



IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2320-6710



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 7, July 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.569



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

AI Assisted Potential Health Risk Prediction Using Machine Learning Computing

Shailaja DS¹, Dr. Kamalakshi Naganna²

Department of Computer Science & Engineering, Sapthagiri College of Engineering, Bangalore, Karnataka, India¹

Department of Computer Science & Engineering, Sapthagiri College of Engineering, Bangalore, Karnataka, India²

ABSTRACT: With the development of society and economy, people pay more attention to their own health. The demand of more personalized health service is gradually rising. However, due to the lack of experienced doctors and physicians, most healthcare organizations cannot meet the medical demand of public. With the widespread use of hospital information system, there is huge amount of generated data which can be used to improve healthcare service. Thus, more and more data mining applications are developed to provide people more customized healthcare service. In this paper, we propose 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. It can predict examinees' risk of physical status next year based on the physical examination records this year. The system provides a user-friendly interface for examinees and doctors. Examinees can know their potential health risks while doctors can get a set of examinees with potential risk. It is a good solution for the mismatch of insufficient medical resources and rising medical demands.

I. INTRODUCTION

Many healthcare organizations (hospitals, medical centers) in China are busy in serving people with best-effort healthcare service. Now a days, 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, charts and images. 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. Data mining is suitable for processing large datasets from hospital information system and finding relations among data features. It takes only a few researchers to analyze data from hospital information systems, and provide huge medical knowledge which can be used to support clinical decision making. Also, we could use data mining to provide a self-service healthcare system, which can serve lots of people at the same time. The self-service healthcare system is of great significance to solve the problem of imbalance between limited medical resources and demands. Healthcare data mining has been most widely used for diagnosis, prognosis or treatment planning.

Organization of the Paper

This paper is organised as follows. Related works is discussed in section 1, section 2 includes Performance parameters related to data survey prediction.

II. RELATED WORKS

Literature Survey:

1. Title: Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks

Author: Srinivas K, Rani B K, Govrdhan A.

Abstract: 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, Naïve 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 Naïve Bayes classifier (ODANB) and naive credal 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.

Disadvantages:

In this work for predicting heart attack significantly 15 attributes are listed but we won't get the accuracy of prediction results using only 15 attributes.

4. Title: Predicting patient acuity from electronic patient records

Author: Kontio E, Airola A, Pahikkala T,

Abstract: The ability to predict acuity (patients' care needs), would provide a powerful tool for health care managers to allocate resources. Such estimations and predictions for the care process can be produced from the vast amounts of healthcare data using information technology and computational intelligence techniques. Tactical decision-making and resource allocation may also be supported with different mathematical optimization models.

METHODS: This study was conducted with a data set comprising electronic nursing narratives and the associated Oulu Patient Classification (OPCq) acuity. A mathematical model for the automated assignment of patient acuity scores was utilized and evaluated with the pre-processed data from 23,528 electronic patient records. The methods to predict patient's acuity were based on linguistic pre-processing, vector-space text modeling, and regularized least-squares regression.

Disadvantages:

Less accuracy in prediction

Less efficiency

Problem Statement:

The demand of more personalized health service is gradually rising. However, due to the lack of experienced doctors and physicians, most healthcare organizations cannot meet the medical demand of public. With the widespread use of hospital information system, there is huge amount of generated data which can be used to improve healthcare service.

Challenges:

There are three challenges in our system:

1. The data is in various forms. We should preprocess data precisely for the data mining, including data cleaning, normalization and so on. Diagnostic data is in the form of sentences of natural language. Thus, proper structured method should be applied before further processing. Natural language processing (NLP) techniques are widely used in AI assistants, but rarely used in medical terms. It is important to find a proper way to process diagnostic data using NLP methods.

2. The data is in high dimension Physical examination data usually includes examinees basic information, laboratory results and some diagnostic data. It means original data is high dimensional and we should apply appropriate dimensionality reduction method or otherwise we may encounter high computational complexity. Also, high dimension input may also bring low accuracy problem when dataset is not huge enough.

3. 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. This mechanism is of vital importance to reduce the number of maintenance personnel and make the system operating automatically.

III.METHODOLOGIES

Naive Bayes: Naive Bayes has been studied extensively since the 1950s. It was introduced under a different name into the text retrieval community in the early 1960s, and remains a popular (baseline) method for text categorization, the

problem of judging documents as belonging to one category or the other (such as spam or legitimate, sports or politics, etc.) with word frequencies as the features. With appropriate pre-processing, it is competitive in this domain with more advanced methods including support vector machines. It also finds application in automatic medical diagnosis.

KNN Classifier:

In pattern recognition, the k-nearest neighbors algorithm (k-NN) is a non-parametric method used for classification and regression.[1] In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression:

In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of that single nearest neighbor.

System Architecture:

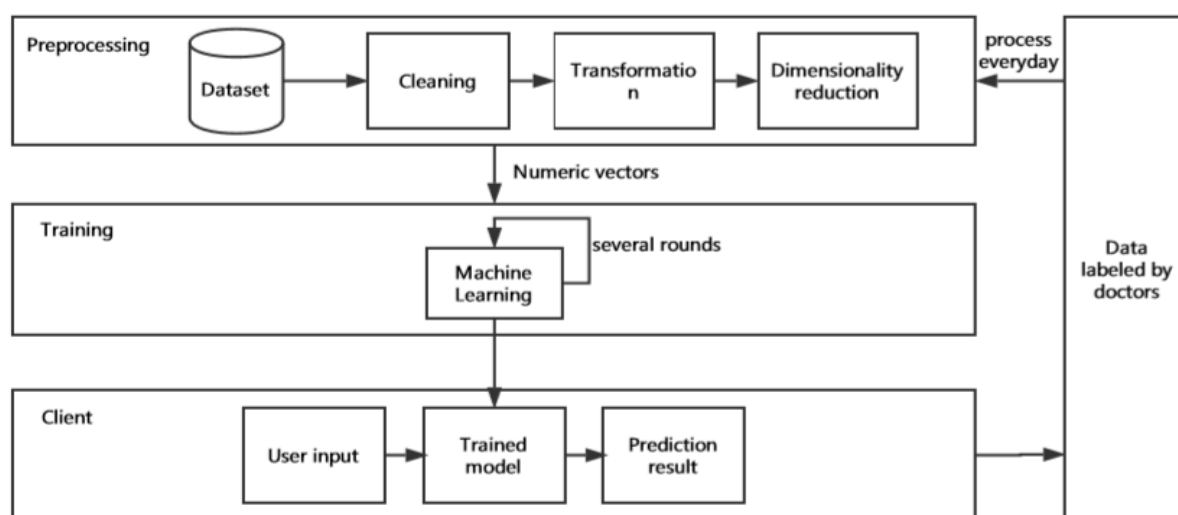


Figure1:Architecture Flow

FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and **business** proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- ◆ ECONOMICAL FEASIBILITY
- ◆ TECHNICAL FEASIBILITY
- ◆ SOCIAL FEASIBILITY

ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

IV.RESULTS AND PERFORMANCE ANALYSIS

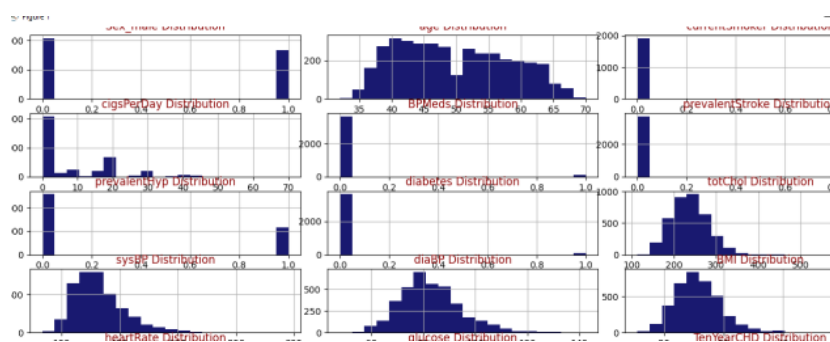


Figure2: Output analysis

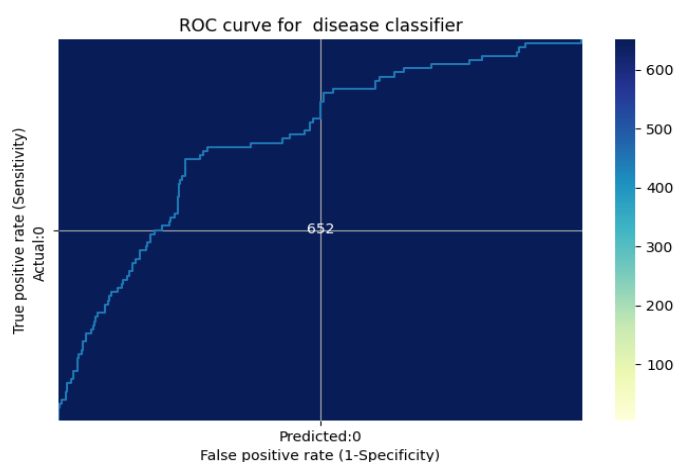


Figure3: Histogram curve

Our system is a self-service system, which can provide personalized healthcare service to examinees with few maintenance personnel. It is a good solution for the mismatch of insufficient experienced doctors and rising medical demands. It will gather more training data and improve precision automatically, which releases a huge amount of manpower and contains great potential for application.

We have developed with only the text dataset in future we can be able to use deep learning algorithms to improve the accuracy and predict from the image data.

V.CONCLUSION

- We propose new machine learning based multimodal disease risk prediction (CNN-MDRP) algorithm using structured and unstructured data from hospital.
- To the best of our knowledge, none of the existing work focused on both data types in the area of medical big data analytics. Compared to several typical prediction algorithms, the prediction accuracy of our proposed algorithm reaches 94.8% with a convergence speed which is faster than that of the machine learning-based unimodal disease risk prediction (CNNUDRP) algorithm.



REFERENCES

- [1] Srinivas K, Rani B K, Govrdhan A. Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks[J]. International Journal on Computer Science & Engineering, 2010, 2(2):250-255.
- [2] Delen, D., & Demirkan, H. Data, information and analytics as services. Decision Support Systems, 2013: 55, 359363.
- [3] Malik M M, Abdallah S, AlaRaj M. Data mining and predictive analytics applications for the delivery of healthcare services: a systematic literature review[J]. Annals of Operations Research, 2016:1-26.
- [4] Sittig D, Wright A, Osheroff J, et al. Grand challenges in clinical decision support.[J]. Journal of Biomedical Informatics, 2008, 41(2):387.
- [5] Anderson J E, Chang D C. Using Electronic Health Records for Surgical Quality Improvement in the Era of Big Data[J]. Jama Surgery, 2015, 150(1):1-6.
- [6] Gheorghe M, Petre R. Integrating Data Mining Techniques into Telemedicine Systems[J]. Informatica Economica Journal, 2014, 18(1):120-130.
- [7] Kontio E, Airola A, Pahikkala T, et al. Predicting patient acuity from electronic patient records[J]. Journal of Biomedical Informatics, 2014, 51:35-40.
- [8] Amarasingham R, Patzer R E, Huesch M, et al. Implementing electronic health care predictive analytics: considerations and challenges.[J]. Health Aff, 2014, 33(7):1148-1154.
- [9] Koh H C, Tan G. Data mining applications in healthcare.[J]. Journal of Healthcare Information Management Jhim, 2005, 19(2):64-72.
- [10] 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.
- [11] Yoo I, Alafaireet P, Marinov M, et al. Data mining in healthcare and biomedicine: a survey of the literature.[J]. Journal of Medical Systems, 2012, 36(4):2431-2448.
- [12] Miller R A. Medical Diagnostic Decision Support SystemsPast, Present, And Future A Threaded Bibliography and Brief Commentary[J]. Journal of the American Medical Informatics Association Jamia, 1994, 1(1):8.



**Impact Factor:
7.569**



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details