Medication Diagonising Guidance Tool Using Machine Learning

N Bharath Kumar
Bachelor of Engineering, Under Graduate, Department of Computer Science and Engineering, Misrimal Navajee Munoth Jain Engineering College(Affiliated to Anna University),Chennai, Tamil Nadu, India.

Janaki E
Bachelor of Engineering, Under Graduate, Department of Computer Science and Engineering, Sri Muthukumaran Institute of Technology(Affiliated to Anna University), Chennai, Tamil Nadu, India.

Abstract – Medication errors are common in general practice and in hospitals. Errors in the act of prescription writing and prescribing faults due to erroneous medical decision can result in harm to patients. Medication errors cause at least one death everyday and injures approximately 5.2 million people annually in the India. Clinical choice emotionally supportive network, which utilizes progressed information mining procedures to help clinician make legitimate choice, has gotten extensive consideration as of late. The purpose of the project is to explore potential knowledge from diagnosis history records for guiding doctors to prescribe medication correctly and it also provide prescription based on the drug-drug interaction to decrease medication error effectively. The focal point of clinical choice emotionally supportive network incorporate not just enhancing analysis exactness additionally lessening conclusion time. Most diagnosis case data in hospitals is still kept untouched and has not been used for mining, so the value behind the data is hidden. Technologies as data mining and machine learning provide possibilities to explore knowledge from diagnosis history records. Experiments are done to evaluate the knowledge from the data mining , finally Support Vector Machine is selected for the medicine recommendation model. CDSS with support vector machine classifier has offered many advantages over a traditional health care systems and opens a new way to predict patient's disease.

Index Terms – Support Vector Machine Learning, Drug Drug Interaction, Clinical Decision Support System.

1. INTRODUCTION

1.1DOMAIN INTRODUCTION

1.1.1 MACHINE LEARNING

Machine learning is the subfield of computer science that gives computers the ability to learn without being explicitly programmed .Evolved from the study of pattern recognition and computational learning theory in artificial intelligence, machine learning explores the study and construction of algorithms that can learn from and make predictions on data such algorithms overcome following strictly static program instructions by making data driven predictions or decisions, through building a model from sample inputs. Machine learning is employed in a range of computing tasks where designing and programming explicit algorithms is infeasible; example applications include spam filtering, detection of network intruders or malicious insiders working towards a data breach, optical character recognition (OCR), search engines and computer vision.

Machine learning is closely related to (and often overlaps with) computational statistics, which also focuses in prediction-making through the use of computers. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning is sometimes conflated with data mining, where the latter subfield focuses more on exploratory data analysis and is known as unsupervised learning. Machine learning can also be unsupervised and be used to learn and establish baseline behavioral profiles for various entities and then used to find meaningful anomalies.

Within the field of data analytics, machine learning is a method used to devise complex models and algorithms that lend themselves to prediction; in commercial use, this is known as predictive analytics. These analytical models allow researchers, data scientists, engineers, and analysts to "produce reliable, repeatable decisions and results" and uncover "hidden insights" through learning from historical relationships and trends in the data.

1.1.2 DATA MINING

Data mining is the computational process of discovering patterns in large datasets involving methods at the intersection of artificial intelligence, machine learning, statistics and database systems. It is an interdisciplinary subfield of computer science. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data preprocessing, model and inference considerations,interestingnessmetrics, complexity c onsiderations, post processing of discovered

structures, visualization, and online updating. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD. The term is a misnomer, because the goal is the extraction of patterns and knowledge from amounts of data, not the extraction (mining) of data itself. It also is a buzzword and is frequently applied to any form of large scale data or information processing (collection, extraction, warehousing, analysis, and statistics) as well as any application of computer decision support system, including artificial intelligence, machine learning, and business intelligence. The book Data mining*:* Practical machine learning tools and techniques with Java (which covers mostly machine learning material) was originally to be named just Practical machine learning, and the term data mining was only added for marketing reasons. Often the more general terms (large scale) data analysis or, when referring to actual methods, artificial intelligence and machine learning are more appropriate.

The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown, interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection), and dependencies (association rule mining, sequential pattern mining). This usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system. Neither the data collection, data preparation nor result interpretation and reporting is part of the data mining step, but do belong to the overall KDD process as additional steps.

The related terms data dredging, data fishing, and data snooping refer to the use of data mining methods to sample parts of a larger population data set that are (or may be) too small for reliable satistical inferences to be made about the validity of any patterns discovered. These methods can however be used in creating new hypotheses to test against the larger data populations.

1.2 PROBLEM DEFINITION

A medication error is a failure in the treatment process that leads to, or has the potential to lead to, harm to the patients. To solve this problem, this system assists by presenting the likely disease and the possible medicine when a set of inputs are given. The prediction is made based on four symptoms, age, gender and certain biological factors.

1.3 PROJECT INTRODUCTION

One of the software system that have been enhancing in participation on recommendations and making decisions is recommender system. The system are able to adopt, select or recommend one element between big volumes of relative information. Diagnosis guidance system using machine learning(SVM) is about a recommendation method to recognize and treat disease by classification. This system is able to learn by the information cashed from the patients.

2. LITERATURE SURVEY

1. An intelligent medicine recommender framework Authors: YoujunBao, Xiaohong Jiang Paper Published: 2016

Youjun bao and xiaohong uses technologies as data mining and recommender technologies to provide possibility to explore potential knowledge from diagnosis history records and help doctors to prescribe medication correctly to decrease the medication error effectively. Universal medicine recommender system framework is implemented that applies data mining technologies to the recommendation system. Finally in the given open dataset SVM recommendation model is selected for the medicine recommendation module. Experimental results show their system can give medication recommendation with an excellent efficiency, accuracy, scalability.

3. SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

Existing system concentrate on the data collected for the particular chronic illness (Heart, Diabetes, Cancer) and provide recommendation. It apply the data mining in the hospital data base to find the pattern for the particular disease and use the machine learning technique to predict the seriousness of the disease. It also provide recommendation based on the information available online , it can leads to medication errors. It fails to concentrate on the heap amount of the patients data that is left in the hospital database.

3.2 PROPOSED SYSTEM

The proposed system is based on applying the data mining to the electronic records stored in the hospital database, and uses machine learning to predict the disease and medicine for the patients. On the contrary to the existing system, it identifies more than ten disease and medicine for those diseases.

Proposed system provide medicine recommendation to overcome the medication error done by the novice doctors. It uses the Support Vector Machine to make the prediction , since it provide the recommendation with higher efficiency.

4. SYSTEM REQUIREMENTS

4.1 HARDWARE REQUIREMENTS

The hardware requirements may serve as the basis for a contract for the implementation of the system and should therefore be a complete engineers as the starting point for the system design.

Table 1: Hardware Requirements

Hardware	Requirement
Processor	INTEL CORE I7 2.40 GHz
Ram	8 GB DD RAM
Monitor	15'' LCD, LCD Monitor
Hard disk	500 GB

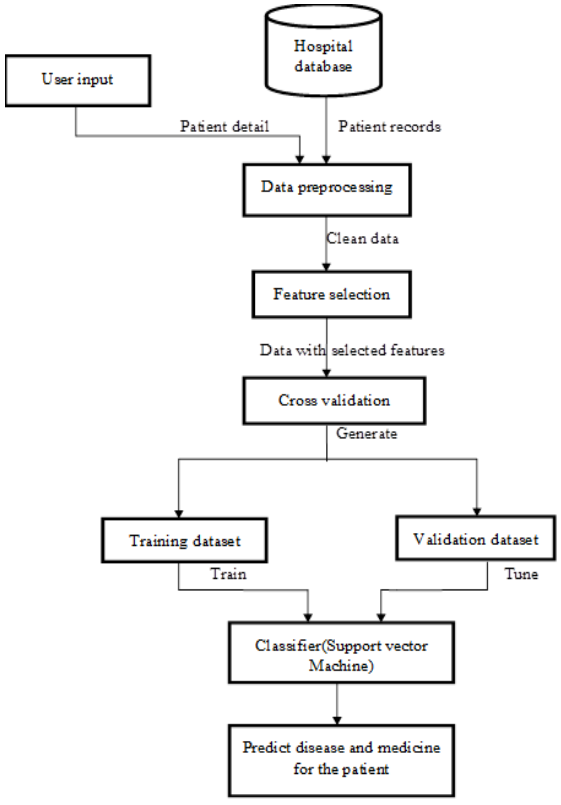
4.2 SOFTWARE REQUIREMENTS

The software requirements documents is the specification of the system. It should include both a definition and a specification of requirements. It is a set of what the system should rather than how it should do it. The software requirements provide a basis for creating software requirements specification.

Table 2 : Software Requirements

Software	Requirement
Front End	MAT LAB
Back End	MySQL
Operating System	Windows, Mac, Linux
IDE	MAT LAB R2016b

5. SYSTEM DESIGN



LIST OF MODULES

- Data pre processing module
- Feature selection module
- Classification module
- Cross validation module

6. SYSTEM IMPLEMENTATION

6.1 SOFTWARE DESCRIPTION

6.1.1 MATLAB

MATLAB is a high-performance language for technical computing. It integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation. Typical uses include:

- Math and computation
- Algorithm development
- Modelling, simulation, and prototyping
- Data analysis, exploration, and visualization
- Scientific and engineering graphics

Application development, including Graphical User Interface building

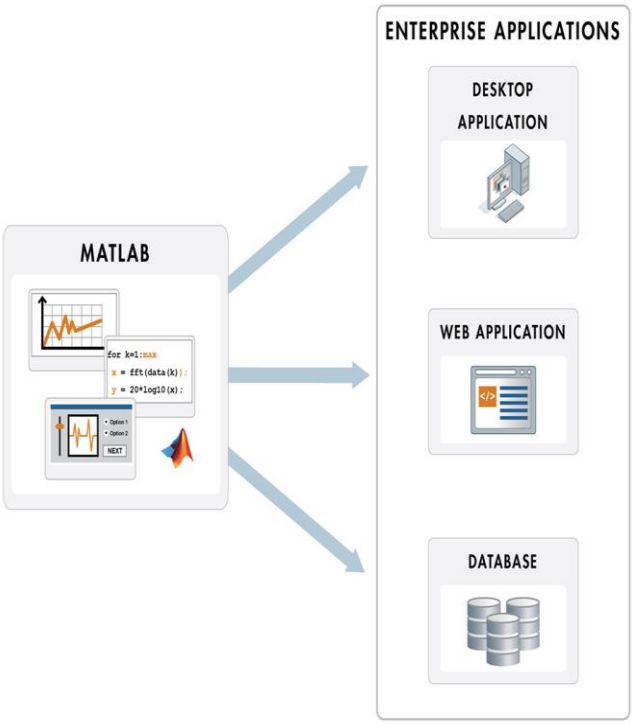


Figure 6.1: Software Development process

6.1.2 The MATLAB System

The MATLAB system consists of five main parts:

The MATLAB language.

This is a high-level matrix/array language with control flow statements, functions, data structures, input/output, and object-oriented programming features. It allows both "programming in the small" to rapidly create quick and dirty throw-away programs, and "programming in the large" to create complete large and complex application programs.

The MATLAB working environment.

This is the set of tools and facilities that you work with as the MATLAB user or programmer. It includes facilities for managing the variables in your workspace and importing and exporting data. It also includes tools for developing, managing, debugging, and profiling M-files, MATLAB's applications.

Handle Graphics.

This is the MATLAB graphics system. It includes high-level commands for two-dimensional and three-dimensional data visualization, image processing, animation, and presentation graphics. It also includes low-level commands that allow you to fully customize the appearance of graphics as well as to build complete Graphical User Interfaces on your MATLAB applications.

The MATLAB mathematical function library.

This is a vast collection of computational algorithms ranging from elementary functions like sum, sine, cosine, and complex arithmetic, to more sophisticated functions like matrix inverse, matrix eigenvalues, Bessel functions, and fast Fourier transforms.

The MATLAB Application Program Interface (API).

This is a library that allows you to write C and Fortran programs that interact with MATLAB. It include facilities for calling routines from MATLAB (dynamic linking), calling MATLAB as a computational engine, and for reading and writing MAT-files.

6.1.3 MYSQL

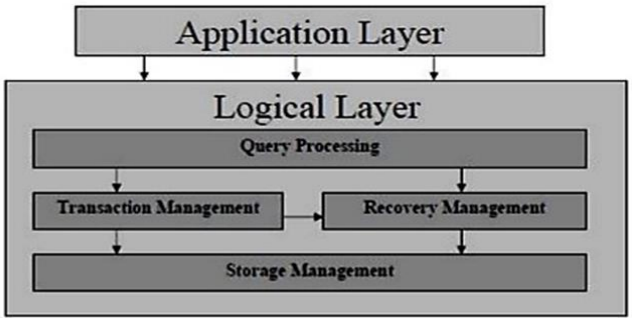


Figure 6.2 MYSQL architecture diagram

7. SYSTEM TESTING

7.1 SYSTEM TESTING

Testing is performed to identify errors. It is used for quality assurance. Testing is an integral part of the entire development and maintenance process. Testing is a set of activities that can be planned in advance and conducted systematically. For this reason a template for software testing, a set of steps into which we can place specific test case design techniques and testing methods should be defines for software process.

Testing often accounts for more effort than any other software engineering activity. If it is conducted haphazardly, time is wasted, unnecessary effort is expanded, and even worse, errors sneak through undetected. It would therefore seem reasonable to establish a systematic strategy for testing software.

8. CONCLUSION AND FUTURE ENHANCEMENT

8.1 CONCLUSION

This project provide a way to rectify the medication error by mining the patients details. Patient records which is stored in database are mined to get the knowledge and then the SVM(Support Vector Machine) is implemented to predict the disease and medicine to the patients. It helps to reduce the medication error and provide decision support to the doctor. It also uses the patient history to provide the recommendation based on the drug-drug interaction so that the medication error can be reduced effectively.

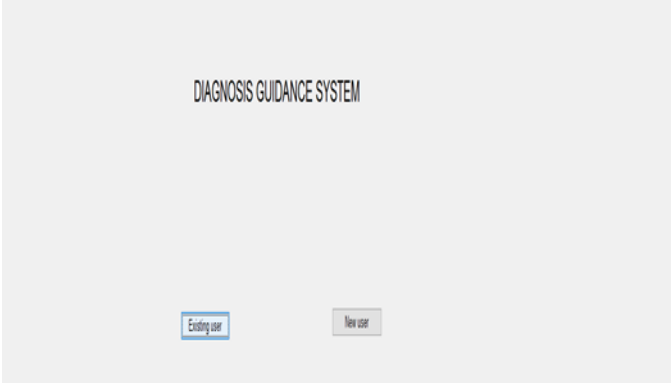
8.2 FUTURE ENHANCEMENT

In future , big data from the healthcare industry are mined using data mining techniques and the map reduce for the machine learning are implemented on the healthcare data to provide recommendation.

9. SAMPLE

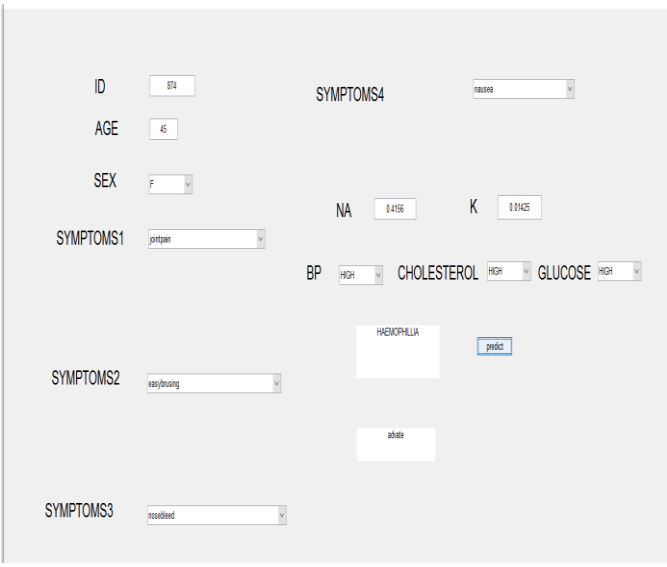
A.2.1 HOME PAGE

This is the home page which contains the options for the new patients and the options for the existing patients.



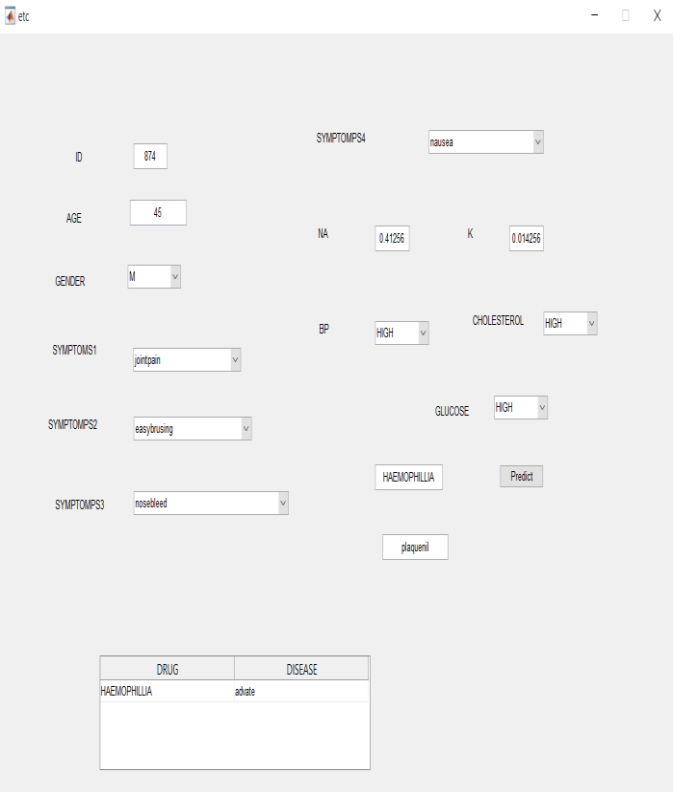
A.2.2 RECOMMENDER PAGE FOR THE NEW PATIENTS

Registration form contains the necessary fields that has to be filled to get the diagnosis guidance for the new patients.



A,2.3 RECOMMENDER PAGE FOR THE EXISTING PATIENTS

This provide the guidance based on the symptoms experienced by the patients and the patient history.



REFERENCES

[1] Chang C C, Lin C J. LIBSVM : a library for support vector machine. ACM Transactions on Intelligent systems and Technology(TIST),2011,2(3):27
[2] O.R. zaiane, Building a recommender agent for e-learning systems, Proceedings of 2002 International Conference on Computers in Educations, 2002, 55-59
[3] Poli R, Cagnoni S, Livi R, et al. A neural network expert system for diagnosis and treating hypertension. Computer,1991,24(3):64-71
[4] Rahul majetha, Varun mishra, Akshit Singhal, Lakshmi Manasa K, Kunchay sahithi, Vijay Nandwani People save: Recommending effective drugs through web crowd sourcing, COMSNET 2016- Net Health Workshop.
[5] Ronda G. Hughes and Eduardo Ortiz , Medication error why they happen , how they can be prevented. State of the Science on State Medicine Administration, 2005.
[6] Sebastian Maldonado and Richard weber , A wrapper method for feature selection using Support Vector Machine on Information science 179(2009) 2208-2217
[7] Soni J, Ansari U, Sharma D, et al Predictive data mining for medical diagnosis : An overview of heart disease prediction. International journal of computer applications, 2011, 17(8): 43-48.
[8] S.Vijayarani, Ilamathi, Nithya , Preproceesing technique for text mining – An Overview , International Journal of Computer Science and Communication Networks Vol-5(1) 7-16.
[9] Vijay hudder, Babu Koundinya desiraju, Vaibhav Rajan,Sakyajit Bhattacharya, Predicting Complications in critical care using heterogenous clinical data , IEEE Transaction 2016.
[10] Ximeng Liu, Rongxing Lu, Jianfeng Ma, Le Chen and Baodong Qin , Privacy-Preserving Patient-Centric Clinical Decision Support System on Naïve Bayesian Classification. IEEE Journal of Biomedical And Health Informatics, Vol .20, NO.2, 2016
[11] www.mathworks.com
[12] www.coursera.org/learn/machine-learning
[13] https://www.ncbi.nlm.nin.gov
[14] www.kaggle.com
[15] www.drugs.com