RESEARCHERID
THOMSON REUTERS

[NCRAES-2019] ISSN 2348 – 8034
Impact Factor- 5.070

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

AI-ASSISTED PREDICTION ON POTENTIAL HEALTH RISKS WITH REGULAR PHYSICAL EXAMINATION RECORDS

Sonali Patil[*1], Deepali Patil[2] & Supriya Yadav[3]
[*1,2,3]Assistant Professor, Computer Engineering, NMIET, Pune, India

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.

Keywords: *Navii Bayes, Machine Learning.*

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 [1]. 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 [2]. 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 [3]. In this paper, we aim to build a self-service prediction system to identify at-risk people using their regular physical examination records. It can serve both examinees and doctors with prediction of potential risk. Typical physical examination records consist of examinees' basic information, laboratory results, medical history and some diagnostic data. These records are usually in different forms and in high dimension, which brings difficulties in data processing. Additionally, our system can automatically collect new data while running, with which we re-train the model and improve the system performance. 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 and is difficult to be processed. 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.

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.

We collect our dataset from the Health Management Center of the Peking University First Hospital (PUFH). We leverage data cleaning approach, dimensionality reduction methods and several machine learning algorithms to process the data and setup a prediction system. Firstly, we apply 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 relationship between physical examination records and potential health risks. With these techniques, we build up our prediction system. The system provides a user-friendly interface for examinees checking their health risks after 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 everyday, which automatically improves the performance of the system.

The rest of this paper is organized as follows. In Section II, we briefly introduce the related works about data mining and medical diagnostic decision support systems. Then in Section III, we introduce our dataset and formulate the health risk prediction problem. Next we present system architecture and data processing details in Section IV. We evaluate the performance of proposed system in Section V. Finally, we conclude our work in Section VI.

II. RELATED WORK

Data mining is an efficient way of analyzing big data. As continuous and voluminous growth of data is being generated from various hospital information systems and on-body de-vices, many works using data mining methods have been done in the medical field. One of the most important applications is the clinic decision support (CDS) systems, which provides clinicians, staff, patients, or other individuals with knowledge and person-specific information, intelligently filtered or presented at appropriate times, to enhance health and health care [4]. Srinivas et al. make prediction of heart attacks [1], and Anderson and Chang identify and classify at-risk people in surgery using electronic health records [5]. Gheorghe and Petre integrate data mining techniques into telemedicine systems [6], and Kontio et al. determine patient acuity from electronic patient records [7]. Increasing healthcare costs also draws much attention, many applications focus on cutting costs or planning budget on healthcare. Amarasingham et al. bring out the use of predictive modeling for real-time clinical decision as a way of enhancing patients experiences as well as reducing healthcare costs [8]. Other data mining and predictive analytics applications in healthcare include customer relationship man-agement [9], detection of fraud [10], [11] and evaluation of treatment effectiveness [3], [9].

The medical diagnostic decision support (MDDS) system is one of the most important programs of CDS. MDDSs have become an established component of medical technology and their use will continue to grow, fueled by electronic medical records and automatic data capture [12]. The purpose of an MDDS is to augment, not replace, the natural capabilities of human diagnosticians in the complex process of medical diagnosis [13]. The development of machine learning methods has greatly advanced MDDSs improvement. Methods varying from decision tree to neural network models have been applied to MDDSs. These machine learning methods are used to diagnose on breast sonograms [14], to classify breast tumour [15], [16], to detect gastrointestinal adenomas [17], [18], to identify

risk factors in medication management in Australian residential aged care (RAC) homes [19] and to predict drug efficacy for diabetes treatment [20]. Among these applications, logistic regression and neural network are widely used, reflect-ing strong ability of machine learning methods in classification and prediction tasks. Those applications have proved that ma-chine learning algorithms are practical in medical diagnostic decision support area.

There are also many healthcare management applications based on data mining. Data mining applications can be developed to better identify and track chronic disease states and high-risk patients, design appropriate interventions, and reduce the number of hospital admissions and claims [21]. An application predicting the risk of in-hospital mortality in cancer patients with nonterminal disease [22] is developed using neural network. Another system leverages neural network to predict the disposition in children presenting to the emergency room with bronchiolitis [23]. Also, logistic regression models are used to compare hospital profiles based on risk-adjusted death with 30 days of non-cardiac surgery [24]. Healthcare management applications reduce the imbalance of medical resources and public healthcare demand. With the development of data mining, more and more healthcare management applications are put into service. Our proposed system is also one of healthcare management applications, which could be very practical to satisfy the rising demand of healthcare.

In this paper, we build a self-service prediction system to identify people with potential health risk using their regular physical examination records. We apply various supervised machine learning methods, including decision tree, XGBoost and random forests to predict potential health risks of exam-inees using their physical examination records. Experiment shows that overall performance of XGBoost is the best. Additionally, our system can automatically collect new data while running, then re-train the model and improve system performance using these data. This is a unique mechanism and we will describe our system in detail in following sections.

III. DATASET AND PROBLEM FORMULATION

Now we introduce our dataset and formulate the prediction problem in mathematical language.

A. Dataset

We establish cooperation with the Health Management Center of Peking University First Hospital and collect for physical examination data of several months. Peking University First Hospital is a large comprehensive tertiary hospital integrating medical services with teaching, research and preventive medicine. It is a Tertiary Hospital with 36 clinical departments, 16 technological departments, 6 institutes, totally 64 wards and 1574 available beds. Every year, over 3.6 million outpatients and emergency cases are treated, with over 85,000 patients admitted and about 44,000 operations completed. The physical examination center is attached to the Health Management Center, one of the clinical departments of the Peking University First Hospital. As a professional organization, it provides tens of thousands people physical examination service each year.

Our dataset covers examinees' basic information (age, gender, height, etc.), laboratory tests result, medical history and brief diagnosis with sensitive personal data excluded. Examinees are select from the ones attending physical examination in March and April in 2016 and corresponding months in 2017. Since many companies and institutions in China provide their employees welfare of regular physical examination each year, most of the examinees in our dataset participate in physical examination in both 2016 and 2017, providing the physical condition of same group of people for two consecutive years. After filtering out the examinees who only take single year examination, there are 2,637 people remaining in the dataset.

Each laboratory result consists of the examinees identifier, a description of single test, and the result in form of digits or condition levels. Note that for each test performed by each person, there is a record. Not everyone has participated in every test. Medical history is represented by several words of examinee's statement. Brief diagnosis is represented similarly or by a sentence, made by a general practitioner through examinee's record.

B. Problem Formulation

As mentioned above, our task is to predict potential risk of a physical examinee. To achieve this goal, we define a classification problem below.

ISSN 2348 – 8034
Impact Factor- 5.070

P is the examinees set of size *n*. *B* is the basic infor-mation set of examinees. *R* is the laboratory test results set. *H* represents the medical history set and *D* represents the diagnosis set. Then *A*= (*P, B, R, H, D*) represents the whole medical dataset. Let *i*= 1· · ·*n* represents the examinee's ID
and *k*= 1,2 represents the year of physical examination. Then

$P = \{p_i\}_{i=1}^{i=n}$, $B = \{b_i^k\}_{i=1,k=1}^{i=n,k=2}$, $R = R_i^k$, $H = \{h_i^k\}_{i=1,k=1}^{i=n,k=2}$, and $D = \{d_i^k\}_{i=1,k=1}^{i=n,k=2}$. So $E_i^k = (b_i^k, R_i^k, h_i^k, d_i^k)$ is physical examination records of p_i in year k, and $E = \{E_i\}_{i=1}^{i=n}$

represents physical examination records set of all examinees. $E^1$and E^2represent the first year and second year dataset respectively.

We define a mapping *O* from *E* to *C={0,1}*, where *C* represents whether a examinee's single key result in laboratory test goes bad for two consecutive years. Our task is to find an appropriate model $f{:}E^1 \rightarrow O(\cdot)$. *f* can predict whether examinee's result in laboratory test will get worse next year based on physical condition result in previous year. The prediction should be accurate relatively to the actual data from next year and mapping *O*, i.e.

Problem 1:

$$\arg\min_f \sum_{i=1}^{n} |O(E_i) - O'(E_i^1)|$$

IV. METHODOLOGY

In this section, we describe the design of our system, including data processing procedure, algorithms, and the whole architecture. We firstly filter out some examinee features which contain lots of missing values and Secondly, we apply word vector model on diagnostic data and medical history data. Then we evaluate each feature and apply a dimensionality reduction method. Afterwards, we apply machine learning algorithms and build a self-service prediction system, which can predict risk of deterioration of key laboratory indexes for examinees. Finally, a feedback mechanism is built to collect new training data automatically.

A. Data Cleaning and Transformation

We collect original dataset from hospital information system and assign each examinee a unique ID across all the dataset. Only data from examinees who participate in physical exam-ination both years is useful for our task. After filtering out the ones who only take single year examination, there are 2,637 examinees remaining. As mentioned before, we will define a proper mapping $O{:}E \rightarrow \{0,1\}$, which is a critical discriminant of at-risk people classification. After conducting careful survey in medical center and discussing with our cooperative hospital, we define a mapping *O* according to actual work needs. Let $a^k{}_i$ represent a single numeric result of laboratory test of examinee *i* in year *k*, and U_a represent upper bound of normal limits for this result, then:

$$f(x) = \begin{cases} 1 & if\ a_i^1 \le U_a\ and\ a_i^2 > U_a\ or \\ & a_i^1 > U_a\ and\ a_i^2 > U_a\ and\ (a_i^2)/(a_i^1) > 1-\gamma \\ 0 & otherwise \end{cases}$$

γ is a coefficient related to laboratory test. It is determined by actual work of the Health Management Center and medical experience, and is usually set to 0.2 or less.

ISSN 2348 – 8034
Impact Factor- 5.070

Before applying machine learning method, we should trans-form all data into numeric form. As for categorical data in b^k_i and R_i^k, we apply one-hot encoding method to transform them into 0-1 vector. Since medical history data is in form of words or phrases instead of sentences, the NLP semantic analysis methods may not work. So we apply text segmentation to the medical history data and count the number of word occurrence as Figure 1 shows. are 'hypertension', 'fatty liver', 'high cholesterol', 'diabetes' and 'coronary heart disease'. Since other vocabularies occur rarely and play an insignificant role, we remove them to reduce the dimension of input. We take top five words occurrence as five 0-1 features word vector attached to the examinees. We do the same procedure on diagnosis data. Figure 2 shows the top 25 occurrence word in diagnosis data. There are 165 words appearing no less than 20 times in the dataset. We filter out the non-representative diagnoses and select these 165 types of diagnosis records for further processing. We transform them into 0-1 features word vector as additional data for examinees.

Figure 1: Word occurrence count in medical history

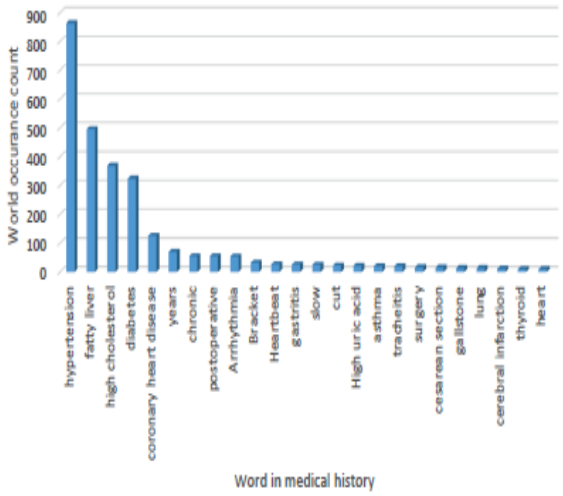

Fig. 1. Word occurrence count in medical history

Figure 2: Word occurrence count in diagnosis data

ISSN 2348 – 8034
Impact Factor- 5.070

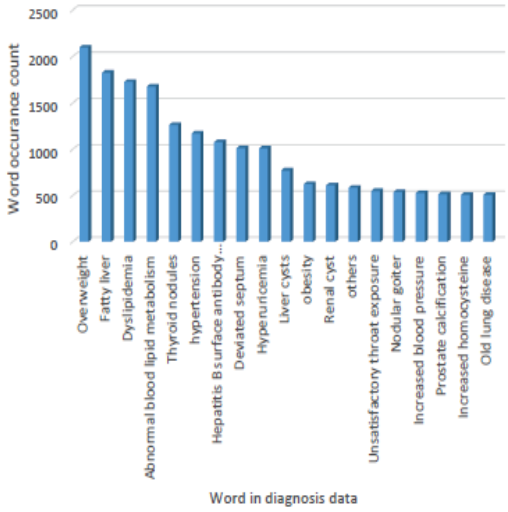

Fig. 2. Word occurrence count in diagnosis data

Since not all the examinees take all laboratory test, we select a set of test and only consider examinees who participate in these tests both years. We make a statistic about the number of participants in each test, as Figure 3 shows.

Considering the dataset size, we select tests more than 2,000 examinees attended in both years. There are 46 out of 110 tests which are qualified. After filtering out examinees who didn't participate in these tests, we get 1,700 examinees remaining. As we discussed above, there are 165 diagnosis features, 5 medical history features, 46 laboratory result features and 18 basic information features. So the each examinee has 234 features.

Figure 3: Participants count in each laboratory test

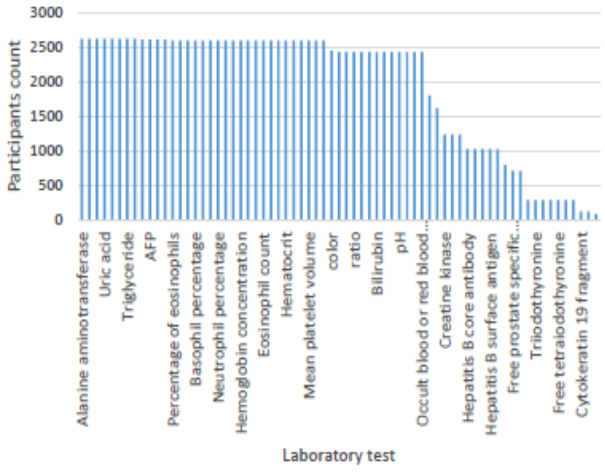

Fig. 3. Participants count in each laboratory test

ISSN 2348 – 8034
Impact Factor- 5.070

B. Dimensionality Reduction (DR)

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 overfitting 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 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.

C. Machine Learning

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

- *Decision Tree*

A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements

- *XGBoost*

XGBoost is an optimized distributed gradient boost-ing library designed to be highly efficient, flexible and portable. XGBoost provides a parallel tree boosting (also known as GBDT, GBM) that solves many data science problems in a fast and accurate way. Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.

- *Random Forest*

Random forests or random decision forests are an en-semble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests can correct for decision trees' habit of overfitting to their training set.

D. Architecture

The system architecture is shown in Figure 4. Data gathered from medical center is cleaned and transformed into numeric vectors, then dimensionality of feature vectors is reduced. Afterwards, vectors are put into the training process, which is periodically executed to make sure effectiveness of model parameters being highly optimized. Then we get the model of machine learning from the training step, which will be used in prediction task. There is a user-friendly interface for examinees and doctors. Examinees can input their physical examination records and the system will give a risk potential of examinees' physical status. Doctors can input a batch of physical examination data and get a set of examinees with potential risk of physical deterioration.

Besides, doctors can fix the prediction result manually, which will be feedbacked and recorded as new training data. Extra training step will be executed each day using new training data to revise model automatically. Thus, the system can improve the effectiveness automatically and meet the needs of doctors more precisely.

V. EXPERIMENT

Now we evaluate the performance of the proposed algorithm using the dataset from Peking University First Hospital.

A. Metrics

As discussed above, we focus on the intersection between the predicted risk examinees set and the real risk examinees set in second year. Thus, we introduce some metrics in predictionproblems to evaluate performance of proposed system. We define

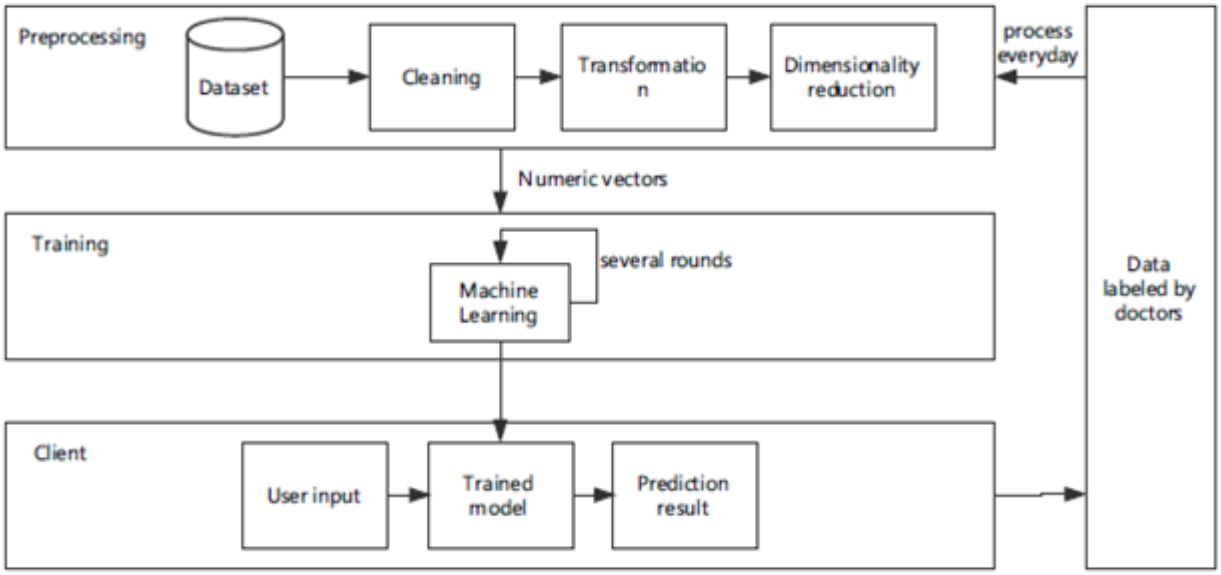

Figure 4: Architecture of the System

The true positive set as $TP=\{i|O(E_i) = 1, O(E_i^1) =1\}$,
The false positive set as $FP = \{i|O(E_i) = 0, O(E_i^1) = 1\}$,
The true negative is $TN = \{i|O(E_i) = 0, O(E_i^1) = 0\}$,and
The false negative is $FN=\{i|O(E_i) = 1, O(E_i^1) = 0\}$.

Then precision and recall is:

$$precision = \frac{|TP|}{|TP| + |FP|}$$

$$recall = \frac{|TP|}{|TP| + |FN|}$$

Precision and recall are the typical metrics in prediction problems. F1 score is a metric which is harmonic mean of precision and recall as follows:

$$F1 - score = \frac{2 \times recall \times precision}{recall + precision}$$

Table 1: the precision, recall and f1-score of the systems with different algorithms in different tasks

Task	Algorithm	Precision	Recall	F1 score
LDL-C	DT	0.7997	0.8128	0.8022
	XGBoost	0.7997	0.8146	0.7944
	RF	0.7866	0.8039	0.7747
UA	DT	0.8426	0.8538	0.8463
	XGBoost	0.8498	0.8610	0.8530
	RF	0.8395	0.8538	0.8429
TG	DT	0.7779	0.7611	0.7673
	XGBoost	0.7803	0.7807	0.7805
	RF	0.7697	0.7825	0.7718

In statistical analysis of binary classification, the F1 score is a measure of accuracy. It considers both the precision and the recall of the test. F1 score will be used as main evaluation of our experiment.

B. Results and Performance

We divide the dataset into training set and testing set by 2 to 1. Then we apply the three algorithms mentioned above in risk-prediction task of three laboratory tests. These three laboratory tests include low density lipoprotein cholesterol (LDL-C), uric acid (UA) and triglyceride (TG), which are key indicators in diagnosis of chronic disease. The precision, recall and F1-score are as Table I shows. Note that all the metrics are weighted averages of corresponding value of risk people prediction problem and non-risk people prediction problem. Figure 5 shows the comparison of F1 scores between algo-rithms and baseline, which is the result of random prediction algorithm. F1 score of all algorithm is above 0.75, far above the baseline, which means prediction using machine learning is of great practical value. The overall performance of XGBoost is better than the others', random forests and decision tree goes well in certain circumstances. Thus, we choose XGBoost as main machine learning algorithm of the system for public use.

To evaluate the function of dimensionality reduction, we compare the performance of systems with and without the dimensionality reduction step in LDL-C risk prediction task. As Table II shows, the performance of system with dimension-ality reduction is better than system without dimensionality reduction in most circumstances, which means dimensionality reduction step is of vital importance in our system. It can reduce the computational complexity as well as raise overall performance.

Then we focus on the role of medical history and diagnosis data. We compare two system with same settings but different inputs. One runs with all data while the other runs with only basic and laboratory data. Table III shows the performance of two system. Apparently, system with all data performs slightly better than the other in XGBoost. The medical history and diagnosis data play a role in prediction system and we make use of it in our system.

VI. CONCLUSION

With the rapid development of information technology, the healthcare technologies are making great progress. However, due to the lack of experienced doctors and physicians, most healthcare organizations cannot meet the demand of higher quality and more personalized healthcare service.

Table 2: The precision, recall and f1-score of the systems with and without dimensionality reduction

Settings	Algorithm	Precision	Recall	F1 score
with DR	DT	0.7997	0.8128	0.8022
	XGBoost	0.7997	0.8146	0.7944
	RF	0.7866	0.8039	0.7747
without DR	DT	0.7811	0.8004	0.7707
	XGBoost	0.7780	0.7986	0.7716
	RF	0.7947	0.8039	0.7640

Table 3: The precision, recall and f1-score of the systems with and without medical history and diagnosis data

Dataset	Algorithm	Precision	Recall	F1 score
all data	DT	0.7997	0.8128	0.8022
	XGBoost	0.7997	0.8146	0.7944
	RF	0.7866	0.8039	0.7747
without medical history and diagnosis data	DT	0.7985	0.8111	0.8014
	XGBoost	0.7944	0.8075	0.7751
	RF	0.7911	0.8075	0.7812

In this paper, we build a system that can predict examinees' risk of physical status next year based on the physical examination records this year. The system transforms basic information, laboratory data, medical history and diagnosis of examinees into numeric vectors. Several machine learning algorithms are applied to predict whether physical status of a examinee will be in danger of physical deterioration next year. With the system, it is practical for examinees to get to know their physical conditions, as well as for doctors to filter out people who really need intervention. The algorithms in system are proved to be precise and suitable for task in our experiments. It can achieve 75%~84% on F1 score, which means accuracy of the system is good enough for practice. 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 everyday.Our system is a self-service system, which can provide personalized healthcare service to examinees with few maintenance personnel.

Figure 5: F1-score of algorithms

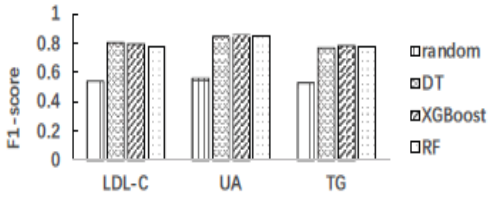

Fig. 5. F1-score of algorithms

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 huge amount of manpower and contains great potential for application. It points out a proper way to implement AI-assist medical care system.

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 Informatio 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]. Artificial 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 artificial 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 breast tumour classification in clinical screening with machine learning methods[J]. Artificial Intelligence in Medicine, 2005, 34(2):129- 139.*
16. *Nattkemper T W, Wismller A. Tumor feature visualization with unsupervised learning[J]. Medical Image Analysis, 2005, 9(4):344.*
17. *Iakovidis D K, Maroulis D E, Karkanis S A. An intelligent system for automatic detection of gastrointestinal adenomas in video endoscopy[J]. Computers in Biology & Medicine, 2006, 36(10):1084-1103.*
18. *Billah M, Waheed S, Rahman M M. An Automatic Gastrointestinal Polyp Detection System in Video Endoscopy Using Fusion of Color Wavelet and Convolutional Neural Network Features.[J]. International Journal of Biomedical Imaging, 2017,(2017-8-14), 2017, 2017(1):1-9.*
19. *Jiang T, Qian S, Hailey D, et al. Text Data Mining of Aged Care Accreditation Reports to Identify Risk Factors in Medication Management in Australian Residential Aged Care Homes[J]. Studies in Health Technology & Informatics, 2017, 245:892.*
20. *Gandhali Upadhye, Trupti Dange, cloud resource allocation as a non-preemptive approach [J] Second International Conference on Current Trends In Engineering and Technology,2014*
21. *Seokho Kang, Personalized prediction of drug efficacy for diabetes treatment via patient-level sequential modelin[J]. Artificial Intelligence in Medicine, 2018.*
22. *Koh H C, Tan G. Data mining applications in healthcare[J]. J HealthcI nf Manag, 2005, 19(2):64-72.*
23. *Miss. Pratiksha Bhegade, Miss. prajkta Dhore, Miss. Prachi Mane, Mr. Namdev Mansukh, Prof. Deepali Patil, Real Time Communication for Emergency Ambulance System:Application for Healthcare[J]International Journal of Scientific and Engineering Research,2018*
24. *Bozcuk H, Bilge U, Koyuncu E, et al. An application of a genetic algorithm in conjunction with other data mining methods for estimating outcome after hospitalization in cancer patients.[J]. Medical Science Monitor International Medical Journal of Experimental & Clinical Research,2004, 10(6):CR246.*
25. *Miss. Pratiksha Bhegade, Miss. prajkta Dhore, Miss. Prachi Mane, Mr. Namdev Mansukh, Prof. Deepali Patil, Real Time Communication for Emergency Ambulance System: Application for Healthcare[J]International Journal for research in Applied Science and Engineering Technology, 2018*
26. *Walsh P, Cunningham P, Rothenberg S J, et al. An artificial neural network ensemble to predict disposition and length of stay in children presenting with bronchiolit is.[J]. European Journal of Emergency Medicine,2004, 11(5):259-264.*
27. *Geraci J M, Johnson M L, Gordon H S, et al. Mortality after cardiac bypass surgery: prediction from administrative versus clinical data.[J].Medical Care, 2005.*