

# Prediction of Best Drug Indications in Healthcare Sector using Classification Algorithm

R. Preethi  
M.E. CSE  
SCSVMV

M. Gayathri  
Assistant Professor  
Dept of CSE  
SCSVMV

## ABSTRACT

In the field of healthcare and medicine, several precautions should be taken by the healthcare professionals when prescribing drugs for patients. The side effects and effectiveness of drugs depends on the characteristics of patients such as age, gender, lifestyle and genetic profile. Existing techniques does not consider the side effects of drugs effectively with respect to the patient's personal medical profile. The goal is to provide a tool to assist the professionals and patients in finding and choosing the right drug using data mining technique. A hybrid approach is used which combines graph based approach and SVM algorithm to get even more effective drug Prediction. Graph based approach is used for disease prediction and as level 1 drug retrieval process and for further optimization we apply SVM Learning algorithm to predict the side effects association of the drug.

## Keywords

Graph based approach, SVM Learning Algorithm, Disease Prediction, Drug Prediction.

## 1. INTRODUCTION

For providing drug related information to medical professionals as well as to general people in the digital form, several efforts have been made. Web interfaces are used where the queries are submitted and answered by the experts [1][12]. DIC (Drug Information Center) plays an important role in answering the drug related queries. DIC's are majorly operated by Medical Schools and Government. Generally, several online sites are available for discussing and providing drug related information and queries to the public.

There are lots of drug reference resources available that gives guidance and detailed information about various drugs. But major problem is that it cannot be accessed by all people at all the time. Also, some people have difficulty of finding and using the drug available in their own medical box during emergency situation. Further, manual search may take more processing time and many people do not have much knowledge to use what kind of drug in particular situation.

In such a situation, where a patient or anyone else, wants to know details of drug such as its usage, side effects, directions, dosage, interactions, etc., [4]. And further the doctor-to-patient ratio is very low in our country. Availability of DIC information is very important to handle certain situations without contacting the medical practitioners anytime and anywhere.

Suggesting a drug to the patient is very tedious job because the medical practitioners have to consider all the factors related to patients health such as patient's past medical history, lifestyle, gender, age, allergic reactions, genetic profile, etc., [18]. Lots of online websites are available to prefer drugs but it does not consider all the factors and this

may results in the chance of inaccuracy in suggesting drugs[1].

Our aim is to provide a decision support tool to automatically analyze and predict the disease and suggest the corresponding drug by effectively considering all the factors in a timely and accurate manner [1]. To achieve this, we need a comprehensive drug information to answer the drug queries [1][12]. Different drug information sources such as DrugBank [2], SIDER2[4], KEGG[5] drug are available. But these data sources are available in different aspect, different coverage and may have some errors. So, we have to compare and combine these data sources into comprehensive drug information so as to reduce the noisiness and incompleteness.

## 2. RELATED WORK

Samamon Khemmarat and Lixin Gao proposed [1]. The aim of this proposal is to provide a tool to assist medical practitioners in prescribing drugs. It proposes an approach to query for drugs that satisfy a set of conditions based on drug properties and Addresses one of the major problems in supporting prescription, which is how to obtain and rank the answers. But it has following drawbacks associated with it. This system assumes that a query graph is given by users. For better usability, a module that assists users in constructing the query graphs, such as a form-based user interface or a module that translates natural language queries to query graph, is required. Also the answers obtained need to be reviewed by experts.

Charalampos Doulaverakis, George Nikolaidis, Athanasios Kleontas et al proposed [3]. It provides an efficient drug recommendations service where Semantic Web technologies are coupled with traditional business rule engines. It consumes less memory as the rule base is loaded during query time and with Drools being optimized to handle large rule bases efficiently. Following are the drawbacks that were addressed in this proposal. For enabling this service, medical information and terminology had to be translated to ontological terms and be appropriately coupled with medical knowledge of the field. Lack of probabilistic reasoning and the evaluation of framework is not carried out in real world data.

Akhil Langer, Rohit Banga, Ankush Mittal and et al proposed [8]. It presents the design of a drug QA system that could be used for providing information about medicines over short message service (SMS). The approach provides drug related information to people through natural language queries in a manner that meets their immediate need in a cost-effective manner. No installation of any kind is required on the user's mobile phone and the QA system can be quickly incorporated by a network service provider. The system meets emergency needs in an inexpensive manner which otherwise would have required consulting a doctor. But this system is having the

following problems. Dissemination of medical information using mobile phones is still in a nascent stage because of their limited features – lack of penetration of mobile internet, small screen size etc. This paper does not consider thoroughly the intricacies of SMS texting language (besides handling common spelling mistakes). Also the system requires minimal human involvement.

Qian zhu , cui tao , feichen shen ,christopher g. Chute proposed [9]. It uses pharmacogenomics data generated from pharmacogenomics studies, applied informatics and Semantic Web technologies to address the drug repositioning problem. However following drawbacks are addressed in this proposal. Traditional drug development is costly and labor-intensive, and scientists are devoted to finding an alternative way to facilitate the drug discovery process. In the current study, we explored only PharmGKB as a PGx resource, which is not enough to identify more novel associations for BCDs.

Jiahui Jinyz , Samamon Khemmaratz , Lixin Gaoz , Junzhou Luoy proposed [13]. It proposes an efficient algorithm for finding the best k answers for a given query without pre-computing graph indices. This approach efficiently answer the queries on billion-node information networks. But as far as Information networks are considered they are incomplete and noisy in nature. So it is important to discover answers that match exactly as well as answer that are similar to queries.

Asma Ben Abacha and Pierre Zweigenbaum proposed [12]. It provides a complete question analysis approach including medical entity recognition, semantic relation extraction and automatic translation to SPARQL queries. This approach deals with different types of questions, including questions with more than one expected answer type and more than one focus. It allows a deep analysis using different information extraction methods which allow extracting medical entities but also semantic relations and even additional information about the patient. But designing a question answering systems requires efficient and deep analysis of natural language questions. More specific processes are still required to deal with complex questions (e.g. why, when) and questions with new semantic relations that are not defined in our reference ontology.

Jiahui Jinyz, Samamon Khemmaratz, Lixin Gaoz, Junzhou Luoy proposed [14], in which the algorithm utilizes the bound-based top-k detection mechanism to obtain the answers quickly without indexing. The approaches provide meaningful answers and can efficiently compute answers for a query on massive information networks with billions of nodes. It is based majorly on indexing. That makes it a drawback of this proposal. The time and storage costs of indexing increase significantly when the indexing depth increases from 2-hop to 3-hop. 3-Hop still cannot provide accurate answers. To increase the accuracy, we have to increase the indexing depth further. As a result, the indexing costs would be unaffordable.

### 3. PROPOSED SYSTEM

#### 3.1 Architecture of Proposed System

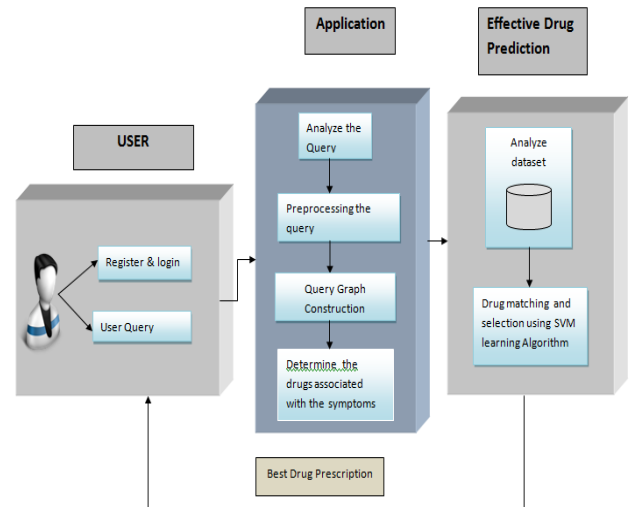


Fig 1: Architecture of the proposed system

#### 3.2 Description

The proposed system uses a machine learning algorithm called Support Vector Machine (SVM) algorithm to determine the maximum likelihood of the match. We use a score function for ranking the answers, which estimates the correlation between drugs and drug properties. Hybrid approach is used which uses both graph based approach and SVM algorithm to increase the efficiency

#### 3.3 Objectives of Proposed System

The main aim of the research study is to automatically analyze the disease and suggest the corresponding drugs and doctors. To effectively consider the interaction of drug with patient's medical profile such as allergic reactions, side effects, age, etc in a timely and accurate manner.

#### 3.4 Advantages

The Suggested System effectively analyzes the disease and predicts the best drug exactly. It considers the drug side effects, interactions and patient's medical profile effectively while predicting the drug. It suggests best doctor and arranges appointments with them through e-mail.

### 4. ALGORITHM/METHODOLOGY

#### 4.1 Graph Based Approach

In graph based approach, user query is represented as a Graph. Each node of the graph is called as Query node. The graph is represented as  $G(V_G; E_G; \text{type}_G; \text{key}_G)$ , where  $V_G$  represents the nodes and  $E_G$  represents the edges.  $\text{Type}_G$  maps the node to its Characteristics such as drug or drug properties.  $\text{key}_G$  maps the node to the keywords that identifies and describes the node.

Query nodes are classified as 2 types – variable node and reference node. Variable node shows the information that is searched by the user. Reference node contains the keywords specified by the user which acts as a pointer to the variable node. A positive or negative sign is associated with each edge of the query graph. If the connection should be available between two nodes, then it is indicated by a positive edge, otherwise it will be indicated as a negative edge.

#### 4.1.1 Graph Schema

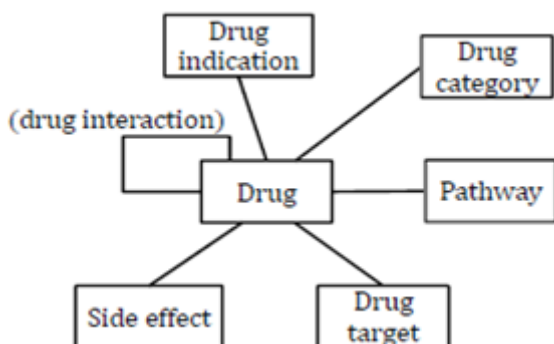


Fig 2: Schema of the Drug Graph

#### 4.1.2 Example Query Graph

Find a drug for fever and allergy that does not interact with Enoxaparin and does not cause drowsiness and nausea. In this query graph, there is one variable node,  $d_1$ , representing the drug to find. The drug  $d_1$  is connected to the two indications with positive edges. To avoid side effects and interactions, negative edges are used to connect  $d_1$  to the side effect nodes and another drug node.

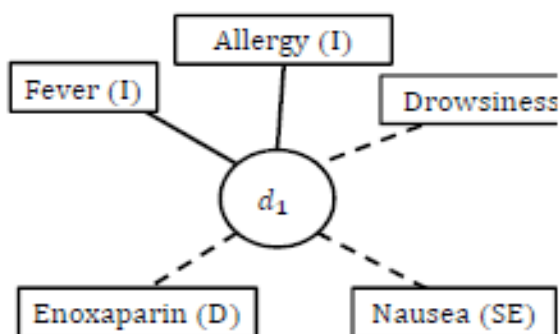


Fig 3: Query Graph Example

#### 4.2 Support Vector Machine (SVM)

SVM is combined with FG (Fine Grained Drug Features). Mainly helps in predicting drug-side effect associations. SVM+FG represent each drug with a binary vector which describes various types of drug properties such as indications, side effects, targets and chemical structure. SVM Learning algorithm trains a classifier from the various drug feature vectors. Selection is made by retaining only the features that falls under positive sample. For a drug  $d$ , the value in a dimension  $i$  is equal to 1 if drug  $d$  has the property associated with that dimension. Otherwise, the value is 0.

### 5. RESULTS AND DISCUSSIONS

For illustrating how the answers obtained by our approach are personalized according to a given patient profile, consider the following Query : **Find the drug for pneumonia without causing the side effect drowsiness.**

Table 1. Without Patients' profile

Rank	Drug	Score
1	Rocephin	0.99
2	Amoxil	0.99
3	<b>Biaxin</b>	0.98
4	<b>Azithromycin</b>	0.96
5	Cipro	0.95
6	Axert	0.76
7	Relpax	0.65
8	CiproXR	0.48
9	Chloromycetin	0.25
10	Procaine Penicillin	0.25

Table 2. With Patients' profile

Rank	Drug	Score
1	Rocephin	0.99
2	Amoxil	0.99
3	Cipro	0.95
4	Axert	0.76
5	<b>Azithromycin</b>	0.72
6	Relpax	0.65
7	<b>Biaxin</b>	0.59
8	CiproXR	0.48
9	Chloromycetin	0.25
10	Procaine Penicillin	0.25

The top 10 results obtained are compared when a user profile is not given and when the user profile is specified as female, elder in the above table. When the user profile is given, three of the drugs which are **Biaxin**, and **Azithromycin**, receive lower scores and thus lower ranks. These two drugs were reported as potential causes of Drowsiness in elder female patients. Our approach takes into account this fact and adjusts the scores of the drugs accordingly.

### 6. CONCLUSION

In this paper, we propose an approach for answering drug queries to support drug prescription. Our focus is on how to obtain and rank answers based on incomplete information and provide personalization. To cope with incomplete and noisy data, we allow both exact and close matches when answering queries. We also present an intuitive approach to display answers to users, which aims to help users to understand the ranked results and possibly refine their queries.

### 7. REFERENCES

- [1] S. Khemmarat and L. Gao, "Supporting drug prescription and personalized query system," in PervasiveHealth. IEEE, 2015.
- [2] C. Knox et al., "Drugbank 3.0: "A comprehensive resource for omics research on drugs," Nucleic acids research, vol. 39, no. supply 1, pp. D1035–D1041, 2011.

- [3] Charalampos Doulaverakis et al., “A Semantic-enabled Framework for Drug Recommendations”.
- [4] M. Kuhn et al., “A side effect resource to capture phenotypic effects of drugs,” *Molecular systems biology*, vol. 6, no. 1, p. 343, 2010.
- [5] M. Kanehisa and S. Goto, “Kegg: kyoto encyclopedia of genes and genomes,” *Nucleic acids research*, vol. 28, no. 1, pp. 27–30, 2000.
- [6] T. Fawcett, “An introduction to roc analysis,” *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.
- [7] K. Sangkuhl et al., “Pharmgkb: understanding the effects of individual genetic variants,” *Drug Metab. Rev.*, vol. 40, no. 4, pp. 539–551, 2008.
- [8] A. Langer et al., “A text based drug query system for mobile phones,” *Int. J. Mob. Commun.*, vol. 12, no. 4, pp. 411–429, Jul. 2014.
- [9] Qian Zhu et al., “Exploring the pharmacogenomics knowledge base (PHARMGKB) for repositioning breast
- [10] C. Doulaverakis et al., “Panacea, a semantic-enabled drug recommendations discovery framework,” *J. Biomed. Semant.*, vol. 5, p. 13, 2014.
- [11] M. Dumontier and N. Villanueva-Rosales, “Towards pharmacogenomics knowledge discovery with the semantic web,” *Briefings in bioinformatics*, vol. 10, no. 2, pp. 153–163, 2009.
- [12] A. Ben Abacha and P. Zweigenbaum, “Medical question answering: translating medical questions into sparql queries,” in *Proceedings of the 2nd ACM SIGHIT*. ACM, 2012, pp. 41–50.
- [13] Jiahui Jinyz, Samamon Khemmaratz, Lixin Gaoz, Junzhou Luoy., “Querying Web-Scale Information Networks Through Bounding Matching Scores”. *ACM 978-1-4503-3469-3/15/05.*, WWW 2015, May 18–22, 2015, Florence, Italy.
- [14] J. Jin et al., “A distributed approach for top-k star queries on massive Information networks,” in *ICPADS 2014*, 2014, pp. 9–16.
- [15] A. Khan et al., “Neighborhood based fast graph search in large networks,” In *SIGMOD 2011*. ACM, 2011, pp. 901– 912.
- [16] A. Khan et al., “Nema: Fast graph search with labelsimilarity,” in *Proc.of the VLDB Endowment*, vol. 6, no.3, 2013, pp. 181–192.
- [17] J. Sun, H. Xu, and Z. Zhao, “Network-assisted investigation of antipsychotic Drugs and their targets,” *Chemistry & biodiversity*, vol. 9, no. 5, pp. 900–910, 2012.
- [18] F. Cheng and Z. Zhao, “Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties,” *J. Am. Med. Inform. Assoc.*, vol. 21, no. e2, pp. e278–e286, 2014.