)  
 • Gene co-expression networks (transcript–transcript association networks)  
 • Metabolic networks  
 • Signaling networks  
 • Neuronal networks  
 • Food webs  
 • Between-species interaction networks  
 • Within-species interaction networks |
| **Classification**     | Two groups of statistical and Machine learning algorithms, that provides possibility of predicting unseen biological behaviors according to known biological facts. | • Logistic regression  
 • Naïve Bayes  
 • Support vector machine  
 • Neural networks  
 • Decision trees  
 • Random Forest |
| **Clustering**         | Clusters in biology used for grouping a set of biological elements in such a way that elements in the same group (cluster) are more close/similar to each other. It helps to stratifies and grading biological evidences too. | • Connectivity-based clustering (hierarchical clustering)  
 • Centroid-based clustering  
 • Distribution-based clustering  
 • Density-based clustering  
 • Recent developments |
| **Data/text mining**   | Knowledge discovery in databases (KDD) vastly used to pattern discovery and knowledge discovery among large amount of biological data/texts. | • Anomaly/outlier/change detection  
 • Association rule learning  
 • Factor analysis  
 • Sequence mining  
 • Structured data analysis  
 • Text mining  
 • Agent mining |
| **Modeling/simulations** | A task in systems biology and mathematical biology, with the goal of visualization and computer modelling of biological systems including cellular subsystems. | • Molecular modelling  
 • Model of the immune system  
 • Pharmacokinetics and pharmacodynamics modeling  
 • Brain modelling  
 • Protein folding  
 • Virtual liver |
| **Registries and clinical systems** | All software systems used in health pharmaceutical industries to provide information retrieval resources. | • ADR/DDI registries  
 • Drug monitoring systems  
 • Pharmaceutical information systems  
 • Computerized physician order entry systems |

---

**Fig. 6. Different categories and measures for binary feature vector similarity.**
providing specialists, while unknown DDIs are often neglected causing undesired outcomes. Thus, there exists an intensive need for the development of automated DDI detection systems. Two types of resources can be used for the automated DDI detection systems, including (a) medical records and (b) social networks. Medical records contain information about received medication and outcomes of the treatment, for which huge amount of data is available by the aid of electronic medical records. Accordingly, different computational approaches such as data-mining techniques, machine-learning techniques and pattern recognition protocols can be applied to these data to detect and collect all emerged DDIs. Further, crawling web and social network mining of user posts and comments about their experienced side effects of medications may lead to some new DDI discovery though such information needs to be revalidated.

Computerized modeling for management of DDIs
By prediction of DDIs, especially for the investigational new drugs (INDs), emergence of ADRs and DDIs can be circumvented resulting in much more improved clinical corollaries and patient safety. Computational PK and/or PD modeling of drugs behaviors within body and use of drugs associated biological signaling pathways and elements such as drugs target proteins and target ligands help to discovering new DDIs. Patterns and algorithms such as artificial neural networks (ANN) can reveal the hidden prototypes of DDIs that may result in ADRs and hence failure of the pharmacotherapy modality. Given that the complexity of human metabolism makes interpretation and etiology of DDIs an intricate task, computer based modeling an in silico approaches provides a set of high throughput information in favor of understanding DDIs. However, the resultant DDIs need to be fully understood and interpreted prior to putting them into clinical practice. For example, data found for the cellular behaviors and drug biological functions for a designated drug must be interpreted in terms of the etiology of the interactions occurred.

DDI alarming system
Computer-based DDIs prediction of co-administrated drugs is useful in their prevention, and accordingly there is a tendency towards computerized prescription in health industry. Use of the computerized physician order entry (COPE) can assist medical professionals to avoided DDIs as much as possible. However, as part of public education systems, the DDI detection platforms could be used by patient who has to undergo multiple therapies. Further, the pharmaceutical companies are another users of DDI systems to inform their customers upon possible occurrence of DDIs for a designated drug.

Prediction of outcome and seriousness of DDI
Despite knowledge about known DDI, in some cases co-administration of several drugs is inevitable. For example, in the case of chemotherapy or patients who have to use a special drug for a long time period or even permanently, some DDIs seem to be deliberately neglected. In these cases, having a good estimation from the outcomes and seriousness of interactions is vital for the success of treatment modality. Besides, some interactions may directly endanger the patient’s life. As a result, estimating the outcome, diversity and risk of drug interactions with the aid of computational and algorithmic approaches could result in higher patient safety in case of multi therapies.

Final remarks and outlook
Many clinical complications can emerge from DDIs after inaccurate combination therapy. To avoid such problems, several steps can be helpful including (a) use of computerized models for prediction of possible DDIs qualitatively and quantitatively, (b) careful design of treatment modality, (c) interpretation of patients pharmacotherapy profile, and (d) translation of preclinical data to rationalized clinical use. Ideally, the main objective of any treatment modality is to maximize the therapeutic effects with minimized
side effects. Thus, in addition to the in vitro and in vivo correlation approaches, biological simulation and modeling can be of great help in terms of drug discovery. Accordingly, accurate computational modeling and simulation helps to foster the drug discovery approaches and in particular to understand the DDIs mechanisms.

The relation between clinical and molecular information can be revealed with the aid of data graph and data networks constructed with different type of information. For instance, we are now aware of existence of direct relations between pharmacogenetics/pharmacogenomics and drug-induced toxicities such as nephrotoxicity in renal transplant recipients.\(^{46}\) Therefore, to attain rational treatment modalities in various diseases, all possible detrimental DDIs/ADRs, drug-protein interactions, pharmacogenetics and drug metabolite interactions must be taken into consideration. However, from translational and personalized medicine viewpoints, effective standard protocols need to be codified and applied for these matters.\(^{47}\)

Taken all, incorporation of data science with computerized techniques of modeling and simulation brings new facilities for researchers to predict and understand DDIs and hence avoid the occurrence of ADRs that may jeopardize patient safety especially when combination therapy is inevitable. This could be deemed as one of the main steps in terms of personalized medicines. Finally, careful considerations upon the potential undesirable drug interactions in drug development and administration processes (i.e., drug design and discovery, drug prescription and selection of the most effective treatment modalities in multi-therapies), will significantly warrant the success of treatment modalities, improve the patients’ safety and lower the burden of health systems.

Ethical approval
There is none to be disclosed.

Competing interests
Authors declare no competing interests.

Acknowledgments
Authors like to acknowledge the Research Center for Pharmaceutical Nanotechnology at Tabriz University of Medical Sciences for the financial support.

References


