

Disease Prediction System using Machine Learning

Jalindar Vidhate¹, Shubham Gore², Murad Shaikh³, Priti Rokade⁴ and ⁵Prof. Suryavanshi Y.G.

^{1,2,3,4} Student, ⁵ Assistant Professor,
ATC, SPPU, PUNE, Maharashtra, INDIA

Abstract:- People are more concerned with their own well-being in our public at large. Custom well-being management is increasing step by step. Due to the absence of experienced doctors and specialists, the health demand of the public cannot be met to most social insurance associations. The public also need a gradually correct result. In line with this, a growing number of applications in information mining are being developed to give more altered medical services to individuals. A decent response to the confusion of the lack of therapeutic resources and the increasing demands for rehabilitation. To discover the link between usual physical test records and the potential well-being risk presented by or open to the customer, we propose a framework that uses information mining technologies. The main concept is to determine health problems according to the side effects, the daily routine of the user, together with the ability to provide the nearest doctor's office in this area. The framework offers examiners and specialists an easy-to-understand interface. Examiners can determine the different side effects in your body while experts can obtain many examinations with a potential risk. A critical component could spare work and therefore improve the implementation of the framework. The specialist can determine the results of forecasts via an interface, which gathers latest information from specialists. All this information will be regularly enabled by an additional preparation procedure. Our framework could thus naturally improve the execution of the forecast.

Keyword:- Query Suggestion, Keyword Searching, Prediction of Disease, Health Prediction, Medical Services, Health Monitoring.

I. INTRODUCTION

Numerous social insurance associations (clinics, therapeutic focuses) in China are occupied in serving individuals with best possible healthcare services. These days, individuals give careful consideration on their physical conditions. They need higher quality and increasingly customized medicinal services benefit. Nonetheless, with the confinement of number of talented specialists and doctors, most medicinal services associations can't address the issue of open. The key issue is how to provide higher quality healthcare services to more individuals with limited mean power. The medicinal services condition is commonly seen as being 'data rich' yet information poor. Doctor's facility data frameworks commonly produce immense measure of information which appears as numbers, content. There is a great deal of hidden information in this data. Information Numerous social insurance associations (clinics, therapeutic focuses) in China are occupied in serving individuals with

best-exertion medicinal services benefit. These days, individuals give careful consideration on their physical conditions. They need higher quality and increasingly customized medicinal services benefit. However, due to the limited number of skilled specialist and doctors, most healthcare organizations cannot meet the needs of the public. Step by step instructions to give higher quality human services to more individuals with restricted labor turns into a key issue. The medicinal services condition is commonly seen as being 'data rich' yet information poor. Doctor's facility data frameworks commonly produce immense measure of information which appears as numbers, content. There is a great deal of concealed data in this information immaculate.

II. RELATED WORK

The environment of health [1] is seen as "informative" yet "without knowledge." The healthcare systems have a wealth of data available. There is, however, a lack of effective tools for analysing hideous relationships and data trends. In business and scientific areas, knowledge discovery and datamining have found numerous applications. Use of data mining techniques in the health system can discover valuable knowledge. In this study we are looking brilliantly at the potential use in massive medical volume of data-based clustering techniques including rule-based, decision-making trees, naive bays, and artificial neural networks. The health sector collects enormous amounts of health information, which, unfortunately, is not "mined" to discover hidden information. The ODANB and NCC2 algorithms are used for data pre-processing and decision-making purposes. The ODANB algorithm enhances the Naïve Bayes classification approach by including a single dependency between the features while the NCC2 algorithm improves the decision-making accuracy by combining multiple Naïve Bayes classifiers. Both algorithms have proven to be effective in improving data analysis and decision making in various applications. This is an extension to imprecise probabilities in Bayes that seeks to provide strong classifications, even if small or incomplete data sets are being dealt with. Often hidden patterns and relationships are untapped. It can forecast the likelihood of cardiac disease in patients using medical profiles such as age, sex, blood pressure and blood sugar. It allows for the establishment of significant knowledge, e.g., patterns, links between heart disease medical factors.

There is a pressing need [2] to develop, develop, implement, evaluate, and keep high-quality, efficient ways of supporting all kinds of clinical decision support for clinicians, patients, and consumers. Using a process of iterative consensus building, a ranking list of the top ten major challenges has been identified in support of clinical decisions. The list was developed for researchers, developers, funders, and policymakers to be educated or inspired. When patients and organisations begin to reap the benefits of these systems to the greatest extent possible, the challenges listed to overcome them include improving the interfaces between humans and computer systems; disseminating CDS-design, development, and implementation best practises; summarising patient information; prioritising and filtering user-specific recommendations; creating an architecture; and ensuring that the systems are secure. It is critical to identify solutions to these problems if support for clinical decision making is to reach its full potential and contribute to improvements in health care quality, security, and efficiency. Many medical facilities [3] enforce security on their electronic medical records using a corrective mechanism: some personnel have nearly unrestricted access to documents, but the process of inappropriate accesses is strictly ex-post facto, i.e., accesses that violate security and confidentiality policies in the facility are only detected after the fact has already occurred. As a result, this process is ineffective because every suspected access point must be examined by a security expert, which occurs after potential damage has occurred. This encourages the use of automated approaches that are based on historical data and machine learning. Successful applications of supervised learning models, such as SVMs and logistic regression, have been made in previous attempts to develop such a system. While these approaches have advantages over manual audits, they do not take into consideration the identities of registered users and patients. They cannot, as a result, take advantage of the fact that a patient whose record has previously been involved in an infringement is at an increased risk of being involved in another infringement in the future. In this document, we propose a collaborative filtering approach that encourages us to anticipate insufficient accesses. Our solution incorporates explicit as well as latent functionality for both staff and patients, who serve as custom fingerprints based on their previous access patterns and act as custom fingerprints. The proposed method, when applied to real EHR access data from two tertiary hospitals as well as a file-access data set from Amazon, not only provides significantly better performance than existing modes, but it also provides insight into what an improper access indicates.

The provision of telemedicine services has become a significant part of medical development because of recent advancements in information and computer technologies, according to authors [4]. With the ability to predict future trends and make decisions based on patterns and trends discovered, data mining, a dynamic and rapidly expanding domain, has, in fact, improved many aspects of human life, including healthcare and education. Data mining is made possible by the diversity of data and the variety of data mining technologies available. This includes applications for the health organisation, among other things. As part of the

medical development process, integrating data-mining technology into telemedicine systems would help to improve the efficiency and effectiveness of healthcare organisations, as well as the overall efficiency and effectiveness of the healthcare system. In this paper [5], we propose a method for generating a list of related requests based on a search engine query that has been submitted to the search engine. They can be based on previously issued queries and can be sent to the search engine so that it can adjust the search process or redirect it if necessary. Query clustering is used in this method to identify groups of semi-related queries, which is based on a query clustering process. During the clustering process, the contents of historical user preferences that have been recorded in the search engine query log are used. The procedure not only detects related questions, but it also ranks them according to a criterion of relevance that is determined by the procedure. In the end, we use experiments in the search engine query log to demonstrate the effectiveness of the method.

In this paper [6], our system was focused on comparing a variety of techniques, approaches, and tools, as well as how they affect the healthcare industry. The goal of a data mining application is to transform data into knowledge or information using computer algorithms and programmes. Primary objective is to develop an automated tool that will identify and disseminate relevant information throughout the healthcare system. The purpose of this article is to produce a detailed survey report on various types of data mining applications in the health sector, as well as to reduce the complexity of the study of transactions in medical data. Additionally, a comparative study of various applications, techniques, and methodologies used in the mining of information from databases in the healthcare industry is presented. To conclude, the current data mining techniques are discussed in detail, with an emphasis on data mining algorithms and their application in the healthcare industry.

In most health-care organizations [7], the security of their electronic health records is implemented through an adjustment mechanism: while certain employees have nominally almost unrestricted access to the record, inappropriate access, that is, accesses that are not in accordance with the security and privacy policies, is subject to a strict post-facto auditing process. As a result, this process is ineffective because every suspected access point must be examined by a security expert, which occurs after potential damage has occurred. This encourages the use of automated approaches that are based on historical data and machine learning. Successful applications of supervised learning models, such as SVMs and logistic regression, have been made in previous attempts to develop such a system. While these approaches have advantages over manual audits, they do not take into consideration the identities of registered users and patients. They cannot, as a result, take advantage of the fact that a patient whose record has previously been involved in an infringement is at an increased risk of being involved in another infringement in the future. In this document, we propose a collaborative filtering approach that encourages us to anticipate insufficient accesses. Our solution incorporates explicit as

well as latent functionality for both staff and patients, who serve as custom fingerprints based on their previous access patterns and function as custom fingerprints. After being tested on real data from two tertiary hospitals as well as an Amazon file-access dataset, the proposed method not only demonstrates significantly improved performance when compared to existing procedures, but it also provides insight into what constitutes improper access.

This study [8] investigated the risk factors associated with the management of medicinal products in Australia's residential aged care (RAC) homes. Only 18 out of 3,607 RAC homes failed to meet the standard of aged care accreditation in medicines management during the period from 7 March 2011 to 25 March 2015. Text data mining techniques were employed to determine the root causes of failure. In the absence of medication management, this resulted in the identification of 21 RAC risk indicators. These indicators have also been subdivided into a total of ten categories. They include the overall management of drugs, drug assessment, ordering, dispensing, storage, stocking, and disposal, management, incident reporting, surveillance, personnel, and the resident's satisfaction with the facility. The following are the top three risk factors: "ineffective monitoring process" (18 households), "non-compliance with professional standards and regulations," and "non-compliance with professional standards and regulations" (10 homes).

When it comes to visualising the signal structure, k-means [9] and self-organizing maps (sOM) are used to analyse and visualise the structure. In a CAD approach, we classify features using k-classified neighbourhoods (k-nns), supporting vector machines (SVMs), and decision-making bodies (DTs), as well as supporting vector machines and decision-making bodies (DTs). Female breast cancer [10] is one of the most common cancers to occur in women of reproductive age. Solography is now routinely used in conjunction with other imaging modalities to obtain images of the breasts. Despite significant efforts to improve imaging techniques, such as solography, the final determination of whether a solid breast lesion is malignant or benign is still made by biopsy, despite the advances in technology.

III. OPEN ISSUES

Due to non-accessibility to public health care and low quality of health care services, a majority of people in India turn to the local private health sector as their first choice of care. If we look at the health landscape of India 92 percent of health care visits are to private providers of which 70 percent is urban population. However, private healthcare is expensive, often unregulated and variable in quality.

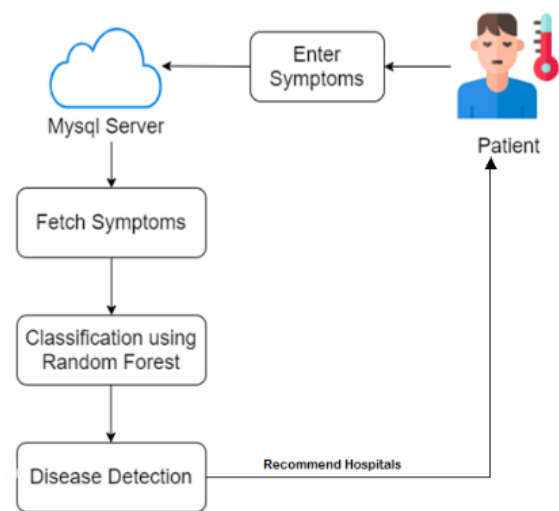


Fig. 1: System Architecture

Control the spread of diseases and reduce the growing rates of mortality due to lack of adequate health facilities, special attention needs to be given to the health care in rural areas. The key challenges in the healthcare sector are low quality of care, poor accountability, lack of awareness, and limited access to facilities.

IV. PROPOSED SYSTEM

Main Concept: When a user searches for a hospital, they are directed to the hospital that is closest to their current location based on the symptoms and daily routine they provide. The system offers an easy-to-use interface for examiners and physicians. Tests may know their symptoms that have added in the body, and doctors can get several examiners at risk. A feedback mechanism can automatically save workplace and improve system performance. An interface can fix the prediction result, which collects the input of physicians as new training information. Using these data, a further training process is initiated daily. Our system could therefore automatically improve the performance of the forecast model.

➤ *Advantages are:*

- Increases the number of interactions between humans and computers.
- The User's current location has been found.
- Referred the patient to the appropriate hospital and doctor based on the diseases predicted.
- Provided medicine for diseases that were predicted to happen.
- A fast prediction system that is scalable and low-cost, and that has comparable quality to experts.

V. RESULT

In our system, we divide the data set into two parts: the training set and the testing set, divided by a factor of two to one. Then, in a risk-prediction task involving three symptoms, Our System employs the two algorithms mentioned above to make predictions. When the system meets these three symptoms, it will ask the user a question about the symptoms. When the system asks a question, the

user will receive an answer. When a user searches for a hospital in our system using keywords such as specialisation, doctor name, and hospital name, the user is presented with a list of hospitals that are closest to their current location.



Fig. 1: Login page

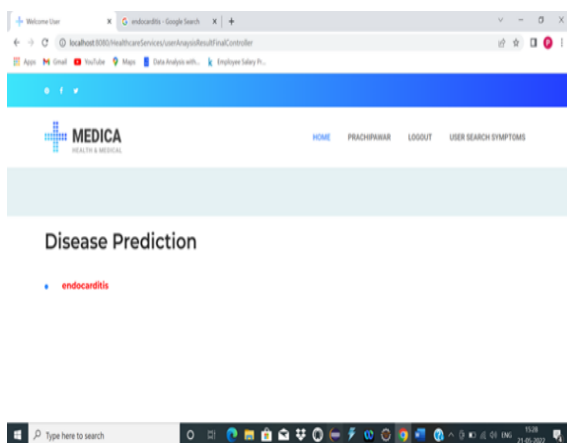


Fig. 2: Disease Prediction Result

VI. CONCLUSION

This project makes use of data mining techniques to discover the relationship between regular physical test records and the potential health risks presented by the user or by the public to improve public health. The physical condition of an examination is likely to deteriorate in the coming year because of the application of various machine learning algorithms. When a system user or a patient search for a hospital, the results are routed to the user's or patient's location that is closest to the search engine. It is possible to express user/patient symptoms, and the system predicts diseases and supplies medicines in response. In addition, we are developing a feedback system for physicians to use in determining classification results or entering new training data, and the system will restart the training process every day to improve performance.

REFERENCES

- [1.] 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.
- [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.] Gheorghe M, Petre R. "Integrating Data Mining Techniques into Telemedicine Systems" *Informatica Economica Journal*, 2014.
- [4.] 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.
- [5.] Koh H C, Tan G. Data mining applications in healthcare. [J]. *Journal of Healthcare Information Management Jhim*, 2005, 19(2):64-72.
- [6.] 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.
- [7.] Accreditation Reports to Identify Risk Factors in Medication Management in Australian Residential Aged Care Homes[J]. *Studies in Health Technology & Informatics*, 2017, 245:892.
- [8.] 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.
- [9.] 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):