Prediction on Health Care based on Near Search by Keyword

Alice Mary Marshal Rajan
Department of Computer Engineering
P.E.S Modern College of Engineering

Prof Mrs. Deepti Nirwal
Department of Computer Engineering
P.E.S Modern College of Engineering

ABSTRACT:-

In our society, humans have more attention to their own fitness. Personalized fitness service is regularly rising. The lack of skilled doctors and physicians, maximum healthcare corporations cannot meet the clinical call for of public. Public expects 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. We advocate an AI-assisted prediction device 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 on every day basis. The main concept is to decide medicinal ailments as per given side effects and every day routine given by the user, along with the ability of providing the closest doctor's facility of that present area. The system provides a user-friendly interface for examinees and doctors. Examinees can know the various side effects occurring in their body while the specialists can get a lot of examinees with potential hazard. The machine gives a user-friendly interface for examinees and medical doctors. The doctor should restore prediction result via an interface, which will accumulate medical doctors' entry as new training information. Thus, enhancing the overall performance of prediction.

Keyword: - *Data Mining, Machine Learning and diseases prediction*

I. INTRODUCTION

Many healthcare companies (hospitals, medical facilities) are busy in serving humans with quality-attempt healthcare carrier. Nowadays, humans pay extra interest on their bodily situations. They need higher first-class and more customized healthcare provider. However, with the limitation of number of skilled medical doctors and physicians, most healthcare agencies cannot meet the need of public. How to offer better first-class healthcare to more people with restrained manpower turns into a key problem. The healthcare environment is usually perceived as being 'facts wealthy' but 'understanding bad'. Hospital facts structures usually generate big amount of records which takes the form of numbers, text. There is lots of hidden records in these data untouched. Data mining and predictive analytics goal to reveal patterns and policies through applying advanced facts analysis strategies on a large set of facts for descriptive and predictive purposes. Data mining is suitable for processing large datasets from hospital records machine and locating family members amongst facts features. It takes only some researchers to investigate information from sanatorium records. The Main Concept is to determine medical sicknesses consistent with given signs and every day routine. The machine provides a person-pleasant interface for examinees and medical doctors.

1.1 Motivation:-

1. Previous medical examiner only used basic symptoms of particular diseases but in our application, examiner examines on the word count, laboratory results and diagnostic data.
2. 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.
3. When the user visits hospital physically, then user's personal record is saved and then that record is added to the examiner data set.

1.2 Aim & Objective:-

1. The Main concept is to determine medical diseases according to given symptoms and daily routine and when user search the hospital, the hospital which is nearest to their current location is given.
2. Determine medical diseases according to given symptoms and daily routine.
3. Prediction is done on the word count, laboratory results and diagnostic data.

II. RELATED WORK:-

1]. **"Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks" Author,-Srinivas K, Rani B K, Govrdhan A.** The healthcare environment is generally perceived as being 'information rich' yet 'knowledge poor'. 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. 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, we briefly examine the potential use of classification based data mining techniques such as rule based, decision tree, naive bayes and artificial neural network to massive volume of healthcare data. 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 Naive Bayes classifier (ODANB) and naive creedal classifier 2 (NCC2) are used. This is an extension of naive 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.

Disadvantage:

- For predicting heart attack significantly 15 attributes are listed
- Besides the 15 listed in medical literature we can also incorporate other data mining techniques, e.g., Time Series, Clustering and Association Rules.
- Categorical data is used
- Text mining is not used for unstructured data.

2]. **"Grand challenges in clinical decision support" Author- Sittig D, Wright A, Osheroff J, et al.**

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.

Disadvantage:-

- 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.

3]. **"Using Electronic Health Records for Surgical Quality Improvement in the Era of Big Data" Author-Anderson J E, Chang D C.** Many healthcare facilities enforce security on their electronic health records (EHRs) through a corrective mechanism: some staff nominally has 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, we propose a collaborative filtering inspired approach to predicting inappropriate accesses. 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.

4]. **"Data Mining Techniques into Telemedicine Systems" Author-Gheorghe M, Petre R** Providing care services through telemedicine has become an important part of the medical development process, due to the latest innovation in the in-formation 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.

5]. **"Query recommendation using query logs in search engines" Author-R. Baeza-Yates, C. Hurtado, and M. Mendoza.** In this paper we propose a method 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, we show with experiments over the query log of a search engine the effectiveness of the method.

6]. **"Data Mining Applications In Healthcare Sector: A Study " Author -M. Durairaj, V.** 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.

7]. **"Detecting Inappropriate Access to Electronic Health Records Using Collaborative Filtering" Author-Aditya Krishna Menon ,** 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, we propose a collaborative filtering inspired approach to predicting inappropriate accesses. 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.

8]. **"Text data mining of aged care accreditation reports to identify risk factors in medication management in Australian residential aged care homes" Author-Tao Jiang & Siyu Qian,** This 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).

9]. **"Evaluation of radiological features for breast tumour classification in clinical screening with machine learning methods" Author-Tim W. Nattkempera,, Bert Arnrich** The k-means clustering and self-organizing maps (SOM) are applied to analyze the signal structure in terms of visualization. We employ k-nearest neighbor classifiers (k-nn), support vector machines (SVM) and decision trees (DT) to classify features using a computer aided diagnosis (CAD) approach.

10]. **"Comparative Analysis of Logistic Regression and Artificial Neural Network for Computer-Aided Diagnosis of Breast Masses" Author-Song J H, Venkatesh S S, Conant E A,** Breast cancer is one of the most common cancers in women. Solography 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.

III. PROPOSED SYSTEM

The main concept is to decide medicinal ailments as per given side effects and every day routine given by the user, along with the ability of providing the closest doctor's facility of that present area. The system provides a user-friendly interface for examinees and doctors. Examinees can know the various side effects occurring in their body while the specialists can get a lot of examinees with potential hazard. 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.

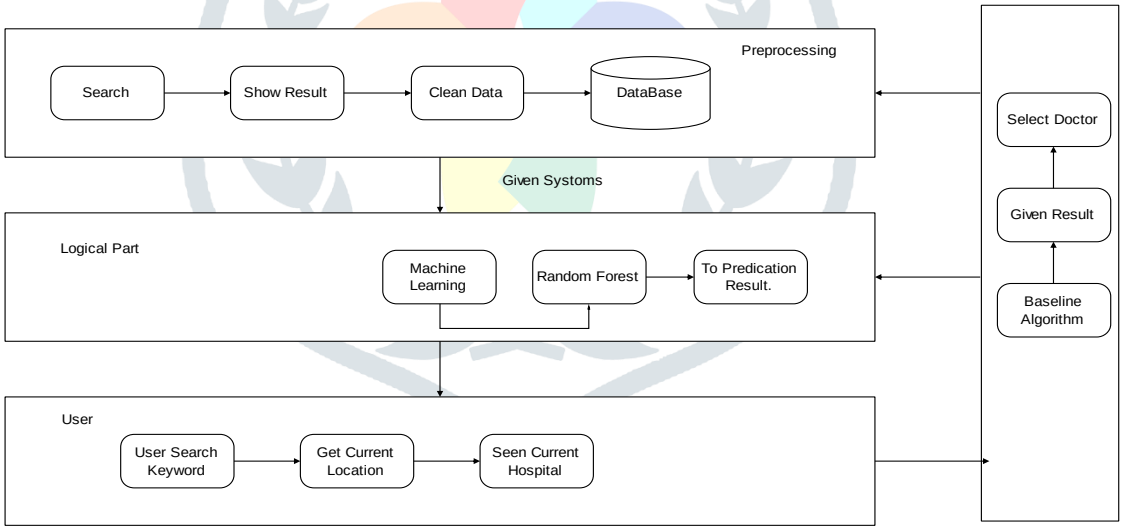

Fig 1: System Overview

Advantages are:

- Increases human-computer interactions
- Location of User is detected.
- Recommends the hospital and doctor to the patient according to the diseases predicted.
- Provides medicine for diseases which is predicted .
- Fast Prediction system
- Scalable, Low-cost
- Comparable quality to experts

IV. RESULT

In Our System, the dataset is divided into training set and testing set by the ratio 2 to 1. Application of the two algorithms mentioned above in risk-prediction is done. Based on the symptoms the system asks few questions and the user gives the answer to the questions generated by the system. The User also searches with keyword like specialization, doctor name and hospital name and gets the nearest hospital as a result according to current Location.

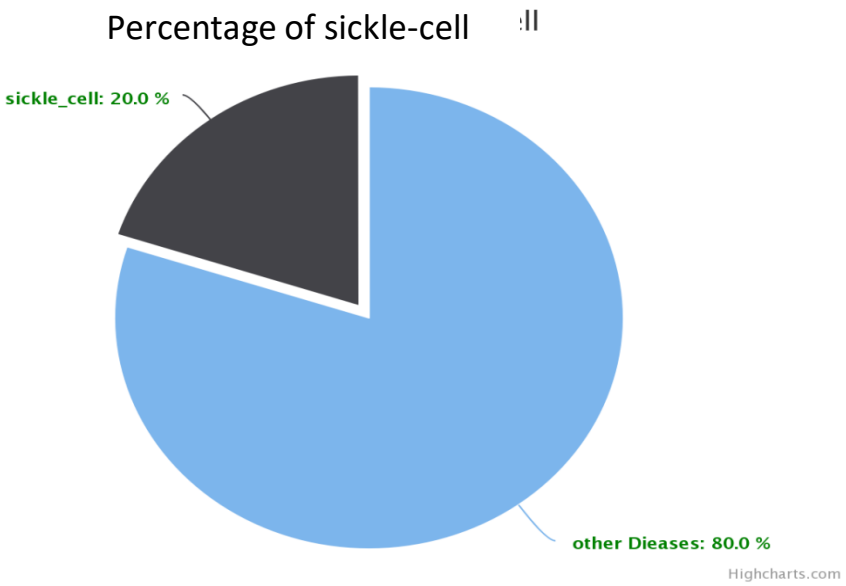

Fig 1. Result Analysis According to Precision and Recall

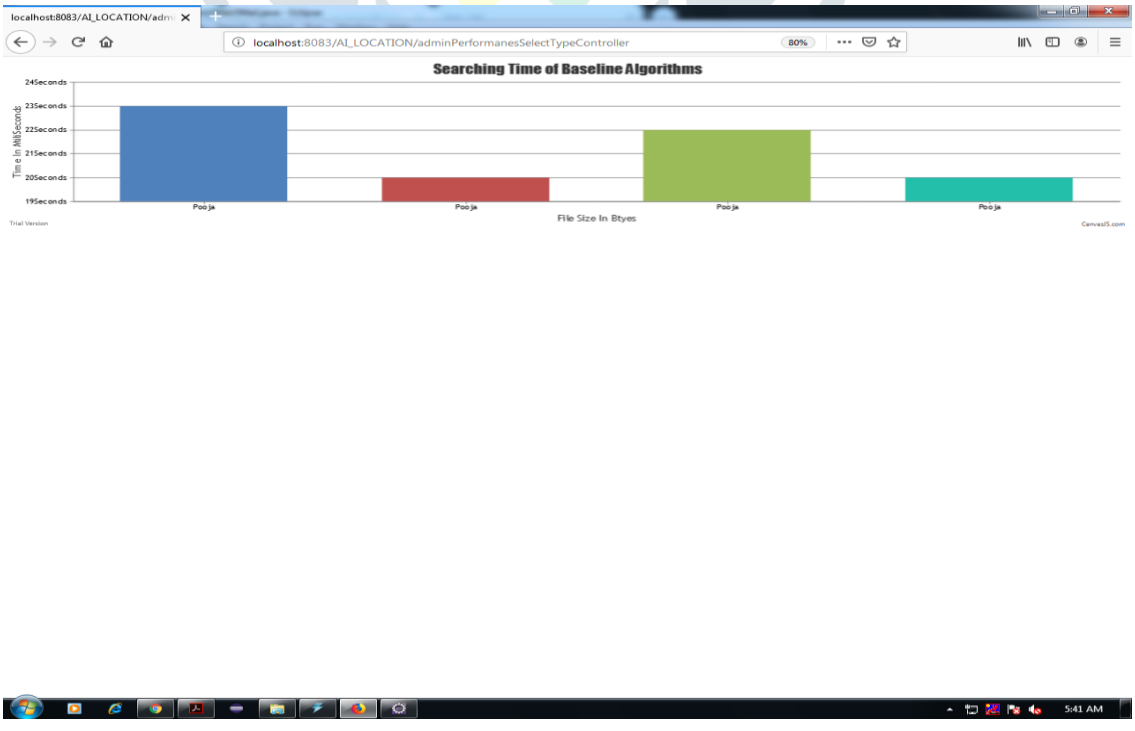

Fig 2. Result Analysis of Baseline Algorithms

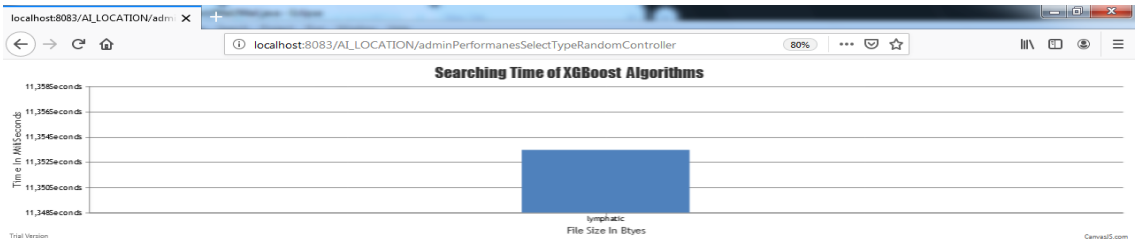

Fig 3. Result Analysis of XG Boost Algorithms

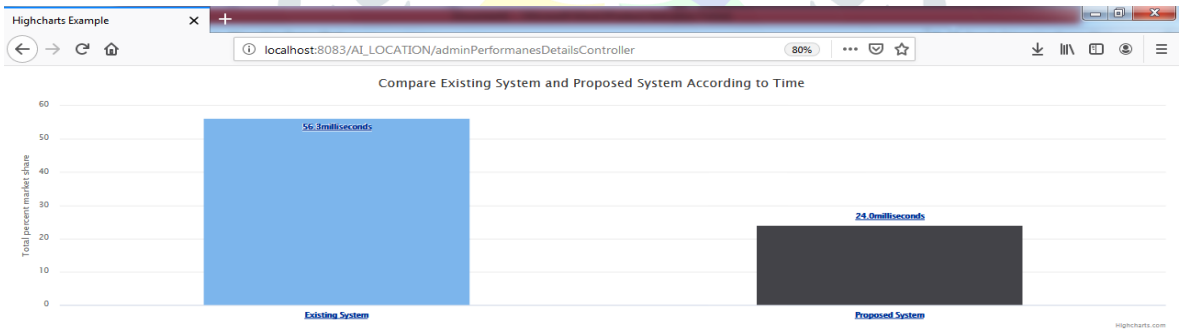

Fig 4 . Comparison of Existing System and Proposed System

V. 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 who will be in danger of physical deterioration next year. In our System user or patient search the hospital, and the results given are according to the nearest location of current location of user/patient. User / Patient gives symptoms and the system predicts the disease and provide the medicines. We also design a feedback mechanism 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.

REFERENCES

[1] Zhaoqian Lan, Guopeng Zhou, Yichun Duan , Wei Yan , "AI-assisted Prediction on Potential Health Risks with Regular Physical Examination Records", IEEE Transactions On Knowledge And Data Science, 2018.

[2] 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.

[3] Sittig D, Wright A, Osheroff J, et al. "Grand challenges in clinical decision support". Journal of Biomedical Informatics, 2008.

[4] Anderson J E, Chang D C. "Using Electronic Health Records for Surgical Quality Improvement in the Era of Big Data"[J]. Jama Surgery, 2015.

[5] Gheorghe M, Petre R. "Integrating Data Mining Techniques into Telemedicine Systems" Informatica Economica 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] Accreditation Reports to Identify Risk Factors in Medication Management in Australian Residential Aged Care Homes[J]. Studies in Health Technology & Informatics, 2017, 245:892.

[10] Nattkemper T W, Arnrich B, Lichte O, et al. Evaluation of radiological features for breast tumour classification in clinical screening with machine learning methods[J]. Artificial Intelligence in Medicine, 2005, 34(2):129- 139.

[11] 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.