Drug Discovery Based On Model Driven Architecture

Miss. Shruti Banpatte[1], Miss. Ujjwala Shinde[2], Miss. Rajashree Patil [3], Miss. Kiran Manole[4], Miss. Rasika Patil [5], Asst. Prof. Mr. K.M.Aldar[6]

*[1]Miss. Shruti Banpatte Department Of Computer Science and Engineering
Sou. Sushila Dyanchand Ghodawat Charitable Trust's
Sanjay Ghodawat Of Institutions , Atigre, Maharashtra, India.*

*[2]Miss. Ujjwala Shinde Department Of Computer Science and Engineering
Sou. Sushila Dyanchand Ghodawat Charitable Trust's
Sanjay Ghodawat Of Institutions , Atigre, Maharashtra, India.*

*[3]Miss. Rajashree Patil Department Of Computer Science and Engineering
Sou. Sushila Dyanchand Ghodawat Charitable Trust's
Sanjay Ghodawat Of Institutions , Atigre, Maharashtra, India.*

*[4]Miss. Kiran Manole Department Of Computer Science and Engineering
Sou. Sushila Dyanchand Ghodawat Charitable Trust's
Sanjay Ghodawat Of Institutions , Atigre, Maharashtra, India.*

*[5]Miss. Rasika Patil Department Of Computer Science and Engineering
Sou. Sushila Dyanchand Ghodawat Charitable Trust's
Sanjay Ghodawat Of Institutions , Atigre, Maharashtra, India.*

*[6]Asst.Prof.Mr.K.M.Aldar Department Of Computer Science and Engineering
Sou. Sushila Dyanchand Ghodawat Charitable Trust's
Sanjay Ghodawat Of Institutions , Atigre, Maharashtra, India.*

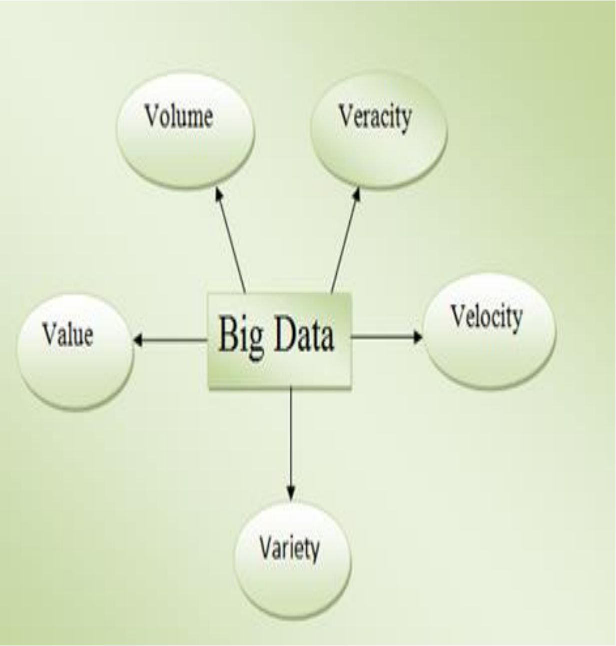
---***---

Abstract - *The object of this paper is to develop a Drug Discovery system in MDA(Model Driven Architecture). In this paper, we define a set of models to get proper guidelines for the identifications of drug, genes, disease, ATC, side effect. A Model-driven architecture is a kind of the domain model. The Drug Discovery system, is a powerful tool for users to suggest drugs for different diseases and also a side effect of drugs. This paper proposes to incorporate details of drugs and data mining into our drug discovery system. Our first aim is to find new uses for old drugs and the second aim is to predict side effect. This prediction model constructed by using different type of biological data. A service for new uses of old drugs can found in data models. We demonstrate that the prediction model for drug discovery is implemented as a prototype system to verify those models and their practicality.*

Key Words : **Drug, Diseases, Side Effect, ATC Code, Genes, MDA.**

1.INTRODUCTION

The paper "DRUG DISCOVERY IN MDA" is a website which provides information about diseases, drugs, gens and it's side effects for doctors as well as pharmacy users also. It also includes Registration for doctors and they can add their new research of drugs. Doctors can refer old paper and research papers for drug discovery. It is designed for drug discovery and its side effects. We can also search drugs for related diseases. In our paper there are certain fields which specify details like ATC_code, drugs, diseases, genes, side effects of drugs.



Figur-1 Illustration of Big Data Properties

Big data is a large and complex data structure. The term "big data" often refers simply the use of predictive analytics, user behavior analytics to a particular size of the data set. Big data require large amounts of data that requires the data which is in the form of structured and unstructured data. The key challenges of big data are specified as volume, velocity, variety, veracity, and value [1]. The value (business value) is crucial to offer services for business. The characteristics of big data are volume, velocity and variety. Here, the "veracity" of big data properties an important problem to develop an application for drug discovery because all of the medical data is not offered and the data in drug discovery is sparse.

We can develop a database application model like side effect prediction of drug, which is our first aim of big data application, and for drug discovery, which finds new uses from the older drugs, introducing the proposed system development process in MDA into our research process. Drug side effect prediction is one of screening methods in drug discovery and drug candidates can be ranked by this prediction [7]. In drug discovery Anatomical Therapeutic Chemical (ATC) is a classification System, this System is used for the classification of active ingredients of drugs according to the organ or system in which they react and their therapeutic, pharmacological and chemical properties. This pharmaceutical coding system separates drugs into different groups, according to the organ or system on which they act like chemical characteristics. ATC code will indicate that one letter, which specify the anatomical main group.

1.1 Literature Review

As time is proceeding next, technology is developing every single moment. Two of the basic fundamental purpose of technology are to make objects that are not complicated to understand by the user and makes working of the user more convenient. Things are basic when the interface between human and technology is less complex. It is more and more convenient for the users. A revolution made by the mobile technology made our modern life easier as it is found that new services and related commerce with more and more available.

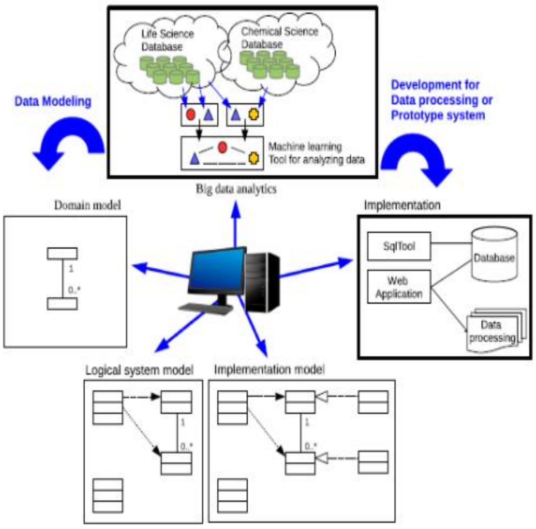
In recent years, a research on drug discovery to find the new medical suggestion from the old drugs and to redevelop them as a treatment for another disease has attracted attention as drug re-positioning in order to minimize cost and a development period for efficiency. One reason is complex biological processes, is one of the human diseases which is one-drug-one-gene approach is not effective to treat. The side effects in official reports of prescription drugs which have increased substantially over the past decade. The existing drugs are designated for all other diseases and were joined to disease genes of rheumatoid arthritis, and those existing drugs were possible drug candidates for new uses of rheumatoid arthritis. So, we considered those results in the clue of drug discovery, and proposed database application model to discover the new drug candidates by new uses of the old drugs.

1.2 Existing System

The Main complex things in medical is multiparameter in drug discovery and balancing numerous activities. The characteristics are in series of chemicals[4].To manage large varieties of property, it requires designers and that data ensure incorporation into their design strategies. To understand the complexity of biological processes to a sufficient degree Of a relational drug design. In that 'one-target–one-drug' model is a multiple-target approach, predicted by many. It will add further to this complexity [8] as networks of interactions grow. It needs to work effectively in larger teams from multiple disciplines and at multiple locations.

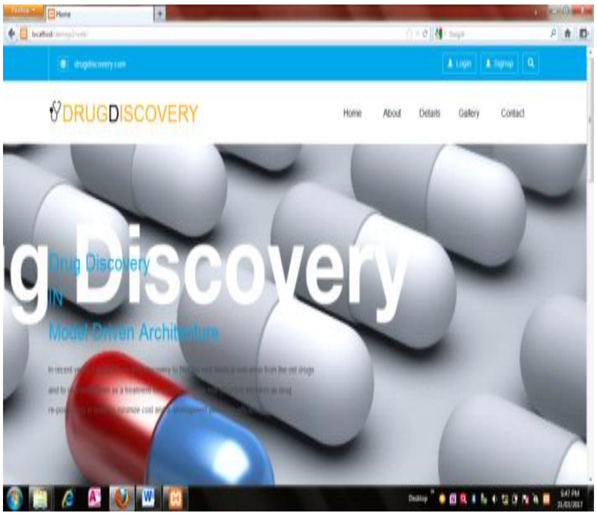
The papers in medical Drug Discovery are a part of collaborative terms, including pharmacologists, molecular biologists, and others[1].For example, In other fields, like particular engineering, large multifactorial problems can be divided into smaller tasks and divided between optimal research and development teams who work independently [9]. In drug discovery application data will be available in large amount. So veracity is an important challenge in big data.

2. System Architecture



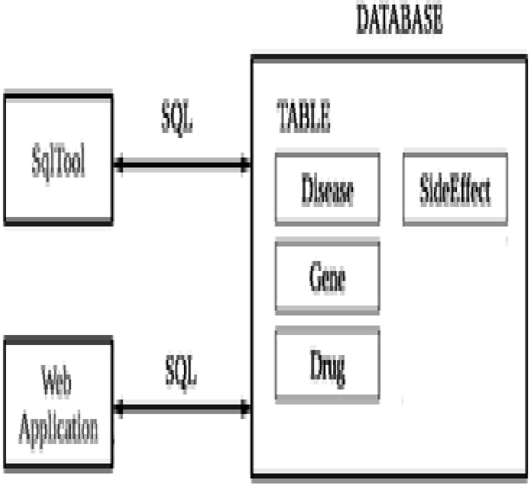
Figur-2. System Architecture[1]

The system shows an implementation that consists of a database which is accessed by MYSQL and Web application. It contains a large amount of medical data and shows software development environment. As for web application open source is adopted [1]. It presents a static view for the objects and classes that makes a design and analysis pace.



Figur-3 System Implementation

It contains drug discovery information. It provides you searching, surfing facility and also provide registration for doctor to login this website and see new researches as well as old researches, they will add their new research into existing system. Another user is admin it check registered doctor's document that is new research approved it only if it contains valid data. Another user is a viewer he can only see the data/information related to drug, diseases, genes and its side effects. Doctors can add their comments, suggestion into a blog which will provide on our website.



Figur-4 Implementation model shown in the database Application[1]

2.1 Proposed System:

In the proposed system, we are developing a database application model and its service for side effects prediction of drug. Our first goal of big data application, and for drug discovery, which finds new uses for old drugs, introducing the proposed system development process in MDA(Model-driven architecture).

2.1.1 Data Repositories

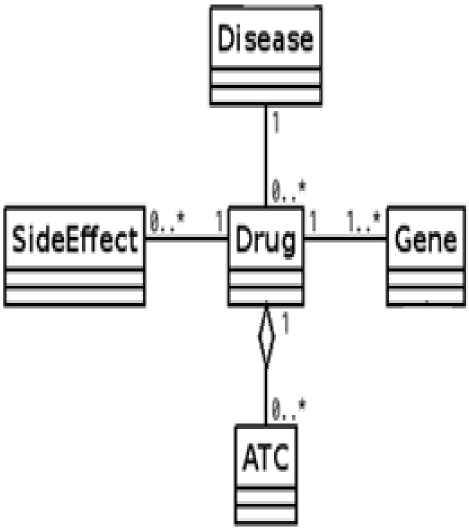
Data collected from different resources are publicly available on different websites.

- **KEGG:**

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a database source that contains chemical, biological, genomic and systemic functional information. In particular, gene brochure forms completely sequenced genomes is connected to higher-level systemic functions of the cell, the organism and the ecosystem. In addition to preserve the aspects to support basic research, KEGG is being enlarged towards more practical applications, integrating human diseases, drugs and other health-related substances [15]. This biological term includes different Human diseases, genes, drugs and ATC codes of that drug.[7]

- **SIDER 2:**

SIDER consist information about medicines of some disease and their recorded adverse drug reactions. The information is extracted from public documents and package inserts. The available information consists of drugs and its side effects containing all information connected to further data, for example drug–target relations. The side effect resource to capture phenotypic effects of drugs [12, 13]. Chemical (drug), ATC code, drug side effect, and its occurrence are extracted. The number of drugs has increased from 996 to 1430. Additional side effects have been recovered. Side effects that are indicates on the label as either potential or not occurring are removed from it.

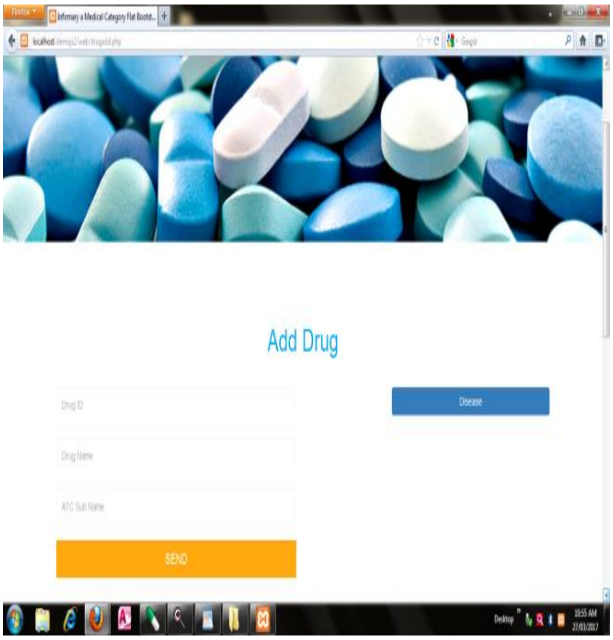


Figur-5 Illustration of Data Attributes consist five Domain objects.[1]

2.1.2 Data Attributes

- **Drug**

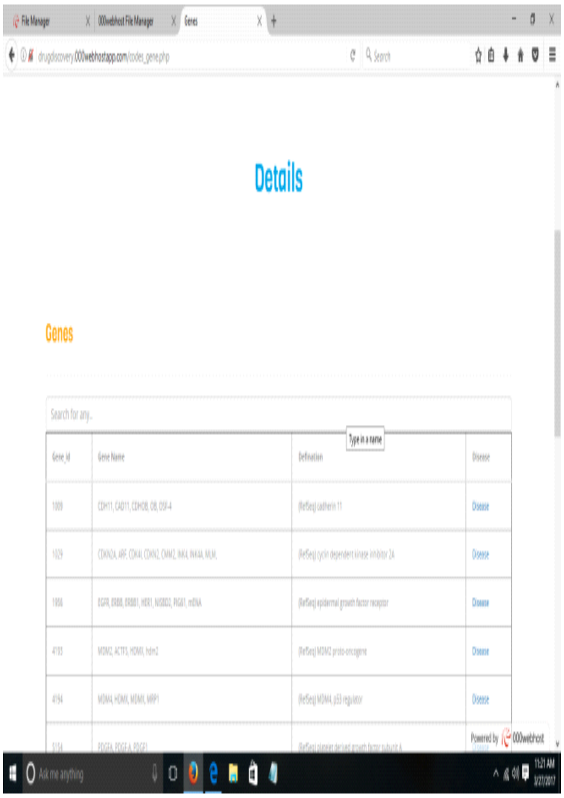
It is a chemical material that influences the processes of the mind or body. In the diagnosis, treatment, or prevention of diseases or other abnormal condition any chemical compound is used. A substance used recreationally for its effects on the central nervous system, such as a narcotic, to administer a drug to. The properties of drugs are biological, anatomical, genetic, and clinical features. The term Protein interaction exists biological feature, the ATC (Anatomical Therapeutic Chemical) code as anatomical feature, Gene ID as genetic feature, and side effect as clinical feature is used.


Infirmary a Medical Category Flat Bootst...
localhost/demoji2/web/drugadd.php
Add Drug
Drug ID
Drug Name
ATC Sub Name
SEND
Disease
Desktop
10:55 AM
27/03/2017

Figur-6 Logical system shows inserting Drugs.

- **Gene**

According to the Human Gene Nomenclature, a gene is defined as "a DNA segment that contributes to phenotype/function. In the absence of demonstrated function a gene may be characterized by order, transcription or homology". A gene is a sequence (a string) of bases. It is a combination which consists A, T, C, and G. These unique combinations decides the function of genes, as a letters are connected together to form words. Each human has thousands of genes billions of base pairs of DNA or bits of information repeated in the nuclei of human cells which determine individual characteristics (genetic traits). Genes are extracted from the KEGG database. In our body there are a number of genes. 2385 disease genes, in which 801 genes have the drugs [1, 7, 9].



Details

Genes

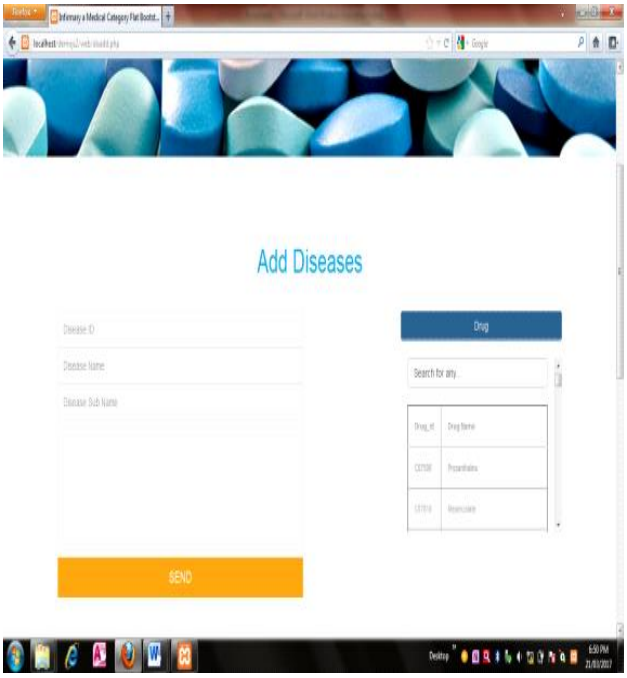
Search for any..

Gene_Id	Gene Name	Defination	Disease
1009	CDH11, CAD11, CDHOB, OB, OSF-4	(RefSeq) cadherin 11	Disease
1029	CDKN2A, ARF, CDK4I, CDKN2, CMM2, INK4, INK4A, MLM,	(RefSeq) cyclin dependent kinase inhibitor 2A	Disease
1956	EGFR, ERBB, ERBB1, HER1, NISBD2, PIG61, mENA	(RefSeq) epidermal growth factor receptor	Disease
4193	MDM2, ACTFS, HDMX, hdm2	(RefSeq) MDM2 proto-oncogene	Disease
4194	MDM4, HDMX, MDMX, MRP1	(RefSeq) MDM4, p53 regulator	Disease
5154	PDGFA, PDGF-A, PDGF1	(RefSeq) platelet derived growth factor subunit A	[illegible]

Figur-7 Logical system shows Genes Details

- **Disease**

A specific pathological process having a feature set containing signs and symptoms. It may affect the complete body or any part of it, and its etiology, prognosis, and pathology may be known or unknown. For particular diseases, see under the specific name, as Addison's disease. Diseases may be classified by pathogenesis (mechanism by which the disease is caused), etiology (cause), or by symptom(s). Alternatively, diseases may be organized according to the organ system involving, though this is often complicated since many diseases are affected to one or more organs. A rare disease is one that affects fewer than 50,000 patients or about 1 in 2,500 people. From 717 diseases, in which 174 diseases have the drugs are chosen from the KEGG database. 64 diseases, which are also nominated in the KEGG database, are found in JSNP database [1, 2].



Add Diseases

Disease ID

Disease Name

Disease Sub Name

SEND

Drug

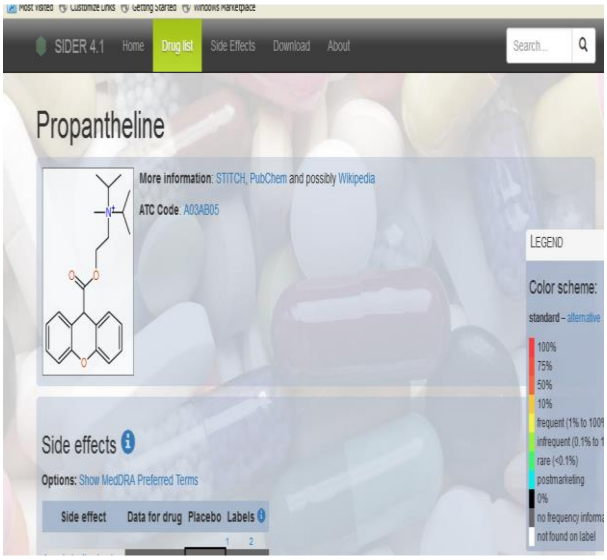
Search for any..

Drug_id	Drug Name
C07[illegible]	Propantheline
C07[illegible]	[illegible]

Figur-8 Logical system shows inserting Diseases.

- **Side Effects**

In medicine, a **side effect** is an effect, whether therapeutic, that is secondary to one intended and also the term is mostly employed to describe effect of particular drug, it can also be useful to apply, but there are consequences of using a drug. Drugs are prescribed or performed specifically for their side effects.



Propantheline

More information: STITCH, PubChem and possibly Wikipedia

ATC Code: A03AB05

Side effects

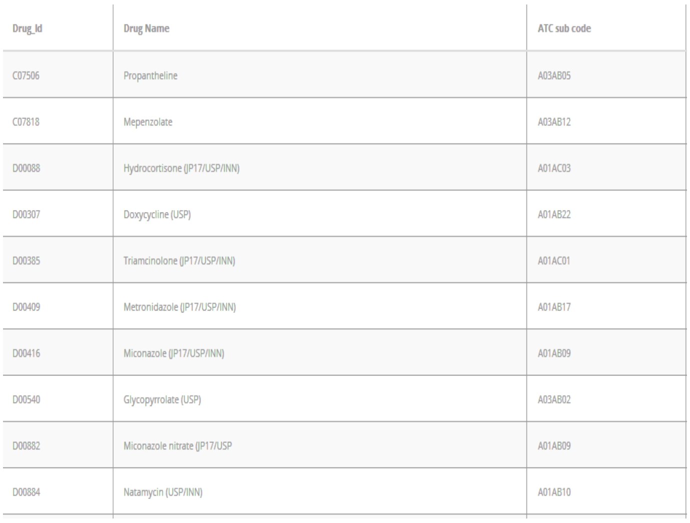
Options: Show MedDRA Preferred Terms

Side effect	Data for drug	Placebo	Labels

Figur-9 Logical system shows side effects.

- **ATC**

The Anatomical Therapeutic Chemical (ATC) Classification System is used in vital components for categorization of drugs. Following is the organs or system on which they act and their therapeutic, pharmacological and chemical premises. It is controlled for Drug Statistics by the World Health Organization Collaborating Centre Methodology (WHOCC), which was published first time in 1976. This pharmaceutical coding system split drugs into different groups according to the systems or organs on which they show therapeutic and chemical characteristics. Each bottom-level ATC code stands for a pharmaceutical used medium, or a combination of different substances, in a single demonstration. This means that one drug can have more than one code, like: acetylsalicylic acid.



Drug_Id	Drug Name	ATC sub code
C07506	Propantheline	A03AB05
C07818	Mepenzolate	A03AB12
D00088	Hydrocortisone (JP17/USP/INN)	A01AC03
D00307	Doxycycline (USP)	A01AB22
D00385	Triamcinolone (JP17/USP/INN)	A01AC01
D00409	Metronidazole (JP17/USP/INN)	A01AB17
D00416	Miconazole (JP17/USP/INN)	A01AB09
D00540	Glycopyrrolate (USP)	A03AB02
D00882	Miconazole nitrate (JP17/USP	A01AB09
D00884	Natamycin (USP/INN)	A01AB10

Figur-10 Logical system shows ATC Codes

The ATC system is mainly based on the (ACS) Anatomical Classification System, which is intended as a tool to classify pharmaceutical products for the pharmaceutical industry. This system, confusingly also called as ATC, it was initiated in 1971 by the European-Pharmaceutical Market Research Association (EphMRA) and is being maintained by the EphMRA and the Pharmaceutical Business Intelligence and Research Group (PBIRG).

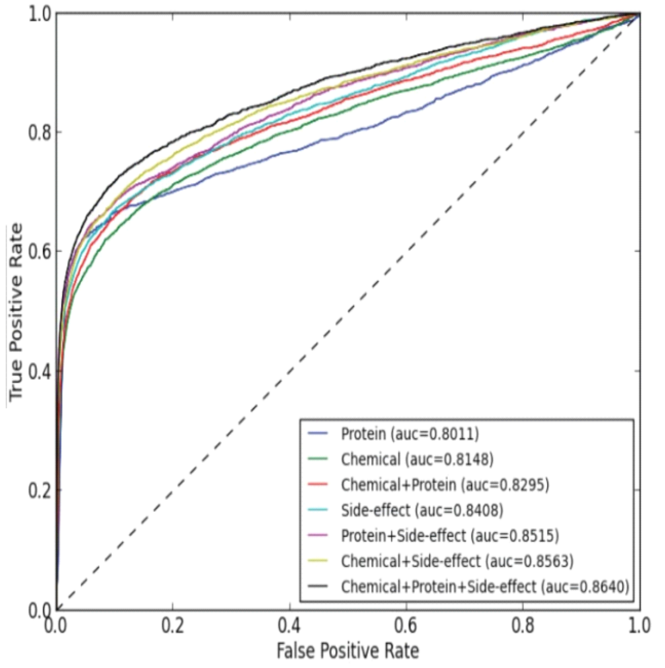
2.2 Results and discussion

The accuracy in the proposed Web application is dependent on the quality of the data that underpin them. This provides specific challenges within pharmaceutical research. Agreement for the new scientific attempt, without enough focus on the IT systems needed to underpin these investments, results in technology islands and should be considered malpractice. The veracity is a challenge in some data-driven decisions and resulted in an increase in the number of compounds retested. Researchers were now all working from identical data with equally rapid access and resulting from the same analysis tools. It was still necessary for experts, particularly those generating data, to give their opinions following the maximum that 'behind every data point there is a story'. Hence it was decided to create a project and data visualization suite, to backing cross-disciplinary research and data-driven decision making. There is also a crisis in the increasingly complicated world of early drug discovery that scientists become increasingly specialized, attention only on their particular sub regulation and the networks they preserve therein. As stated previous, discovery is increasingly made at the interface of disciplines by cooperative researchers from multiple fields. We need to develop more broad-oriented scientists which able to link regulation and with a deep comprehension of the drug discovery process and the generic scientific expertise in data analysis and visualization to accomplish this awareness.

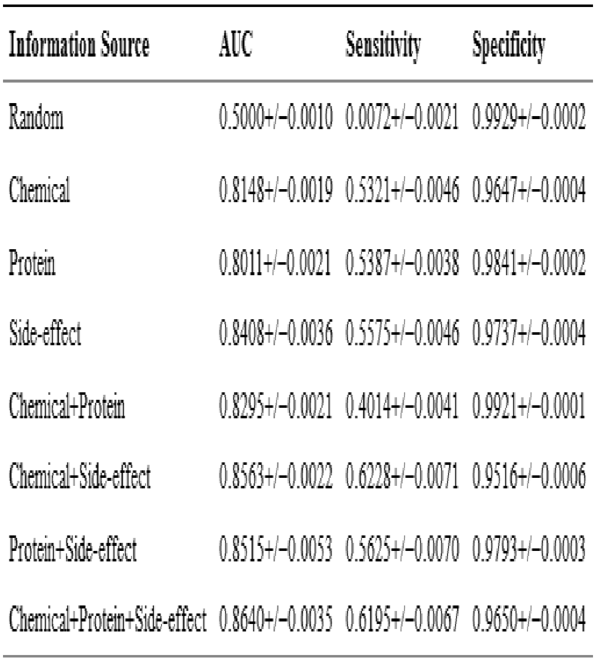
The ability to deviation, creativity and experience of the researchers themselves. A data resource of a paper is of greatest benefit. It means building assurance in the quality of that data by uniformly searching for difference and errors in the exploratory method. The process of drug discovery has assuredly become more complexty in recent years as a result were increased in the diversification and area of technologies planted within the process. Acknowledging some public data are agreed to well-maintained to continual databases, the majority in unorganized publications, typically it made available in a PDF format and exceptionally difficult to incorporate into other analyses. A result of the increased complications, volume and variations of the data sets presented in research. Which has been a new absorption on the development of visualization and analytics approximate [24].

Expanding the capability of medicinal chemistry is assumed in data driven research. It is possible to improve decision making in drug discovery and to ensure the most welfare result can be determined from the data resource which are available internally or externally. A Specialist should be honored for the rapid and constant addition of their data to such a resource with to senior management taking responsibility for the quality of these data. Analysis tools should be efficient with the storehouse and adequate training provided to all users.

For therapeutic-indication prediction task, Figure shows the ordinary cross validation for different data sources based on cross-validation demonstrations. Table review the concrete values of those evaluation results. When the information sources were equating independently, side-effect is the most instructive (AUC of 0.8408), chemical structure grade as the second (AUC of 0.8148), followed by target protein information (AUC of 0.8011). While grooming any two data sources will upgrade the AUC, combing all three information sources, we obtained the highest AUC score (AUC of 0.8640).



Figur-9 The ordinary differentiation of therapeutic make predictions for various data source combinations using in 10-fold cross validation. Information sources are classified in the legend of the figure according to their AUC score.



Information Source	AUC	Sensitivity	Specificity
Random	0.5000+/−0.0010	0.0072+/−0.0021	0.9929+/−0.0002
Chemical	0.8148+/−0.0019	0.5321+/−0.0046	0.9647+/−0.0004
Protein	0.8011+/−0.0021	0.5387+/−0.0038	0.9841+/−0.0002
Side-effect	0.8408+/−0.0036	0.5575+/−0.0046	0.9737+/−0.0004
Chemical+Protein	0.8295+/−0.0021	0.4014+/−0.0041	0.9921+/−0.0001
Chemical+Side-effect	0.8563+/−0.0022	0.6228+/−0.0071	0.9516+/−0.0006
Protein+Side-effect	0.8515+/−0.0053	0.5625+/−0.0070	0.9793+/−0.0003
Chemical+Protein+Side-effect	0.8640+/−0.0035	0.6195+/−0.0067	0.9650+/−0.0004

Table-1 Showing comparison of drug therapeutic-indication prediction with different data sources.

3. CONCLUSIONS

This paper developed a database application model and its service for drug discovery introducing our proposed .Software development process in MDA. We can discover two new services for drug discovery by determining new uses for old drugs in logical system model with big data analytics. In drug discovery application, we combine five Data Repositories into two main databases to reduce the complexity of the database. The rapid increase available data and optimizing their research processes a precondition for future success. Connoisseur should be rewarded for the rapid and consistent addition of their data to such a repository with senior management taking responsibility for the quality of this data. We need to improve the information literacy of medical at a faster rate than we are currently doing so, will have to become comfortable working in data -rich environment.

REFERENCES

[1] The official report on the side effects of prescription drugs that have increased dramatically over the past decade. In 2011, U.S. Food and Drug Administration (FDA) has received about 500,000 reports of health hazards and the death related to medical products per year.

[2] Lusher, S.J. et al. (2011) A molecular informatics view on best practice in multi-parameter compound optimization. Drug Discovery. Today 16, 555–568.

[3] Nicolaou, C.A. and Brown, N. (2013) Multi-objective optimization methods in drug design. Drug Discovery. Today Technol. 10, 427–435

[4] Nicolaou, C.A. et al. (2007) Molecular optimization using computational multi-objective methods. Curr. Opin. Drug Discovery. Dev. 10, 316–324.

[5] Segall, M.D. (2012) Multi-parameter optimization: identifying high quality compounds with a balance of properties. Curr. Pharm. Des. 18, 1292–1310.

[6] Slater, T. et al. (2008) Beyond data integration. Drug Discovery. Today 13, 584–589.

[7] Anatomical Therapeutic Chemical (ATC) Classification (2015). http://www.genome.jp/kegg-bin/get_htext?HYPERLINK "http://www.genome.jp/kegg-bin/get_htext?&"& extend=&htext=br08303.keg

[8] Hirakawa M, Tanaka T, Hashimoto Y, Kuroda M, Takagi1 T, Nakamura Y (2002) Jsnp: a database of common gene variations in the japanese population. Nucleic Acids Res 30(1):158–162

[9] Haga H, Yamada R, Ohnishi Y, Nakamura Y, Tanaka T (2002) Gene-based snp discovery as part of the Japanese millennium genome project : identification of 190,562 genetic variations in the human genome. J Hum Genet 47(11):605–610

[10] JSNP DATABASE (2015). http://snp.ims.u-tokyo.ac.jp/

[11] STITCH 4.0 (2014). http://stitch.embl.de/

[12] 17. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P (2010) A side effect resource to capture phenotypic effects of drugs. Mol Syst Biol 6(343). doi:10.1038/msb.2009.98

[13] SIDER 2 Side Effect Resource (2014). http://sideeffects.embl.de/

[14] "ATC/DDD Index". WHO Collaborating Centre for Drug Statistics Methodology.

[15] Gaulton, A. and Overington, J.P. (2010) Role of open chemical data in aiding drug discovery and design. Future Med. Chem. 2, 903–907

[16] Motherwell, S. (2004) Cheminformatics and crystallography. The Cambridge Structural Database. Database 1, 129–174

[17] Kirchmair, J. et al. (2008) The Protein Data Bank (PDB), its related services and software tools as key components for in silico guided drug discovery. J. Med. Chem. 51, 7021–7040

[18] Rose, P.W. et al. (2013) The RCSB Protein Data Bank: new resources for research and education. Nucleic Acids Res. 41, D475–D482

[19] Nicola, G. et al. (2012) Public domain databases for medicinal chemistry. J. Med. Chem. 55, 6987–7002

[20] Judd, D.B. (2013) Open innovation in drug discovery research comes of age. Drug Discov. Today 18, 315–317

[21] Zhou, Y. et al. (2007) Large-scale annotation of small-molecule libraries using public databases. J. Chem. Inf. Model. 47, 1386–1394

[22] Paolini, G.V. et al. (2006) Global mapping of pharmacological space. Nat. Biotechnol. 24, 805–815

[23] Howe, T.J. et al. (2007) Data reduction and representation in drug discovery. Drug Discov. Today 12, 45–53

[24] Mente, S. and Kuhn, M. (2012) The use of the R language for medicinal chemistry applications. Curr. Top. Med. Chem. 12, 1957–1964