# Drug Side Effects Prediction

RP KELUM PRIYADARSHANA

Reg. No : 2011CS110
Index No : 11001102

**Supervisor: Mrs. M.W.A.C. Rupika Wijesinghe**

Tools: *GrindEQ, LyX, Mendeley, ResearchOMatic, TeX Studio*
Reference Style : IEEE

December 7, 2014

# 1 Chapter 1

## 1.1 Abstract

Drug Side-Effects or Adverse Drug Reaction (ADR) has been become a major public health concern. It is the one of the main causes of failure in the process of drug development, and of drug withdrawal once has reached the market. Harder ADRs that go undetected until the post-marketing phase of a drug often lead to patient morbidity. Faultless prediction of potential ADRs is required in the entire life cycle of a drug, including early stages of drug design, different phases of clinical trials, and post-marketing surveillance. Within this review discuss about nine methods. Those are, Data description, Ordinary canonical correlation analysis (OCCA),Sparse canonical correlation analysis (SCCA),Prediction of side effect profiles for new molecules, Enrichment analyses of targeted proteins, Random assignment (Random),Nearest neighbor (NN) and Support vector machine (SVM)

Above methods can categorize into three categories. As chemical, biological, and phenotypic properties of drugs .this is a review of Computational Approaches in Drug side effects prediction.

# Contents

## 1.2 Introduction

The last few decades have witnessed a steady increase in drug prescriptions for the treatment of -biometric markers rather than overt physiological symptoms. Today, people regularly take multiple drugs in order to normalize serum levels of biomarkers such as cholesterol or glucose, or to reduce blood pressure. All drugs have side effects, which are sometimes debilitating or even life-threatening. When a person taking multiple drugs experiences a new symptom, it is not always clear which, if any, of the drugs or drug combinations are responsible.Predicting and summing the side effects of a new drug during its developmental phase remain important to the drug's overall commercial success. Drug Side effects are responsible for a significant Number of cases where pre-marketed drugs fail during clinical trials. Recognizing the underlying mechanisms of side effects is a challenging task, often because of the drugs' multiple effects on a Biological system. Number of drugs are small compounds that target and interact with proteins to induce perturbations in the proteins network.

The mostly used approach to identify possible side effects for a drug is to use its chemical structure information, based on the observation that drug chemical structures can direct the ligand promiscuity toward protein targets.However, chemical structure-based methods can't supply any biological interpretations regarding the underlying mechanisms at a molecular interaction level.

Governments of most of the Countries spend billions of dollars on prescription drugs every year, resulting in a significant Healthcare burden from adverse drug reactions (ADRs). ADRs are defined as those unintended and undesired responses to drugs beyond the anticipated therapeutic effects during clinical use at Normal doses. It is estimated that 6%-7% of hospitalized patients experience severe ADRs each Year with a potential of 100 000 deaths, which makes it the fourth largest cause of death in the USA.There are multiple international agencies and programs to monitor the safety of approved medical treatments, surveillance programs such as the U.S. Food and Drug Administration's (FDA's) and Adverse Event Reporting System (AERS) are often difficult for general public to reach.Many drugs approved by Food and Drug Administration (FDA) were recalled each year after some unexpected side effects were discovered.

As well as Identifying drug-drug interaction potential early in drug discovery and development is important because drug-drug interactions can cause life threatening changes in drug levels.Early discovery of potential drug-drug interactions for a compound expedites the decision to eliminate that compound from consideration, thus lowering the cost of drug discovery.During the past decade, many compound databases have been constructed, such as Kyoto Encyclopedia of Genes and Genomes (KEGG) and STITCH (Search Tool for Interactions of Chemicals).

### 1.2.1 Importance of studying drug side effects prediction

Bioinformatics is an interdisciplinary field that develops methods and software tools for understanding biological data. As an interdisciplinary field of science, bioinformatics combines computer science, statistics, mathematics and engineering to study and process biological data. Bioinformatics is playing a vital role in development of society by providing quick information and making research fast. It is using today's computer technology and biological research together very efficiently. This field is going to generate more opportunities in future for all people working in different areas.
Bioinformatics tools can be used for three purposes.
1) Protein sequence can be determined by DNA sequencing.
2) If protein structure has to be determined then knowledge of protein sequencing is required.
3) A protein structure enables the determination of protein function.
If there is complete knowledge of these three steps then it is easy to understand that what the biology of an organism is.
***The Research Problem:*** **What are the prevailing methods on Drug Side Effects Prediction?**
Can define Computational Approaches in Drug Side Effects Prediction as Bioinformatics; Bioinformatics is a multi dimensional section and requires people from different working areas. It is the cognation of biology with IT and is a novel emerging section that aggrace in collecting, linking, and manipulating different types of biological data to discover new biological insight. Scientist and researchers spend the whole life in inventing things for human advantages. After few years of improvement, have collected huge amount of valuable data from the experiments all over the world and still this collection is maintain and will always maintain for the better development of human being. Sometimes, need to repeat the old research because either it is hard to obtain aged data or do not know whether it exist or not; this wastes the valuable time. Bioinformatics also helpful in digitalizing the information available on paper or in the form of specimen, so that with the help of Internet it could be easily available to everyone everywhere. Bioinformatics makes information readily available by collecting, linking, and operating.

### 1.2.2 Motivation

Drug Side Effect Prediction is very essential and useful area to study. There are many website like www.Drugs.com and www.Drugwatch.com gives Comprehensive side effect report and adverse reaction information for over 5000 drugs and medications
The mission at Drugwatch.com is to inform people about dangerous drugs and medical devices. Also facilitate people figure out if filing a lawsuit is right for them regarding drugs.
Besides there are crowd sourcing websites and forums that can upload details about effect of patients got when are using it. Such as www.rxisk.org
Experts estimate that only 1–10% of "serious" adverse events (those causing hospitalization, disability, or death) are ever reported.
In Europe, "Adverse drug reactions present a major public health burden. It is estimated that ADR cause 5% of all hospital admissions, 197000 deaths per year in the EUROPE, Total societal cost of 79 euro billion in the EUROPE
There are few organizations and government's authorities that can Collect/Complain detail about drugs and the effects locally and globally. In locally State Pharmacy Corporation (SPC), Cosmetics Devices and Drugs Regulatory Authority (CCDRA). In USA Drugs and Foods Control by FDA which is Food and Drug Administration. Further in Sri Lanka there are some units established within state hospitals to inquiry about medicines. So patient can see if there side effects or allergies involve with such medicines.
There are some ongoing projects those researching drug side effects such as www.projectknow.com
Medical industry still producing drugs with side effects and also those producing alternative drugs with expensive prices. For an example for the cold disease there is drug called Piriton (chlorphenamine) when patient use it , will get headache , sleep and etc. But as replacement to that medicine people use to familiar Cetirizine Hydrochloride Tablets.
When medicine generally use, lot of medicines are known by the trade name not from the scientific name. For an example for "Paracetamol" people use "Panadol" , for "Ketoconazole" which use for Antifungals ,people address as "MyCoral"and for "CETIRIZINE" people use "Cipla". This industry is much useful and very responsible area to study. Each general person should have general idea about side effect of a drug before addict use them.
Throughout this literature survey have been discussed following topics. Predicting Drugs Side Effects Based on Chemical and Chemical Interactions and Protein and Chemical Interactions, High Confidence Predictions of Drug-Drug Interactions: Predicting Affinities for Cytochrome,P450 2C9 with Multiple Computational Methods, Spontaneous Drug adverse Effect Discovery from Online Patient Submitted Opinions, Classification and Mechanisms for Adverse Drug Reactions, Patient online Forums on Drug Side-Effects, Chemical fragment-based approaches for Predicting Drugs Side Effects, Relating drug–protein interaction network with drug side effects, Large amount prediction of effect drug reactions using chemical, biological, and phenotype characteristics of drugs.

### 1.2.3 List of Abbreviations and Acronyms

**DNA** Deoxyribonucleic Acid

**ADR** Adverse Drug Reaction

**KEGG** Kyoto Encyclopedia of Genes and Genomes

**STITCH** Search Tool for Interactions of Chemicals

**SPC** State Pharmacy Corporation

**CCDRA** Cosmetics Devices and Drugs Regulatory Authority

**OCCA** Ordinary canonical correlation analysis

**SCCA** Sparse canonical correlation analysis

**NN** Nearest neighbor

**SVM** Support vector machine

**FDA** Food and Drug Administration

**AERS** Adverse Event Reporting System

**PPI** Protein-Protein Interactions

**GO** Gene Ontology

**NCBI** National Center for Biotechnology Information

# 2 Chapter 2

## 2.1 Computational Approach to Drug Side Effects Prediction

### 2.1.1 Predicting adverse side effects of drugs

Bio-molecular network and gene annotation information can significantly improve the predictive accuracy of ADR of drugs under development. The use of PPI networks can increase prediction specificity and the use of GO annotations can increase prediction sensitivity. Using cardiotoxicity as an example, it can be identified that cardio toxicity-related proteins among drug target expanding PPI networks. The systems pharmacology approach that developed in this study can be generally applicable to all future developmental drug ADR assessments and predictions. This is giving an idea that It is time for drug developers to design new and accurate models to assess unwanted side effects and drug actions before costly human clinical trials.

For comparisons, the prediction models in this study include two independent procedures:
1) machine learning- support vector machines (SVM)
2) Statistical modeling - logistic regression.

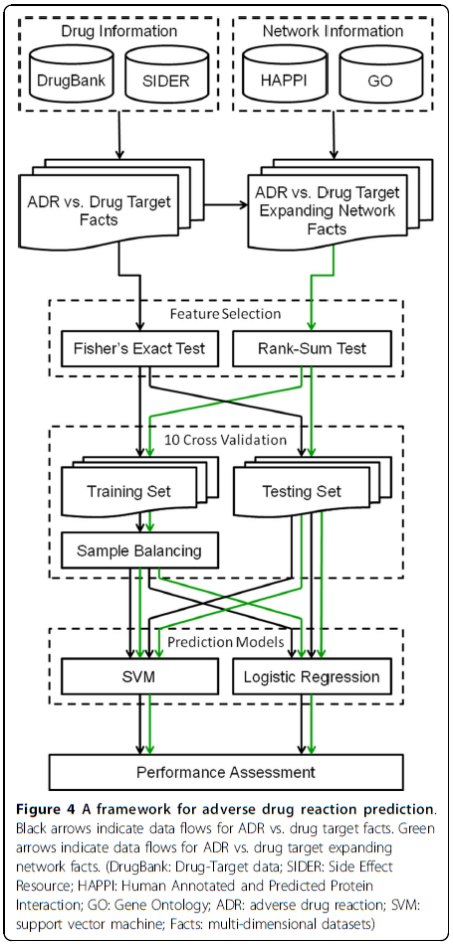



**Figure 4 A framework for adverse drug reaction prediction.** Black arrows indicate data flows for ADR vs. drug target facts. Green arrows indicate data flows for ADR vs. drug target expanding network facts. (DrugBank: Drug-Target data; SIDER: Side Effect Resource; HAPPI: Human Annotated and Predicted Protein Interaction; GO: Gene Ontology; ADR: adverse drug reaction; SVM: support vector machine; Facts: multi-dimensional datasets)

### 2.1.2 Automatic Drug Side Effect Discovery from Online Patient-Submitted Reviews: Focus on Statin Drugs

Within this research article, described the vision of a Web based database providing potential users with a rich facility for exploring the association of prescription drugs with possible side effects. Used the basic strategy of comparing word frequency distributions between two databases as a means to uncover statistically salient phrase patterns. The efforts focused on statin drugs, as these are a widely prescribed medication with diverse side effects. Through standard statistical log likelihood ratio estimation, have shown that statin drugs are very strongly associated with muscle pain and weakness, and that there is as well a statistically significant association between statin drugs and several debilitating diseases, such as ALS, Parkinson's disease, rhabdomyolysis, and heart failure. Many of the findings are supported by the research literature on statins.

**Data description**

To learn the underlying associations between side effects and drug usage from patient-provided reviews, drug reviews have been collected from three drug discussion forums ("AskPatient.com," "Medications.com" and "WebDB.com") which allow users to post reviews on specific drugs and share the experiences. Table 1 gives the statistics on the review data collection. A total of 8,515 statin reviews were collected from the three data sources. Also collected 105K drug reviews from the AskPatient.com, on drugs to treat a broad range of problems such as depression, acid reflux disease, high blood pressure, diabetes, etc. This set includes reviews for non-statin cholesterol lowering drugs.

Table 1. Statistics on drug review data collection.

| Data source | Number of Statin reviews |
|---|---|
| AskPatient.com | 2,647 |
| Medications.com | 4,162 |
| WebMD.com | 1,706 |
| Total | 8,515 |

To build and evaluate the proposed ADR-prediction model, have been used data from SIDER.29 SIDER presents an aggregate of dispersed public information on drug side effects and indications. SIDER extracted information on marketed medicines and the recorded ADRs from public documents and package inserts, which resulted in a collection of 888 drugs and 1385 side-effect keywords. There are a total of 61 102 associations between drugs and side-effect terms in SIDER, and each drug has an average of 68.8 side effects. The chemical structures of drugs were collected from PubChem,30 31 biological properties were obtained from the DrugBank32e34 and KEGG,35e37 and phenotypic data were from SIDER.29 To link these databases, have been mapped drugs in SIDER to DrugBank.32e34 Fifty-six drug names from SIDER could not be mapped to the respective Drug Bank IDs, resulting in a final dataset of 832 drugs, each of which has a 'Yes' or 'No' label for each of the 1385 side effects, indicating whether a drug has a specific side effect or not. The PubChem, Drug Bank, and KEGG databases comprise data that are available during chemical and animal trials, and are available before or during phase I clinical trials. However, the phenotypic data from SIDER are collected from phase I all the way through phase IV post-marketing surveillance. As such, this work describes a surveillance framework that allows pre-human association detection all the way through pre-marketing clinical trial phases to post-marketing surveillance.

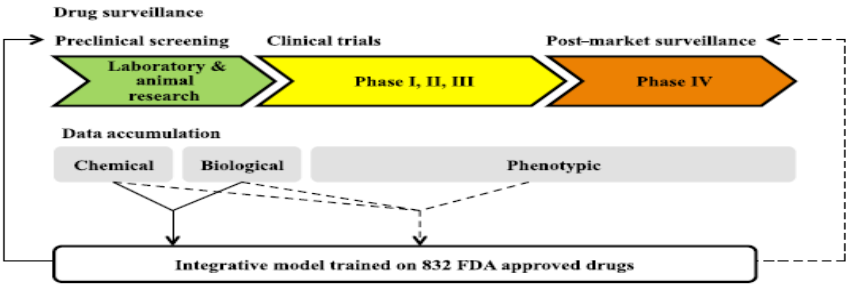



**Figure 1** Overview of the proposed framework for drug surveillance. Different combinations of features can be used for different phases of drug surveillance. Chemical structures and relevant proteins of drugs can be combined to predict potential adverse drug reactions (ADRs) in the early phase of drug development. As drug indication and other ADRs become available, they can be integrated with chemical and biological information for post-market surveillance.

Figure 1 provides a visualization of the proposed ADR-prediction framework at different phases of drug surveillance. Ordinary canonical correlation analysis (OCCA) Sparse canonical correlation analysis (SCCA) Prediction of Drug side effect profiles for new molecules Enrichment analyses of targeted proteins Random assignment (Random)

### 2.1.3 Drug Side-Effects: What Do Patient Forums Reveal?

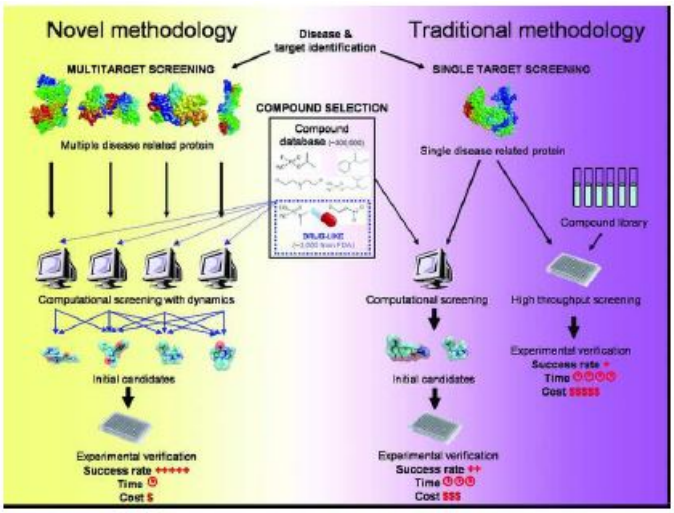
In this research article, Discussed value of Patient forums, reporting side-effects .Highlighted that Patient forums provide a platform where people can discuss the medical conditions, medicines take, outcomes, and side-effects experience. Also proposed propose finding patterns of reporting side-effects, both using heuristics and automatically extracted rules, to be able to extract the entities and the relations.. The goal is creation of a gold standard collection that allows a platform for fair comparison of different techniques for text mining researchers. Provided few entities related with side effects and the examples. Also we can collect many resource and useful links relating to drug side effects

### 2.1.4 Algorithmic Framework for Predicting Side Effects of Drugs

This article is contributed in fourfold ways (i) showed that computational prediction of side effects of drugs is possible. Have been presented an approach that combines correlation based analysis with network diffusion, achieving very high retrieval accuracy. In cross validation, able to accurately predict side effects for up to two thirds of the drugs; in a blind test, able to confirm the predictions for almost half of the drugs. (ii) Demonstrate the use of different data sets, such as chemical structure and cell line response, for the prediction task. The use of different data sets could potentially increase the sensitivity and specificity of the predictions. (iii) find a significant correlation between the similarity of the predicted side effects of drugs and the targets, indicating the potential utility of the algorithm in drug target identification. (iv) Show that analyzing multiple side effects together improves on a simple approach that considers each side effect independently

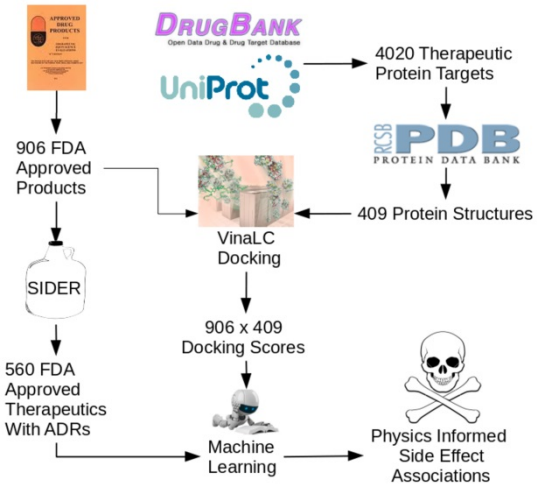
### 2.1.5 NOVEL PARADIGMS FOR DRUG DISCOVERY: COMPUTATIONAL MULTI TARGET SCREENING

Develop a novel computational approach to drug screening and identification that takes advantage of state of the art protein folding and small molecule docking with dynamics methods. Furthermore, have been benchmark and improve the method via rigorous experimental assessment of prospective predictions.




### 2.1.6 Prediction of Drug coalition by Integrating Molecular and Pharmacological Data

In this approach to predict drug combinations by representing drug combinations as combinations of molecular and pharmacological features, including target proteins, therapies, and indication areas, not only led to the proposal of new drug combinations but also allowed mechanistic insights into existing ones. The overlap between the predictions and those reported in the literature demonstrate that this approach can effectively identify new drug combinations with the enriched feature patterns as an indicator for the mode of action underlying both marketed and predicted drug combinations. A limitation of this method is that it relies on the feature patterns enriched in approved drug combinations, which limits the predictions to those combinations that are similar to existing ones to some extent. Nevertheless, the new combinations are far from being obvious given the vast space of possible solutions. the methods proposed here can limit the search space of possible drug combinations as a guide for experimental screens and provide an alternative starting point towards repurposing old drugs.



### 2.1.7 Adverse Drug retro action Prediction Using Values Produced by Huge Scale Drug and Protein Target Docking on High Performance Computing Machines

Depicted on this analysis that molecular docking may enable cost effective, comprehensive,reliable, high throughput screening of a drug candidate for binding across multiple known targets to provide predictions of clinically responsible ADRs. Presented a major principles approach to in silico ADR prediction for drug compounds that leverages physics-based models and HPC by docking 560 small molecule drugs to 409 structures of identified Drug Bank protein targets. Only 21% (87 out of 409) of the drug-protein binding features involve known targets of the drug subset, providing a significant probe of off-target effects. The median AUCs obtained during 10-fold cross-validation were comparable between the VinaLC off-target models (AUC = 0.60–0.69) and the Drug Bank on-target models (AUC = 0.61–0.74) across the ten ADR groups. Most importantly, the VinaLC off-target model outperformed the Drug Bank on-target model for predicting two ADR groups, neoplasms and vascular Disorders

### 2.1.8 Drug Off Target upshot Predicted Using Structural Analysis in the affinity of a Metabolic Network architecture

Pharmaceutical science is only opening to scratch the surface on the exact mechanisms of drug action that lead to a drug's breadth of patient retribution, both expected and side effects. Few years of clinical trials, molecular analysis, and more recent computational analysis have been sought to characterize the interactions between a drug and the cell's molecular machinery. Have been devised an integrated computational approach to assess how a drug may affect a particular system, in this study the metabolism of the human kidney, and its capacity for filtration of the component of the blood. Applied this obtrude to retrospectively enquire potential causal drug targets leading to increased blood pressure in participants of clinical trials for the drug torcetrapib in an effort to display how the obtrude could be directly useful in the drug development process. The results suggest specific metabolic enzymes that may be directly responsible for the side effect. The drug screening framework have been developed could be used to link adverse side effects to particular drug targets, discover new uses for old drugs, identify biomarkers for metabolic disease and drug response, and suggest genetic or dietary risk factors to help guide personalized patient care.

## 2.2 Non Computational Approaches in Drug Side Effects Prediction

### 2.2.1 Predicting Drugs Side Effects Based on Chemical and Chemical Interactions and Protein and Chemical Interactions

Since A drug side effect is an undesirable effect which occurs in addition to the intended therapeutic effect of the drug. Have been decided to generate this predicting Drugs Side Effects Based on Chemical and Chemical Interactions and Protein and Chemical Interactions research. In this study, Have been proposed a novel prediction method to recognize drugs side effects. For any query drug, its side effects were determined by the following strategy if there exist interactive compounds of in the training group, only chemical and chemical interactions were used to identify its side effects otherwise, both
Chemical and chemical interactions and protein and chemical interactions were employed to make prediction. Good performance of the method on the training and test data-sets indicates that the method is quite effective in identifying drug adverse effect . Hope that the method would assist in the prediction of drugs adverse effects during drug development and screening out drug candidates with undesired side effects.

### 2.2.2 Predicting drug adverse effect profiles , chemical Fragment-based approach

In this paper have been proposed a novel method to predict potential side-effect profiles of drug candidate molecules based on the chemical structures using sparse canonical correlation analysis (SCCA). The proposed method in this research article expected to be useful in various stages of the drug development process. Have been proposed five possible methods to predict drug side effect profiles from the chemical structures. those are Random assignment (Random) ,Nearest neighbor (NN),Support vector machine (SVM),Ordinary canonical correlation analysis (OCCA) and Sparse canonical correlation analysis (SCCA) It and also help to find new indications for known drugs, a process named drug re purposing. side-effects of drugs used in a given pathology can be viewed as a beneficial effect in another pathology. Have been proposed a new method to predict potential side-effects of drug candidate molecules based on the chemical structures, applicable on large molecular data banks. A unique feature of the proposed methods are ability to extract correlated sets of chemical substructures (or chemical fragments) and side-effects. This is made possible using sparse canonical correlation analysis (SCCA).The method is computationally efficient and is applicable on large datasets. The method could help to identify chemical substructures of known drugs that might participate in the appearance of a given side-effect. Have been focused on clear statistics also such as have been mentioned that "As an illustration of the extent of this problem, serious drug side-effects are estimated to be the fourth largest cause of death in the United States, resulting in 100,000 deaths per year" those details have been extracted from Nakamura Y's "When good drugs go bad" research article on NCBI(National Center for Biotechnology Information, National Library of Medicine ,United State).

**Ordinary canonical correlation analysis (OCCA)**

Suppose that have a clique of n drugs with p substructure features and q side effect features.Each drug is represented by a chemical substructure feature vector x = (x1 , ..., xp )T and by side-effect feature vector y =(y1 , ..., yq )T .Consider two linear combinations for chemical substructures and side-effects as ui = aT xi and vi = bT yi (i = 1, 2, ..., n), where a = (a1, ..., ap)T and b = (b1 , ..., bq)T are weight vectors. The goal of ordinary CCA is to find weight vectors a and b which maximize the canonical correlation coefficient:

$$\rho = corr(u, v) = \frac{\sum_{i=1}^{n} \alpha^T \mathbf{x}_i \cdot \beta^T \mathbf{y}_i}{\sqrt{\sum_{i=1}^{n} (\alpha^T \mathbf{x}_i)^2} \sqrt{\sum_{i=1}^{n} (\beta^T \mathbf{y}_i)^2}}, \qquad (1)$$

Where $\sum_{i=1}^{n} u_i = 0$ (resp. $\sum_{i=1}^{n} v_i = 0$)is assumed andu (resp. v) is called canonical component for x (resp. y)
Let × denote the n × p matrix defined as X = [x1 , ...,xn]T , and let Y denote the n × q matrix defined as Y =[y1 , ..., yn ]T . The columns of X and Y are assumed to be centered and scaled. Then maximization problem can be written as follows:

$$\max\{\alpha^T X^T Y \beta\} \quad \text{subject to} \quad \alpha^T X^T X \alpha = 1, \quad \beta^T Y^T Y \beta = 1, \tag{2}$$

In other high dimensional issues, it is known that good results can be obtained by treating the covariance matrix as a diagonal matrix as suggested in [34].
Therefore, substitute identity matrices for XT X and YT Y , and consider the following optimization problem:

$$\max\{\alpha^T X^T Y \beta\} \quad \text{subject to} \quad ||\alpha||_2^2 = 1, \quad ||\beta||_2^2 = 1. \tag{3}$$

**Random assignment (Random)**
To evaluate how difficult the problem considered in this paper is, apply a random assignment procedure, that is, use the 0/1 ratio to assign a binary label to each test drug randomly. For example, if the ratio in given training data is 90%, can assign zero for 90% of examples in test; otherwise 1. This method is used as a baseline method in this study.
**Nearest neighbor (NN)**
The most straightforward approach is to apply the nearest neighbor (NN), which predicts a given drug x to have the same side-effects as those of the drug (in a training set) whose chemical substructure profile is the most similar. For each query drug, look for k nearest neighbors, and if k' of k have a side-effect, assign the prediction score of k'/k to the query drug. Repeat this procedure for q side-effects.
Support vector machine (SVM)
A more sophisticated approach would be to apply a supervised binary classification method for predicting whether a given drug x has a side-effect or not, and repeat this process for all q side-effects. The support vector machine (SVM) is a well-known binary classifier, and it has become a popular classification method in bioinformatics and chemoinformatics because of its high-performance prediction ability. Test several kernel functions such as linear kernel, Gaussian RBF kernel with various width parameters and polynomial kernel with various degree parameters. Note that this strategy needs to construct q individual SVM classifiers for q side-effects, so it will require considerable computational burden, because q is quite huge in practical applications (q is 1385 in this study).

### 2.2.3 Relating drug and protein interaction network with drug adverse side effects

In this article, a novel has been proposed SCCA based, approach to relate drug targeted proteins with drug side effects. Biological hermeneutics have been provided of proteins and side effects extracted in each canonical component. Using a cross validation scheme, highlighted that the suggested approach displays better performance than chemical-structure-based methods for the prediction of drug adverse effects. Overcomes suggest that side effect of drugs is more correlated to the mechanism of action, rather than to the chemical structure, which presents an interesting result. In most drug discovery projects, a therapeutic target playing a role in a given disease is searched for, and once identified; the corresponding pathways can be identified. The components that are enriched in these pathways provide a list of potential side effects that one can expect for future drugs acting on the target of interest.A statistical model have been constructed for

the prediction of drug adverse effect profiles from protein binding profiles , primarily because the number of drugs with side effect information is much less than those with targeted protein information. As Well as , investigated scenarios where a drug candidate molecules' targeted protein informations are available,but not the side effect informations. Giving good Prediction of side effects for uncharacterized drugs

**Sparse canonical correlation analysis - SCCA**

Number of elements in the weight vectors $\alpha$ and $\beta$ in OCCA are non-zeros, which makes it hard to interpret the result. Practically , it is desirable to find weight vectors that have large correlation, but that are also sparse for easier interpretation.To impute the sparsity on $\alpha$ and $\beta$, consider the below maximization function with additional L1 penalty terms:

$$\max\{\boldsymbol{\alpha}^T X^T Y \boldsymbol{\beta}\} \quad \text{subjectto}$$

$$||\boldsymbol{\alpha}||_2^2 \leq 1, \quad ||\boldsymbol{\beta}||_2^2 \leq 1, \quad ||\boldsymbol{\alpha}||_1 \leq c_1\sqrt{p}, \quad ||\boldsymbol{\beta}||_1 \leq c_2\sqrt{q},$$

Where ||||1 is L1 norm (the sum of all absolute values of the vector elements),and c1 and c2 are parameters to manipulate the sparsity (0<c1=1 and 0<c2=1).For simplicity, the equal value is used for c1 and c2 in this study. The CCA with L1 forefeits is referred to as SCCA. The weight vectors $\alpha$ and $\beta$ can be optimised by solving penalized matrix decomposition of the matrix Z=XT Y (Witten et al., 2009).To obtain many canonical components, perform a deflation manipulation iteratively as follows: Z(k+1)?Z(k)−dk$\alpha$k$\beta$Tk , where Z(k) is the input of step k (Z(1)=XT Y), $\alpha$k and $\beta$k are the weight vectors, and dk is singular value obtained in the k-th step (corresponding to the k-th component) (k=1,2,...,m). Eventually, obtain m pairs of weight vectors ($\alpha$1,$\beta$1),...,($\alpha$m,$\beta$m). Proteins and adverse effects with non-zero weights in the weight vectors are extracted as correlated sets.

Prediction of adverse effect profiles for new molecules

Given the profile of targeted proteins Xnew for a drug of unknown side effects, consider predicting its potential side effect profile Ynew based on the weight vectors
{$\alpha$k }mk=1 and {$\beta$k }mk=1.
use the following prediction value for a given molecule:

$$\mathbf{y}_{\text{new}} = \sum_{k=1}^{m} \boldsymbol{\beta}_k \rho_k \boldsymbol{\alpha}_k^T \mathbf{x}_{\text{new}} = B\Lambda A^T \mathbf{x}_{\text{new}},$$

Where A=[$\alpha$1,...,$\alpha$m], B=[$\beta$1,...,$\beta$m] and is the diagonal matrix whose masses are canonical correlation coefficients. If the j-th element in ynew has a high score, the new molecule is predicted to have the j-th side effect (j=1,2,...,q). The same prediction value was proposed in the previous work (Pauwels et al., 2011).

Enrichment analyses of focused proteins

Let Gc denote the set of extracted proteins in constituent c and G denote
the clique of proteins in a functional unit (e.g. KEGG pathway map). Let r =|Gc|,k=|G|,z=|G|and l the total number of proteins in the entire dataset. Assume that z follows a hypergeometric circulation.

probability to look to an intersection of size z between G and Gc is computed as follows

$$p(G,G_c)=\sum_{i=z}^{\min(k,r)}\frac{\binom{k}{i}\binom{l-k}{r-i}}{\binom{l}{r}}.$$

define the enrichment value $s(c)$ of a component $c$ by
$s(c)$=-log10$p$FDR($G,Gc$)
where $p$FDR($G,Gc$) is the corrected value of $p(G,Gc)$ by the False Discovery Rate(FDR) (Benjamini and Hochberg, 1995).

### 2.2.4 Large scale prediction of drug side effects reactions using chemical, biological, and phenotypic properties of drugs

Objective of this article is reviewing Adverse drug reaction (ADR).This is Highlighted that ADR is one of the major causes of failure in drug development and Severe ADRs that go undetected until the post-marketing phase of a drug often lead to patient morbidity. Accurate prediction of potential ADRs is required in the entire life cycle of a drug, including early stages of drug design, different phases of clinical trials, and post-marketing surveillance. This study proposed a new drug surveillance framework for ADR prediction by integrating chemical (i.e., compound signatures), biological (i.e., protein targets, transporters, enzymes, and pathways), and phenotypic (i.e., indications and other known side effects) properties. A set of 1385 side effects for 832 drugs have been used from the SIDER database. ML models have been developed to integrate the different sources of information for prediction. Few methods have been used and presented that many studies have utilized either chemical structures or molecular pathways of the drugs to predict ADRs as methods in the studies. In the result evaluation phase, on a fivefold cross validation have been focused, showed that the support vector machine algorithm outperformed the others. Of the three types of information, phenotypic data were the most informative for ADR prediction. Algorithms have been compared using the full feature set over all versus common ADRs and Clinical validations have been used examples of cerivastatin and rofecoxib.

## 2.3 Examples of adverse effects associated with specific medications

| Adverse Drug Reaction | Number of Drugs | Examples of Brand Names |
|---|---|---|
| Depression | 166 | Accutane, Advil, Catapres, Cipro, Dalmane, Factive, Inderal, Naprosyn, Norpace, Pepcid, Reglan, Tagamet, Talwin, Ultracet, Valium, Xanax, Zantac |
| Psychoses/ hallucinations | 156 | Aldomet, Benadryl, Catapres, Celebrex, Cipro, Dexatrim, Elavil, Halcion, Inderal, Lanoxin, Procanbid, Sonata, Tagamet, Ultracet, Valium, Vioxx |
| Confusion/ delirium | 147 | Amaryl, Ambien, Benadryl, Catapres, Cipro, Compazine, Diabeta, Diabinese, Dymelor, Elavil, Mellaril, Sinemet, Tagamet, Valium, Xanax, Zantac |
| Dementia | 76 | Aldomet, Inderal, Maxzide, Mellaril, Regroton, Restoril, Ser-Ap-Es, Tagamet, Valium, Xanax,Zantac |
| Insomnia | 35 | Avelox, Floxin, Inderal, Lasix, Mevacor, Nicorette, Sudafed, Synthroid, Theo-24 |
| Parkinsonism | 40 | Abilify, Aldomet, Asendin, Cardizem, Compazine, Elavil, Geodon, Haldol, Mellaril, Prozac, Reglan, Regroton, Risperdal, Thorazine |
| Tardive dyskinesia | 19 | Abilify, Asendin, Buspar, Compazine, Geodon, Haldol, Mellaril, Risperdal, Thorazine, Wellbutrin, Zyban, Zyprexa |
| Dizziness on standing | 154 | Abilify, Calan SR, Cardizem CD, Cardura, Catapres, Compazine, Elavil, Geodon, Haldol, Hytrin, Inderal, Isordil, Lasix, Minipress, Nitro-Bid, Prinivil, Procardia, Sonata, Tenormin, Valium, Xanax |
| Falls/hip fracture | 59 | Ambien, Celexa, Compazine, Dalmane, Elavil, Haldol, Isordil, Lexapro, Navane, Nembutal, Prozac, Restoril, Sinequan, Valium, Xanax |
| Automobile accidents | 28 | Ambien, Asendin, Ativan, Celexa, Elavil, Lexapro, Norpramin, Pamelor, Paxil, Prozac, Sinequan, Tofranil, Valium, Xanax, Zoloft |
| Sexual dysfunction | 127 | Abilify, Calan SR, Geodon, Lopid, Lopressor, Norpace, Pepcid, Proscar, Prozac, Sarafem, Tagamet, Tegretol, Transderm-Scop, Zantac |
| Loss of appetite, nausea, vomiting | 63 | Advil, Avelox, Daypro, Demerol, EES, Feldene, Feosol, K-Lor, Lanoxin, Levaquin, Relafen, Sumycin, Theo-24, Ultracet, Ultram |
| Abdominal pain, ulcers, GI bleeding | 48 | Advil, Anaprox, Celebrex, Cortone, Daypro, Decadron, Feldene, Indocin, Motrin, Relafen, Somophyllin, Theo-24, Ultracet, Vioxx, Zithromax |
| Constipation | 107 | Amphojel, Benadryl, Caltrate, Cogentin, Inderal, Lotronex, Maalox, Talwin, Tylenol No. 3, Tylox, Ultram, Urised |
| Diarrhea | 56 | Aciphex, Aldomet, Avelox, Cipro, Dulcolax, Maalox, Phillips' Milk of Magnesia, Nexium, Peri-Colace, Precose, Prilosec, Sporanox, Sumycin, Zelnorm |
| Lung toxicity | 59 | Cordarone, Feldene, Inderal, Prinivil, Tegretol, Vasotec, Visken |
| Blocked urination | 56 | Antivert, Artane, Benadryl, Bentyl, Compazine, Duragesic, Elavil, Felbatol, Haldol, Sinequan, Tavist, Ultram, Zyban |
| Urine leakage | 84 | Aricept, Celexa, Esidrix, Hytrin, Inderal, Lasix, Lexapro, Lithobid, Minipress, Neurontin, Paxil, Restoril, Tenormin, Valium, Xanax, Zaroxolyn, Ziac, Zoloft |

# 3 Chapter 3

## 3.1 Conclusion

This survey has been based on "**What are the prevailing methods on Drug Side Effects Prediction**?" research problem. Twelve research article Have been disused which contain Computational and Non Computational Approaches. Within those Computational and Non Computational Approaches, there are many methods that those 12 research article manipulated. "High Confidence Predictions of Drug-Drug Interactions" research article discussing about following methods. Line Walking Recursive Partitioning (LWRP), Normal Equation Recursive Partitioning (NERP), Substructure Discovery Using Examples SUBDUE), Training and Validation Set Selection. Within this article when all of above methods agree, the predictive accuracy is 94%.
By using Benchmark Dataset, Chemical-Chemical Interactions and Protein-Chemical Interactions,Based Method, Similarity-Based Method and Jackknife Test Methods, "Predicting Adverse Side Effects ,Based on Chemical and Chemical Interactions and Protein and Chemical Interactions" research Manipulated. A training dataset and test datasets were constructed from the benchmark dataset that contains 835 drug compounds to evaluate the method. By a jackknife test on the training dataset, the 1st order prediction accuracy was 86.30%, while it was 89.16% on the test dataset.
"self acting Drug Side Effect Discovery from Online Patient" Research Fed Following methods Data Collection,Side-Effect Extraction ,Side-Effect Onology and Association of Drug Class with Side-Effects
Random assignment (Random), nearest neighbor (NN), Support vector machine (SVM), Ordinary canonical correlation analysis (OCCA), sparse canonical correlation analysis (SCCA) Methods driven by "Predicting drug side-effect profiles ,chemical fragment-based approach "research.
"Relating drug–protein interaction network with drug side effects" research using Ordinary canonical correlation analysis (OCCA), Sparse canonical correlation analysis (SCCA), Prediction of adverse drug effect profiles for new molecules, Enrichment analyses of targeted proteins Data description, Features study, Experimental design and ML algorithms Methods use by "Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs" Research

# 4 Chapter 4

## 4.1 Reference


[01]Hudelson , Ketkar, Holder,Carlson,Peng, Waldher and Jones, M , "High Confidence Predictions of Drug-Drug Interactions ". RetrievedJanuary ,
2014 Available: http://www.eecs.wsu.edu/~holder/pubs/HudelsonJMC08.pdf

[02]Liang-Chin Huang, Xiaogang Wu and Jake Y Chen, L , "Predicting adverse side effects of drugs". RetrievedJanuary ,
2013 Available: http://www.biomedcentral.com/content/pdf/1471-2164-12-S5-S11.pdf

[03]Lei Chen,Tao Huang, Jian Zhang, Ming-Yue Zheng,Kai-Yan Feng,Yu-Dong Caiand Kuo-Chen Chou, L , "Predicting Drugs Side Effects Based on Chemical-Chemical Interactions and Protein-Chemical Interactions". RetrievedJanuary ,
2013 Available: http://www.hindawi.com/journals/bmri/2013/485034/

[04]Liu, Li and Seneff, J , "Automatic Drug Side Effect Discovery from Online Patient". RetrievedJanuary ,
2014 Available: http://people.csail.mit.edu/seneff/IMMM.pdf

[05]Pirmohamed, M , "Adverse Drug Reactions: Classification and Mechanisms". RetrievedJanuary ,
2014 Available:http://www.bps.ac.uk/SpringboardWebApp/userfiles/bps/file/Education/ResourcesForUniversitiesAndStudents/Mechanisms%20of%20Adverse%20Drug%20Reactions_M_Pirmohamed.pdf

[06]Karimi, Su Nam Kim and Lawrence Cavedon, S , "Drug Side-Effects: What Do Patient Forums Reveal". RetrievedJanuary ,
2014 Available: http://meco-project.eu/sites/meco-project.eu/files/Karimi_0.pdf

[07]Pauwels, Vronique Stoven and Yoshihiro Yamanishi, E , "Predicting drug adverse side effect , chemical fragment focused approach". Retrieved January ,
2014 Available: http://www.biomedcentral.com/content/pdf/1471-2105-12-169.pdf

[08]Mizutani, Edouard Pauwels, VÃ©ronique Stoven, Susumu Goto and Yoshihiro Yamanishi, S , "Relating drug and protein interaction network with drug side effects". RetrievedJanuary ,
2014 Available: http://bioinformatics.oxfordjournals.org/content/28/18/i522.full.pdf

[09]Liu,Yonghui Wu, Yukun Chen,Jingchun Sun, Zhongming Zhao, Xue-wen Chen, Michael Edwin Matheny and Hua Xu, M , "Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs". RetrievedJanuary ,
2014 Available: http://jamia.bmj.com/content/19/e1/e28.full.pdf

[10]Samudrala, R , "NOVEL PARADIGMS FOR DRUG DISCOVERY: COMPUTATIONAL MULTITARGET SCREENING". RetrievedJanuary ,
2014 Available: http://cando.compbio.washington.edu/pioneer_application/

[11]ATIAS and RODED SHARAN, N , "Algorithmic Framework for Predicting Side Effects of Drugs". RetrievedJanuary ,
2014 Available: www.cs.tau.ac.il/~roded/se-jcb.pdf

[12]Montiago X. LaBute, Xiaohua Zhang, Jason Lenderman, Brian J. Bennion, Sergio E. Wong, Felice C. Lightstone, N , "Adverse Drug Reaction Prediction Using Values Produced by Huge Scale Drug and Protein Focused Docking on Advance Performance Computing Machines". RetrievedJanuary ,
Available: www.cs.tau.ac.il/~roded/se-jcb.pdfhttp://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0106298#s5

[13]Roger L. Chang, Li Xie, Lei Xie, Philip E. Bourne, Bernhard . Palsson, , "Drug Off Target results Predicted Using Structural Analysis in the Context of a Metabolic Network Model". RetrievedJanuary ,
2014 Available: http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1000938

[14]Xing-Ming Zhao, Murat Iskar, Georg Zeller, Michael Kuhn, Vera van Noort, Peer Bork, , "Prediction of Drug coalition by Integrating Molecular and Pharmacological Data". RetrievedJanuary ,
2014 Available: http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1002323

[15]Public Citizen's Health Research Group, , "Drug-Induced Diseases". RetrievedJanuary ,

2014 Available: http://www.worstpills.org/public/page.cfm?op_id=5