Online Medical Support System

Shivani J. Sawarkar[1], Dr. Gitanjali Shinde[2]
Department of Computer Engineering,
Smt. Kashibai Navale College of Engineering, Savitribai Phule, Pune University, Pune, India.

***Abstract* -** In our society, humans pay more attention to their own fitness. Personalized fitness service is regularly rising. Due to lack of skilled doctors and physicians, maximum healthcare corporations cannot meet the clinical call of public. Public want extra accurate and on the spot result. Thus, increasingly more facts mining packages are evolved to provide humans extra custom designed healthcare provider. It is a good answer for the mismatch of insufficient clinical assets and growing medical demands. Here an AI-assisted prediction device is advocated which leverages information mining strategies to show the relationship between the everyday physical examination information and the capability fitness danger given through the consumer or public. The main concept is to decide clinical illnesses in step with given signs and symptoms & every day Routine where user search the sanatorium then given the closest medical institution of their cutting-edge area.

***Keywords* -** Data mining, machine learning and disease prediction.

I. INTRODUCTION

Many healthcare organizations (hospitals, medical centers) in China are busy in serving people with best-effort healthcare service. Nowadays, people pay more attention on their physical conditions. They want higher quality and more personalized healthcare service. However, with the limitation of number of skilled doctors and physicians, most healthcare organizations cannot meet the need of public. How to provide higher quality healthcare to more people with limited manpower becomes a key issue. The healthcare environment is generally perceived as being 'information rich' yet 'knowledge poor'. Hospital information systems typically generate huge amount of data which takes the form of numbers, text. . There is a lot of hidden information in these data untouched. Data mining and predictive analytics aim to reveal patterns and rules by applying advanced data analysis techniques on a large set of data for descriptive and predictive purposes. There is a wealth of data available within the healthcare systems. However, there is a lack of effective analysis tools to discover hidden relationships and trends in data. This process is inefficient, as each suspicious access has to be reviewed by a security expert, and is purely retrospective, as it occurs after damage may have been incurred. [1] Data mining is suitable for processing large datasets from hospital information system and finding relations among data features. The list of challenges in order of importance that they be solved if patients and organizations are to begin realizing the fullest benefits possible of these systems consists of: improve the human–computer interface; disseminate best practices in CDS design, development, and implementation; summarize patient-level information; prioritize and filter recommendations to the user; create an architecture for sharing executable CDS modules and services; combine recommendations for patients with co-morbidities; prioritize CDS content development and implementation; create internet-accessible clinical decision support repositories; use free text information to drive clinical decision support; mine large clinical databases to create new CDS[2] It takes only a few researchers to analyze data from hospital information.. Knowledge discovery and data mining have found numerous applications in business and scientific domain.[3]The main concept is to determine medical diseases according to given symptoms & daily routine when user search the hospital then given the nearest hospital of their current location. Data mining techniques used in the prediction of heart attacks are rule based, decision trees, artificial neural networks. [4] The related queries are based in previously issued queries, and can be issued by the user to the search engine to tune or redirect the search process. The method proposed is based on a query clustering process in which groups of semantically similar queries are identified [5]. The clustering process uses the content of historical preferences of users registered in the query log of the search engine The system provides a user-friendly interface for examinees and doctors. Examinees can know their symptoms which accrued in body which set as the while doctors can get a set of examinees with potential risk. A feedback mechanism could save manpower and improve performance of system automatically.

II. MOTIVATION

Previous medical examiner only used basic symptoms of particular diseases but in this application examiner examines on the word count, laboratory results and diagnostic data. A feedback mechanism could save manpower and improve performance of system automatically. The doctor could fix prediction result through an interface, which will collect doctors' input as new training data. An extra training process will be triggered everyday using these data. Thus, this system could improve the performance of prediction model automatically. When the user visits hospital physically, then user's personal record is saved and then that record is added to the examiner data set. It consumes lot of time.

III. REVIEW OF LITERATURE

Sittig D, Wright A, Osheroff J, et al. [1] There is a pressing need for high-quality, effective means of designing, developing, presenting, implementing, evaluating, and maintaining all types of clinical decision support capabilities for clinicians, patients and consumers. Using an iterative, consensus-building process we identified a rank-ordered list of the top 10 grand challenges in clinical decision support. This list was created to educate and inspire researchers, developers, funders, and policy-makers. The list of challenges in order of importance that they be solved if patients and organizations are to begin realizing the fullest benefits possible of these systems consists of: improve the human–computer interface; disseminate best practices in CDS design, development, and implementation, summarize patient-level information; prioritize and filter recommendations to the user; create an architecture for sharing executable CDS modules and services; combine recommendations for patients with co-morbidities; prioritize CDS content development and implementation; create internet-accessible clinical decision support repositories; use free text information to drive clinical decision support; mine large clinical databases to create new CDS. Identification of solutions to these challenges is critical if clinical decision support is to achieve its potential and improve the quality, safety and efficiency of healthcare.

Anderson J E, Chang DC. Et al [2] Many healthcare facilities enforce security on their electronic health records (EHRs) through a corrective mechanism: some staff nominally have almost unrestricted access to the records, but there is a strict ex post facto audit process for inappropriate accesses, i.e., accesses that violate the facility's security and privacy policies. This process is inefficient, as each suspicious access has to be reviewed by a security expert, and is purely retrospective, as it occurs after damage may have been incurred. This motivates automated approaches based on machine learning using historical data. Previous attempts at such a system have successfully applied supervised learning models to this end, such as SVMs and logistic regression. While providing benefits over manual auditing, these approaches ignore the identity of the users and patients involved in a record access. Therefore, they cannot exploit the fact that a patient whose record was previously involved in a violation has an increased risk of being involved in a future violation. Motivated by this, in this paper, a collaborative filtering inspired approach to predict inappropriate accesses is proposed. Our solution integrates both explicit and latent features for staff and patients, the latter acting as a personalized "finger-print" based on historical access patterns. The proposed method, when applied to real EHR access data from two tertiary hospitals and a file-access dataset from Amazon, shows not only significantly improved performance compared to existing methods, but also provides insights as to what indicates an inappropriate access.

ZhaoqianLan, Guopeng Zhou, YichunDuan , Wei Yan et al [3] healthcare environment is generally perceived as being 'information rich' yet 'knowledge poor'. There is a wealth of data available within the Health-care systems. However, there is a lack of effective analysis tools to discover hidden relationships and trends in data. Knowledge discovery and data mining have found numerous applications in business and scientific domain. Valuable knowledge can be discovered from application of data mining techniques in healthcare system. In this study, the potential use of classification based data mining techniques such as rule based, decision tree, naïve bayes and artificial neural network to massive volume of healthcare data is briefly examined. The healthcare industry collects huge amounts of healthcare data which, unfortunately, are not "mined" to discover hidden information. For data preprocessing and effective decision making One Dependency Augmented Naïve Bayes classifier (ODANB) and naive creedal classifier 2 (NCC2) are used. This is an extension of naïve Bayes to imprecise probabilities that aims at delivering robust classifications also when dealing with small or incomplete data sets. Discovery of hidden patterns and relationships often goes unexploited. Using medical profiles such as age, sex, blood pressure and blood sugar it can predict the likelihood of patients getting a heart disease. It enables significant knowledge, e.g. patterns, relationships between medical factors related to heart disease, to be established.

Srinivas K, Rani B K, Govrdhan A. et al [4]. In this paper, care services through telemedicine is provided and it has become an important part of the medical development process, due to the latest innovation in the information and computer technologies. Meanwhile, data mining, a dynamic and fast-expanding domain, has improved many fields of human life by offering the possibility of predicting future trends and helping with decision making, based on the patterns and trends discovered. The diversity of data and the multitude of data mining techniques provide various applications for data mining, including in the healthcare organization. Integrating data mining techniques into telemedicine systems would help improve the efficiency and effectiveness of the healthcare organizations activity, contributing to the development and refinement of the healthcare services offered as part of the medical development process.

Gheorghe M, Petre R. et al [5] In this paper a method is proposed that, given a query submitted to a search engine, suggests a list of related queries. The related queries are based in previously issued queries, and can be issued by the user to the search engine to tune or redirect the search process. The method proposed is based on a query clustering process in which groups of semantically similar queries are identified. The clustering process uses the content of historical preferences of users registered in the query log of the search engine. The method not only discovers the related queries, but also ranks them according to a relevance criterion. Finally, with experiments over the query log of a search engine is shown and the effectiveness of the method.

R. Baeza-Yates, C. Hurtado, and M. Mendoza, et al [6], the system have focused to compare a variety of techniques, approaches and different tools and its impact on the healthcare sector. The goal of data mining application is to turn that data are facts, numbers, or text which can be processed by a computer into knowledge or information. The main purpose of data mining application in healthcare systems is to develop an automated tool for identifying and disseminating relevant healthcare information. This paper aims to make a detailed study report of different types of data mining applications in the healthcare sector and to reduce the complexity of the study of the healthcare data transactions. Also presents a comparative study of different data mining applications, techniques and different methodologies applied for extracting knowledge from database generated in the healthcare industry. Finally, the existing data mining techniques with data mining algorithms and its application tools which are more valuable for healthcare services are discussed in detail.

Koh H C, Tan G.et al [7] many healthcare facilities enforce security on their electronic health records (EHRs) through a corrective mechanism: some staff nominally have almost unrestricted access to the records, but there is a strict ex post facto audit process for inappropriate accesses, i.e., accesses that violate the facility's security and privacy policies. This process is inefficient, as each suspicious access has to be reviewed by a security expert, and is purely retrospective, as it occurs after damage may have been incurred. This motivates automated approaches based on machine learning using historical data. Previous attempts at such a system have successfully applied supervised learning models to this end, such as SVMs and logistic regression. While providing benefits over manual auditing, these approaches ignore the identity of the users and patients involved in a record access. Therefore, they cannot exploit the fact that a patient whose record was previously involved in a violation has an increased risk of being involved in a future violation. Motivated by this, in this paper, a collaborative filtering inspired approach to predict inappropriate accesses is proposed. The solution integrates both explicit and latent features for staff and patients, the latter acting as a personalized "finger-print" based on historical access patterns. The proposed method, when applied to real EHR access data from two tertiary hospitals and a file-access dataset from Amazon, shows not only significantly improved performance compared to existing methods, but also provides insights as to what indicates an inappropriate access.

Tao Jiang & Siyu Qian, et al. [8] The study aimed to identify risk factors in medication management in Australian residential aged care (RAC) homes. Only 18 out of 3,607 RAC homes failed aged care accreditation standard in medication management between 7th March 2011 and 25th March 2015. Text data mining methods were used to analyse the reasons for failure. This led to the identification of 21 risk indicators for an RAC home to fail in medication management. These indicators were further grouped into ten themes. They are overall medication management, medication assessment, ordering, dispensing, storage, stock and disposal, administration, incident report, monitoring, staff and resident satisfaction. The top three risk factors are: "ineffective monitoring process" (18 homes), "noncompliance with professional standards and guidelines" (15 homes), and "resident dissatisfaction with overall medication management" (10 homes).

Song J H, Venkatesh S S, Conant E A, et al. [9], the k-means clustering and self-organizing maps (SOM) are applied to analyze the signal structure in terms of visualization. k-nearest neighbor classifiers (k-nn), support vector machines (SVM) and decision trees (DT) are employed to classify features using a computer aided diagnosis (CAD) approach.

Song J H, Venkatesh S S, Conant E A, et al. [10], Breast cancer is one of the most common cancers in women. Sonography is now commonly used in combination with other modalities for imaging breasts. Although ultrasound can diagnose simple cysts in the breast with an accuracy of 96%–100%, its use for unequivocal differentiation between solid benign and malignant masses has proven to be more difficult. Despite considerable efforts toward improving imaging techniques, including solography, the final confirmation of whether a solid breast lesion is malignant or benign is still made by biopsy.

IV. GAP ANALYSIS

Sr. No.	Baseline algorithm	Partition algorithm	Decision tree	XG Boost	Random algorithm	SVM	KNN
1.	N	N	N	N	N	N	N
2.	Y	Y	N	N	N	N	N
3.	N	N	Y	Y	Y	N	N
4.	N	N	Y	N	N	N	N
5.	N	N	N	N	N	N	N
6.	N	N	N	N	N	N	N
7.	N	N	N	N	N	N	N
8.	N	N	N	N	N	Y	N
9.	N	N	N	N	N	N	N
10.	N	N	Y	N	N	Y	Y

V. PROPOSED SYSTEM

After the analysis of the previous system this system's main concept is to determine medical diseases according to given symptoms & daily routine and when user search the hospital then the nearest hospital of their current location is given. The system provides a user-friendly interface for examinees and doctors. Examinees can know their symptoms which occurred in body while doctors can get a set of examinees with potential risk. A feedback mechanism could save manpower and improve performance of system automatically. The doctor could fix prediction result through an interface, which will collect doctors' input as new training data. An extra training process will be triggered everyday using these data. Thus, our system could improve the performance of prediction model automatically.

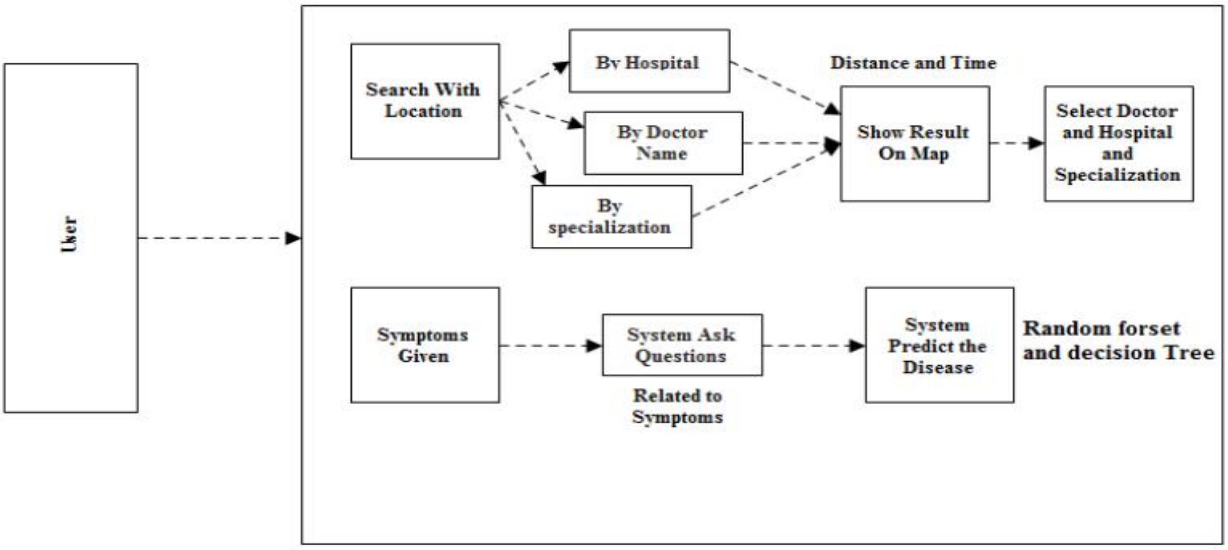


Figure 1: System Overview

This system increases human-computer interactions. Location of user is detected. Also the hospital and doctor is recommended to the patient according to the prediction of the disease. Medicines are provided for the predicted disease. This prediction system is fast, scalable and low-cost.

A. Algorithms:

i). Random Forest Algorithm - The beginning of random forest algorithm starts with randomly selecting **"k"** features out total **"m"** features. In the image, it is seen that features and observations are randomly taken. In the next stage, the randomly selected **"k"** features are used to find the root node by using the best split approach. In the next stage, the daughter nodes are calculated using the same best split approach. The first 3 stages until we form the tree with a root node and having the target as the leaf node. Finally, 1 to 4 stages are repeated to create **"n"** randomly created trees. This randomly created tree forms the **random forest.**

ii). Partition based Algorithm - Implement the algorithm and test itand instrument the algorithm so that it counts the number of comparisons of array elements. (Don't count comparisons between array indices.) Test it to see if the counts "make sense". For values of n from 500 to 10000, run a number of experiments on randomly-ordered arrays of size n and find the average number of comparisons for those experiments. Graph the average number of comparisons as a function of n and repeat the above items 1–4, using an alternative pivot selection method.

iii). Baseline algorithm - Implement the algorithm and test it to find the RWR based top-m query recommend. Start from one unit of active ink injected into node Kq and the order in descending order. Find the weight of each edge e is adjusted based on λq.The algorithm returns the top-m candidate suggestionsother than kq in C as the result.

B. Mathematical Model: Let us consider S as a set of prediction and medicines to be given.

S= { }

INPUT:

Identify the inputs as symptoms and search hospital.

F= $\{f_1, f_2, f_3 \ldots, f_n|$ 'F' as set of functions to execute similar diseases and medicine }

I= $\{i_1, i_2, i_3\ldots|$'I' sets of inputs to the symptoms}//

O= $\{o_1, o_2, o_3\ldots|$'O' set of outputs from the function sets}//

S= {I, F, O}

I = {symptoms and search hospital, Search By doctor name, search by specialization}

O = {predicted the diseases and medicine}

F = {prediction techniques and baseline algorithms}

VI. RESULT AND DISSCUSSION

Now we evaluate the performance of the proposed algorithm using the data set from different Hospital of Pune In our system, we divide the data sets into training set and testing set by the ratio 2: 1. Then our system applies the two algorithms mentioned above in risk-prediction task of three symptoms. Few questions related to the symptoms are asked by the system and the user has to give the required answer.

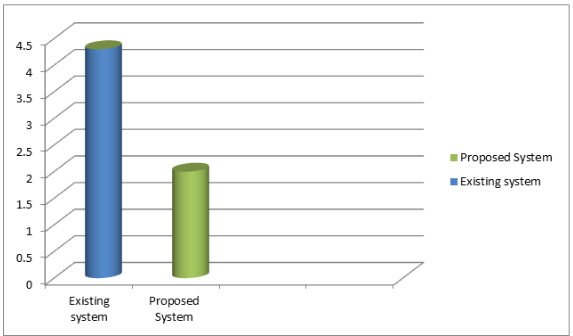
In our system User search with the keyword and gets result according to the current location.
Thus, we introduce some metrics in prediction problems to evaluate performance of proposed system. We define the true positive set as TP = {i|O(E_i) = 1,$O^'(E^1)$ = 1}, the false positive set as FP = {i|O(E_i) = 0,$O^'(E^1)$ = 1}, the true negative is TN = {i|O(E_i) = 0,$O^'(E^1)$ = 0}, and the false negative is FN = {i|O(E_i) = 1,$O^'(E^1)$ = 0}. Then precision and recall is:

Table 1: Compare of Algorithms in Precision and Recall methods

S.No	ALGORITHMS	PRECISION / RECALL	RESULT
01	XG Boost	0.7997 and 0.8146	79.0%
02	Random Forest and Decision Tree	0.7866 and 0.8039	77.7%

Table 2: Compare of Algorithms with different parameter

S.No	ALGORITHMS	PARAMETER	NO OF NODES
01	Baseline Algo-rithms	Slow	n^n
02	Shortest Path Estimation	Large	n^{n+1}
03	Floyd-Warshall's	Very Large	n^{n+n}



VII. CONCLUSION

This project implements an AI-assisted prediction system which leverages data mining methods to reveal the relationship between the regular physical examination records and the potential health risk given by the user or public. Different machine learning algorithms are applied to predict physical status of examinee will be in danger of physical deterioration next year. In this system user or patient search the hospital, then results are given according to the nearest location of current location of user/patients. User / Patients gives symptoms and the system will predict the diseases and will give the medicines. A feedback mechanism is also designed for doctors to fix classification result or input new training data, and the system will automatically rerun the training process to improve performance every day.

VIII. ACKNOWLEDGEMENT

The authors would like to thank the researchers as well as publishers for making their resources available and teachers for their guidance. We are thankful to the authorities of Savitribai Phule University of Pune and, for their constant guidelines and support. We are also thankful to the reviewer for their valuable suggestions. We also thank the college authorities for providing the required infrastructure and support. Finally, we would like to extend a heartfelt gratitude to friends and family members.

IX. REFERENCES

[1]. Sittig D, Wright A, Osheroff J, et al. "Grand challenges in clinical decision support". Journal of Biomedical Informatics, 2008.
[2]. Anderson J E, Chang D C. "Using Electronic Health Records for Surgical Quality Improvement in the Era of Big Data"[J]. Jama Surgery, 2015.
[3]. ZhaoqianLan, Guopeng Zhou, YichunDuan , Wei Yan , "AI-assisted Prediction on Potential Health Risks with Regular Physical Examination Records", IEEE Transactions On Knowledge And Data Science ,2018.
[4]. Srinivas K, Rani B K, Govrdhan A. "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks". International Journal on Computer Science & Engineering, 2010.
[5]. Gheorghe M, Petre R. "Integrating Data Mining Techniques into Telemedicine Systems" InformaticaEconomica Journal, 2014.
[6]. R. Baeza-Yates, C. Hurtado, and M. Mendoza, "Query recommendation using query logs in search engines," in Proc. Int. Conf. Current Trends Database Technol., 2004, pp. 588–596.
[7]. Koh H C, Tan G. Data mining applications in healthcare. [J]. Journal of Healthcare Information Management Jhim, 2005, 19(2):64-72.
[8]. Menon A K, Jiang X, Kim J, et al. Detecting Inappropriate Access to Electronic Health Records Using Collaborative Filtering[J]. Machine Learning, 2014, 95(1):87-101.
[9]. Tao Jiang & Siyu Qian, et al. Accreditation Reports to Identify Risk Factors in Medication Management in Australian Residential Aged Care Homes[J]. Studies in Health Technology & Informatics, 2017
[10]. Song J H, Venkatesh S S, Conant E A, et al. Comparative analysis of logistic regression and artificial neural network for computer-aided diagnosis of breast masses.[J]. Academic Radiology, 2005, 12(4):487-95.