AI-Assisted Prediction on Potential Health Risks

Ms. Anusha
Professor, Department of Computer Science Engineering Jain Deemed to be University
Bangalore, India
G Tharun Reddy
UG Scholar, Department of Computer Science Engineering Jain Deemed to be University
Bangalore, India
G Naveen
UG Scholar, Department of Computer Science Engineering Jain Deemed to be University
Bangalore, India
P Pavanchand
UG Scholar, Department of Computer Science Engineering Jain Deemed to be University
Bangalore, India
M Ravi Kiran
UG Scholar, Department of Computer Science Engineering Jain Deemed to be University
Bangalore, India

Abstract - With the event of society and economy, people pay more attention to their own health. The demand for personalized health services is gradually rising. However, thanks to the shortage of experienced doctors and physicians, most healthcare organizations cannot meet the medical demand of the public. With the widespread use of hospital data systems, there's a huge amount of generated data which may be wont to improve healthcare service. Thus, more and more data processing applications are developed to supply people with more customized healthcare services. In this paper, we are proposing an AI-assisted prediction system, which leverages data processing methods to reveal the connection between the regular physical examination records and therefore the potential health risk. It can predict the examinee risk of physical status next year and supports the physical examination records this year. The system provides a user-friendly interface for both examinees and doctors. The examinee can know their potential health risks while doctors can get a group of examinees with potential risks. It is an honest solution for the mismatch of insufficient medical resources and rising medical demands.

1. INTRODUCTION

Many healthcare organizations (hospitals, medical centers) around the world are busy in serving people with the best healthcare services. Nowadays, people pay more attention to their health. They want higher quality and more personalized healthcare services. However, with the limitation of the number of skilled doctors and physicians, most healthcare organizations cannot meet the needs of the public. How to provide higher quality healthcare to more people with limited manpower becomes a key issue. The healthcare sector is generally perceived as being 'information-rich' yet 'knowledge-poor'. Hospital systems typically generate huge amounts 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 techniques 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 the 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.

2. RELATED WORKS

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

Author: Srinivas K, Rani B K, Govardhan 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 the business and scientific domain. Valuable knowledge can be discovered from the application of data mining techniques in the 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 a 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.

2. Title: Using Electronic Health Records forSurgical Quality Improvement in the Era of Big Data

Author: Anderson J E, Chang DC.

Abstract: To determine whether a number of machine-collected data elements could perform as well as a traditional full-risk adjustment model that includes other physician-assessed and physician-recorded data elements. Rigorous risk-adjusted surgical quality assessment can be performed solely with objective variables. By leveraging data already routinely collected for patient care, this approach allows for wider adoption of quality assessment systems in health care. Identifying data elements that can be automatically collected can make future improvements to surgical outcomes and quality analyses.

Disadvantages:

In this work they are mainly focusing the data analyses for surgeons. it is not useful for all the people.

3. Title: Integrating Data Mining Techniques into Telemedicine Systems

Author: Gheorghe M, Petre R.

Abstract: The medical system is facing a wide range of challenges nowadays due to changes that are taking place in the global healthcare systems. These challenges are represented mostly by economic constraints (spiraling costs, financial issues), but also, by the increased emphasis on accountability and transparency, changes that were made in the education field, the fact that biomedical research keeps growing in what concerns the complexities of the specific studies, etc. Also the new partnerships that were made in medical care systems and the great advances in the IT industry suggest that a predominant paradigm shift is occurring. This needs a focus on interaction, collaboration, and increased sharing of information and knowledge, all of these may in turn be leading healthcare organizations to embrace the techniques of data mining in order to create and sustain optimal healthcare outcomes. Data mining is a domain of great importance nowadays as it provides

advanced data analysis techniques for extracting the knowledge from the huge volumes of data collected and stored by every system on a daily basis. In healthcare organizations data mining can provide valuable information for patient's diagnosis and treatment planning, customer relationship management, organization resources management, or fraud detection. In this article we focus on describing the importance of data mining techniques and systems for healthcare organizations with a focus on developing and implementing telemedicine solutions in order to improve the healthcare services provided to the patients. We provide the architecture for integrating data mining techniques into telemedicine systems and also offer an overview of understanding and improving the implemented solution by using Business Process Management methods.

Disadvantages:

In this system they are improving accuracy and efficiency in data analytics but there is no automated prediction system is not involved

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 healthcare 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 by 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 a 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

5. Title: Implementing electronic health care predictive analytics: considerations and challenges

Author: Amarasingham R, Patzer R E, Huesch M

Abstract: The use of predictive modeling for real-time clinical decision making is increasingly recognized as a way to achieve the Triple Aim of improving outcomes, enhancing patients' experiences, and reducing health care costs. The development and validation of predictive models for clinical practice is only the initial step in the journey toward mainstream implementation of real-time point-of-care predictions. Integrating electronic health care predictive analytics (e-HPA) into the clinical workflow, testing e-HPA in a patient population, and subsequently disseminating e-HPA across US health care systems on a broad scale require thoughtful planning. Input is needed from policymakers, health care executives, researchers, and practitioners as the field evolves. This article describes some of the considerations and challenges of implementing e-HPA, including the need to ensure patients' privacy, establish a health system monitoring team to oversee implementation, incorporate predictive analytics into medical education, and make sure that electronic systems do not replace or crowd out decision making by physicians and patients.

Disadvantages:

In this article we have discussed considerations and challenges in implementing e-HPA in the health care system. This discussion is not comprehensive but instead is intended to highlight specific challenges to introduce the need for a thoughtful set of best practices to guide the rise of e-HPA.

3. PROPOSED SYSTEM

We collect our dataset from UCI Machine Learning Repository.We leverage the data cleaning approach, dimensionality reduction methods, and several machine learning algorithms to process the data and set up a prediction system. Firstly, we apply the word vector model to medical history and diagnosis data, converting them into 0-1 features. Secondly, we design a dimensionality reduction method to address the high dimensionality problem and reduce the computational complexity. Then we use machine learning methods to discover the relationship between physical examination records and potential health risks. With these techniques, we buildup our prediction system. The system provides a user-friendly interface for examinees checking their health risks after a physical examination, and for doctors getting examinees set for intervention. Besides, the system provides a feedback mechanism for doctors to fix the prediction inaccuracy. These latest labeled data will trigger the training step every day, which automatically improves the performance of the system.

4. IMPLEMENTATION

Data Cleaning and Transformation: We collect original dataset from the hospital information system and assign each examinee a unique ID across all the dataset. Only data from examinees who participate in physical examination both years is useful for our task. After filtering out the ones who only take a single year examination, there are 2,637 examinees remaining. As mentioned before, we will de□ne a proper mapping O : E →{ 0,1}, which is a critical discriminant of at-risk people classi□cation. After conducting a careful survey in the medical center and discussing with our cooperative hospital, we de□ne a mapping O according to actual work needs.

Dimensionality Reduction: A high-dimensional feature space can bring many problems to machine learning. Firstly, it significantly increases operating time of prediction.Secondly, as the dimension of features increases, the likelihood of over-fitting increases. Thirdly, the more the independent variable dimension, the more sparse the data is distributed across the input space, and the harder it is to obtain a representative sample of the entire input space. So it is important to apply dimensionality reduction methods before the data mining process. We use logistic regression to predict in advance and use the L1 regularization term to evaluate and select features. L1 regularization can sparse input matrix, and it helps perform feature selection in sparse feature spaces. We choose 0.1 as threshold of L1, and there are 27 to 30 features remaining after dimensionality reduction varying with different tasks.

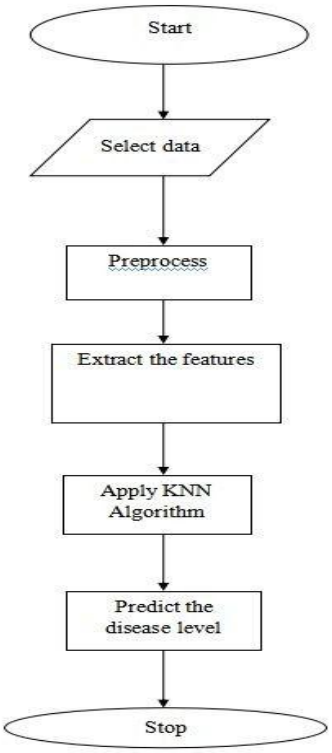
Classi□cation: Now we have prepared the dataset and are ready for applying machine learning methods. This is the core of the whole system, which gives the classi□cation result from the feature vector of the examinee. We select several different algorithms and compare their performance in our experiments.

Naive Bayes: Naive Bayes algorithm is a machine learning technique based on Bayes theorem. It was presented under a special name into the text retrieval community in the early 1960s and remains a well-liked method for text categorization, the matter of judging documents as belonging to at least one category or the opposite with word frequencies because of the features. With appropriate pre-processing, its competitors in this domain with more modern methods including support vector machines. It also finds application in automatic medical diagnosis.

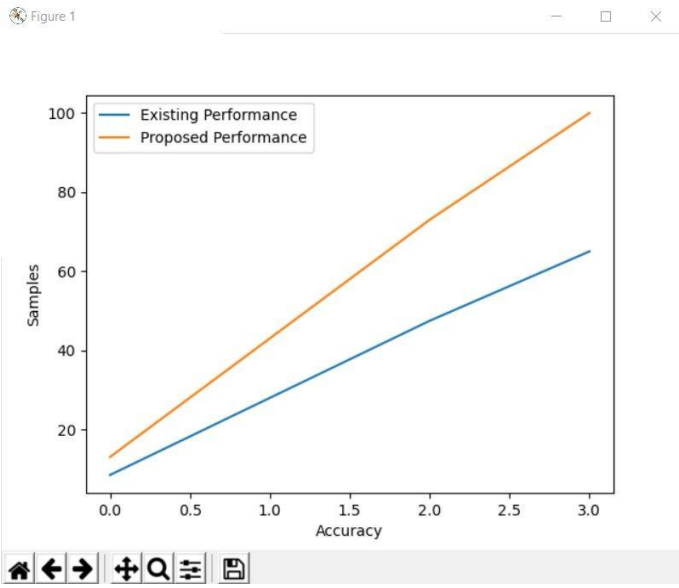
KNN Classifier: The KNN (k-nearest neighbor) is a machine learning algorithm used for classification and regression. The input consists of the k closest training examples within the feature space. The output depends on whether KNN is employed for classification or regression.

In KNN classification, the output may be a class membership. An object is assessed by a majority vote of its neighbors, with the thing being assigned to the category commonest among its k nearest neighbors (k may be a positive integer, typically small). If k = 1, then the thing is assigned to the category of that single nearest neighbor.

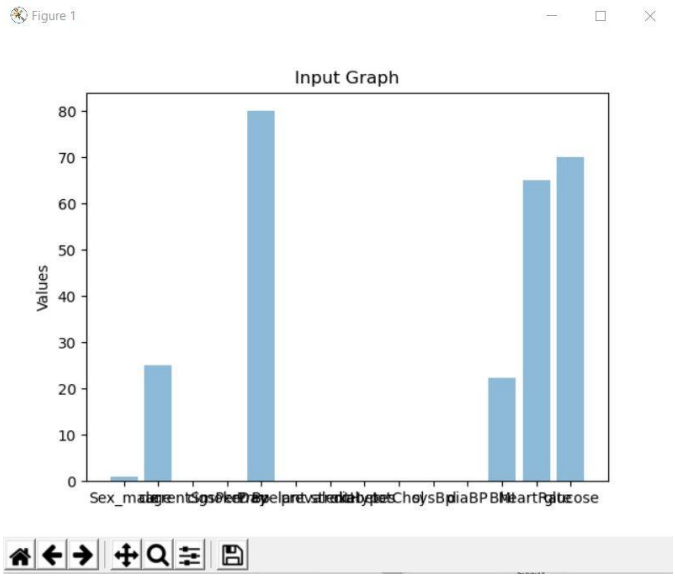
Flow chart:



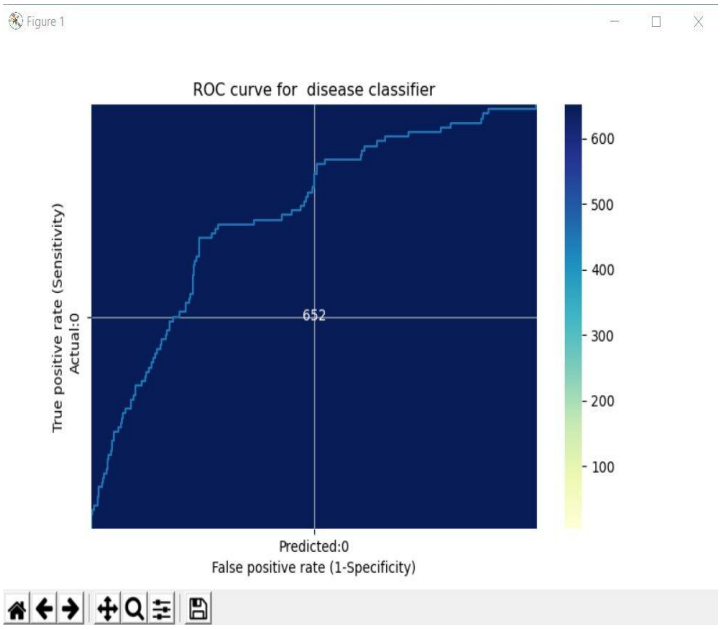
5. RESULTS



Accuracy



Input graph



ROC Curve

6. CONCLUSIONS

A desktop application using python, which can provide personalized healthcare services to examinees. 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.

7. REFERENCES

[1] Srinivas K, Rani B K, Govardhan 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.

[13] West D, West V. Model selection for a medical diagnostic decision support system: a breast cancer detection case.[J]. Arti□cial Intelligence in Medicine, 2000, 20(3):183-204.

[14] Song J H, Venkatesh S S, Conant E A, et al. Comparative analysis of logistic regression and arti□cial neural network for computer-aided diagnosis of breast masses.[J]. Academic Radiology, 2005, 12(4):487-95.

[15] Nattkemper T W, Arnrich B, Lichte O, et al. Evaluation of radiological features for breasttum our classi□cationin clinical screening with machine learning methods[J]. Arti□cial Intelligence in Medicine, 2005, 34(2):129139.