

# Survey on A Smart Medical System on Drug Analysis using ML Algorithms

**Prof. Manjunath N<sup>1</sup>, Ananya S<sup>2</sup>, Anushree S<sup>3</sup>, Chandana S<sup>4</sup>, Kavya S P<sup>5</sup>**

Assistant Professor, Department of Computer Science and Engineering<sup>1</sup>

Student, Department of Computer Science and Engineering<sup>2,3,4,5</sup>

Vidya Vikas Institute of Engineering and Technology, Mysuru, Karnataka, India

**Abstract:** Online health communities continue to offer huge variety of medical information useful for medical practitioners, system administrators and patients alike. In this system we collect real time health posts from reputed websites, where patients express their views, including their experiences and side-effects on drugs used by them. We propose to perform Summarization of user posts per drug, and come out with useful conclusions for medical fraternity as well as patient community at a glance. Further, we propose to classify the users based on their 'emotional state of mind'. Also, we shall perform knowledge discovery from user posts, whereby useful 'patterns' about the triad 'drugs-symptoms-medicine' is done by Association Rule Mining. Association rule mining is a popular and widely-known machine learning task. It is used to find out interesting relations between variables in large database. Rules generated by association have two disjoint set of items having form LHS (Left Hand Side) => RHS (Right Hand Side). The rule says that RHS is likely to occur whenever the LHS set occurs.

**Keywords:** Association Rule generation, Interesting Rule Selection, Symptoms, Disease, Medicine, Age group

## I. INTRODUCTION

With the enormous increase in web, electronic information is also increasing in huge amount which, although good with respect to Information Age, creates overhead of time and space. Also understand-ability of information and consequent knowledge continue to be big challenges For knowledge mining of the health posts, we propose to apply different important operations like - Association Rule Mining, Summarization and Sentiment Analysis on data obtained from the health forum site health-boards.com. Summarization is defined as taking information from the source, extracting content from it, and presenting the most useful content to the user in a condensed form and in a manner suitable to the user's application needs [1]. Summarization is very important in different NLP applications like Information Retrieval, Quality Analysis, and Text Comprehension etc. Commonly there are two types of summaries. First one is Extract in which contents from text i.e. words and sentences are reused. Second one is Abstract which includes regeneration of extracted contents [2] Association rule mining is a popular and widely-known machine learning task. It is used to find out interesting relations between variables in large database. Rules generated by association have two disjoint set of items having form LHS (Left Hand Side) => RHS (Right Hand Side). The rule says that RHS is likely to occur whenever the LHS set occurs [3]. Extraction of association rules includes two steps [4]: 1. Association Rule generation 2. Interesting Rule Selection After the rules have been obtained, they are extracted and post processed. The extracted rules from the health boards data-set could take one or more of the following form- 6 1. Symptoms->disease 2. Disease->disease 3. Medicine->disease 4. Disease->medicines 5. Age group->disease. Sentiment Analysis (SA) or Opinion Mining (OM) is task of finding sentiments from text. These sentiments may take different forms like - opinions from people, attitudes and emotions toward an entity. The entity can represent individuals, events or topics. These topics are most likely to be covered by reviews. Walaa Medhat considered Sentiment Analysis as a classification process. Classification levels considered were - document level, sentence level and aspect level [5]. While doing SA first the important features are selected from text then classification is done using appropriate classifier. We are considering reviews from health posts and in our case represented entity is drug. So, our classification falls in aspect level.

## II. RELATED WORKS

This Paper[1] To diagnose diabetes disease at an early stage is quite a challenging task due to complex inter dependence on various factors. There is a critical need to develop medical diagnostic decision support systems which can aid medical practitioners in the diagnostic process. For detection of diabetes at an early stage, data mining association algorithm is being used. Detecting the cause of diabetes is very important in order to stop diabetes. This paper mainly concentrates on only diabetes disease and related patterns, FP growth algorithm used which takes more time for data processing.

This Paper[2] Now a day there are number of online health communities provide a huge amount medical information that are posted by patients. The health post contains patients view, opinion or experience on particular drug. In this work we proposed to collect real time posts from websites like askapatient.com, healthboards.com etc. We proposed to summarize user posts using simplified Lesk algorithm. While summarizing the importance of each sentence is calculated using online dictionary called WordNet. This paper aims at comparative analysis of different sentiment analysis techniques which is used to process the patients opinions. Doesn't predict the relationship b/w symptoms-diseases-drugs.

This Paper[3] Online health communities offer tremendous medical information that could be available to all. However as is the case with other data intensive applications, this information contains hidden patterns which if explored, analyzed and understood could be very useful for administrators, medical practitioner sand patients alike. There are websites where patients express their experiences or side-effects on drugs used. In this work we collect such real time health posts from reputed websites, and perform data mining to determine the various possible associations from these posts. The methodology used- AIT algorithm, FP algorithm and apriori algorithm. The used algorithms takes more time for processing, less efficient, Small datasets used Take more time for processing.

This Paper[4]This paper proposes a new method of patterns analysis using SOM in medical data sets for medical treatment service under ubiquitous computing environment which is required by real time accessibility and agility. In this paper, it is necessary for us to classify disease patterns in the medical historical record to join the information of patient, using SOM neural network with input vectors of different features, disease code, input factors in order to take the medical treatment service in medical data sets, to reduce patients' search effort to get the information of diagnosis for recovering their health and to improve the rate of accuracy. To verify improved performance, we make experiments with data-set collected in medical center. Methodology used: SOM neural network and Limitations - Less parameters used Neural networks used which leads to less accurate results.

This Paper[5] Due to the sheer volume of opinion rich web resources such as discussion forum, review sites , blogs and news corpora available in digital form, much of the current research is focusing on the area of sentiment analysis. People are intended to develop a system that can identify and classify opinion or sentiment as represented in an electronic text. An accurate method for predicting sentiments could enable us, to extract opinions from the internet and predict online customer's preferences, which could prove valuable for economic or marketing research. Till now, there are few different problems predominating in this research community, namely, sentiment classification, feature based classification and handling negations.

This Paper[6] Automatic recognition of patterns is an important problem in a variety of engineering and scientific disciplines. The most common application in medicine relates to medical imaging for diagnostics' purpose. This paper presents another field of pattern recognition application – mining medical records. Analysis of treatment pathways (in which diagnostic and therapeutic procedures are sequenced on a timeline) as a mean of providing and standardized patient care of high quality, is one of the issues gaining in importance in the last decade in the field of medical informatics. Methodology used: Sequential Pattern prediction algorithms used. Limitations: Sequential patterns prediction algorithms are less efficient algorithms, Small medical datasets used.

This Paper[7] As huge amounts of knowledge are created rapidly, effective information access becomes an important issue. Especially for critical domains, such as medical and financial areas, efficient retrieval of concise and relevant information is highly desired. In this paper we propose a new user query based text summarization technique that makes use of Unified Medical Language System. Methodology used: ontology algorithms used. Limitations: Algorithms are less efficient algorithms, Take more time for processing.

This Paper[8] The paper introduces the significance of adverse drug reactions (ADRs) as a major source of morbidity and mortality. It emphasizes the need for predicting drug risk levels based on ADRs and presents the dataset from the Chinese spontaneous reporting database (CSR D). The paper highlights the imbalance in drug risk levels and introduces the Synthetic Minority Oversampling Technique (SMOTE) to address this issue. The evaluation metrics used for accessing the classification results might not capture the full spectrum of performance

This Paper[9] The paper introduces computational drug repositioning as a solution to the lengthy and expensive drug development process. Drug Repositioning (DR) focuses on finding new uses for existing drugs with established safety records. The authors highlight the increasing interest in DR and mention various methods proposed, including PREDICT, DRRS, NeoDTI, TL-HG BI, and DrugNet. Inspired by neuralnetwork-based approaches, the authors propose HNRD for drug-disease association prediction.

This Paper[10] The goal of this research is to provide a user-friendly, interactive web platform for mining association rules. Filtered Association, Apriori, Frequent Pattern Growth, Predictive Apriori, Generalized Sequential Patterns, Hotspot, and Tertius algorithms are among the association rule approaches employed in the produced program. The experimental findings are analyzed based on the outputs of the generated program, and an association rules mining application is done. Filtered Association, Apriori, Frequent Pattern Growth, Predictive Apriori, Generalized Sequential Patterns, hotspot, and Tertius algorithms are some of the methods used. Not suitable in real time. This paper major objective is on tumor detection (single disease). Objective is to find out no-recurrence-events, recurrence-events.

### III. METHODOLOGY

Step 1: In the 1 st step we collect the medical data (patient opinions[feedback] based on drugs). we referred sources like [www.patientslikeme.com](http://www.patientslikeme.com), [www.healthboard.com](http://www.healthboard.com) and [www.kaggle.com](http://www.kaggle.com).

Step 2: Patient opinions are then preprocessed and irrelevant data removed and only relevant data extracted and inputted to the algorithms.

Step 3: Then we input the necessary things required for algorithms, initially preprocessed patients opinions are summarized using lesk based algorithm and output of lesk based algorithm is inputted to Eclat algorithm to discover the medical patterns (symptomsdiseases-drugs).

Step 4: Eclat algorithm will discover the medical patterns which shows the relationship between symptoms with symptoms, symptoms with disease, disease with disease and disease with drug.

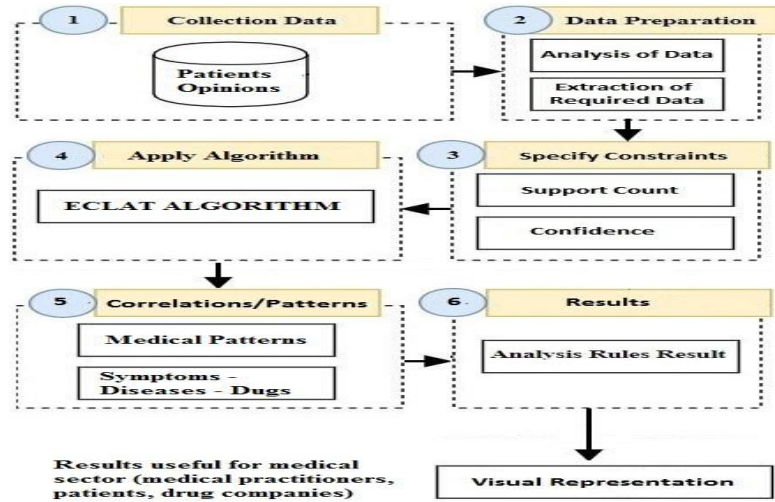
Step 5: Medical patterns displayed on GUI (front end).

Step 6: Results of the data science algorithms analyzed and represented visually

#### A. Figures and Tables

Figures and tables must be centered in the column. Large figures and tables may span across both columns. Any table or figure that takes up more than 1 column width must be positioned either at the top or at the bottom of the page. Graphics must not use stipple fill patterns because they may not be reproduced properly. Please use only SOLID FILL colors which contrast well both on screen and on a black-and-white hardcopy, as shown in Fig. 1.

The project architecture incorporates advanced technologies with a central server, user interfaces, and crucial Medical Practices. These practices create datasets on symptoms, diseases, drugs, and sentiments. User interaction involves an Opinion Mining System and Patients' Prediction for analyzing health opinions and predicting trends. The system efficiently processes large health datasets, and the user interface includes home, about, and contact pages.



#### IV. CONCLUSION

In this work, we collect real time health posts from reputed websites, and perform data mining to determine the various possible associations from these posts and perform knowledge discovery from user posts and detect useful 'patterns' about groups like: disease to disease, disease to drug and drug to symptom. This is done using Association learning algorithm. This will help the doctors to find side-effects of different drugs and with this they can prescribe better drugs to other patients with similar disease. Pharmaceutical companies can the response of several drugs on people and will get a idea about which drug is popular and should be produced. This will also help the patients to know about the opinion of previous users, thus will be in a better position to decide which medicine should be taken for a particular disease and also improve awareness on various side-effects of drugs faced by other people.

#### V. ACKNOWLEDGMENT

We felt great pleasure in submitting this paper on “A Survey on A Smart Medical System for Drug Analysis using ML Algorithms”. We are grateful to the editorial board of International Journal of Advanced Research in Computer Science and Technology(IJARSCT) for providing the opportunity to publish our paper. We also thank our faculty members for giving such chance to prove ourselves. We would like to express our sincere thanks to Professor Manjunath N for the best support, guidance, and preservation.

#### REFERENCES

- [1] JayashreeR,Srikanta Murthy K,Basavaraj .S.Anami, “Categorized Text Document Summarization in the Kannada Language by Sentence Ranking”, 12th International Conference on Intelligent Systems Design and Applications (ISDA), pp 776-781, 2012.
- [2] AlokRanjan Pal, DigantaSaha, “An Approach to Automatic Text Summarization using WordNet”, IEEE International Advance Computing Conference (IACC), 2014.
- [3] JesminNahar, Tasadduq Imam, Kevin S. Tickle, Yi-Ping Phoebe Chen, “Association rule mining to detect factors which contribute to heart disease in males and females”, J. Nahar et al. / Expert Systems with Applications 40 (2013) 1086–1093, Elsevier, 2012.
- [4] Lakshmi K.S, G. Santhosh Kumar, “Association Rule Extraction from Medical Transcripts of Diabetic Patients”,IEEE,2014.
- [5] WalaMedhat, Ahmed Hassan, HodaKorashy, “Sentiment analysis algorithms and applications: A survey”, In press, Elsevier, 2014.

- [6] Rafael Ferreira, Frederico Freitas, Luciano de Souza Cabral, Rafael DueireLins, Rinaldo Lima, Gabriel Franca, Steven J. Simske, Luciano Favaro, "A Context Based Text Summarization System", 11th IAPR International Workshop on Document Analysis Systems, pp 66-70, 2014.
- [7] C. Lakshmi Devasenal and M. Hemalatha, "Automatic Text Categorization and Summarization using Rule Reduction", IEEE- International Conference On Advances In Engineering, Science And Management (ICAESM - 2012), pp 594-598, 2012.
- [8] Sara Keretna, CheePeng Lim, Doug Creighton, "A Hybrid Model for Named Entity Recognition Using Unstructured Medical Text", Proc. Of the 2014 9th International Conference on System of Systems Engineering (SOSE), Adelaide, Australia- June 9-13, pp 85-90, 2014. 21
- [9] Saeed Mohajeri, Afsaneh Esteki, Osmar R. Zaiane and Davood Rafiei, "Innovative Navigation of Health Discussion Forums based on Relationship Extraction and Medical Ontologies", IEEE International Conference on Bioinformatics and Biomedicine, pp 13-14, 2013.
- [10] Yi Chen, Yunzhong Liu, "Connecting the Dots: Knowledge Discovery in Online Healthcare Forums", ICEC'14 August 05 - 06 2014, ACM.