Multivariate Analysis of the Health Status of the Elderly Based on the Decision Support System Model and Countermeasures

Xiaohua Li, M.Sc
Jianbo Li, M.Sc
Xianchu Liu, M.Sc
Wenxue Tan, M.Sc

Xiaohua Li, Lecturer, School of Physical Education, Hunan University of Arts and Science, Changde, 415000, Hunan, China. Jianbo Li, Lecturer,School of Computer and Electrical Engineering, Hunan University of Arts and Science, Changde, 415000, Hunan, China.Xianchu Liu, Lecturer,School of Physical Education, Hunan University of Arts and Science, Changde, 415000, Hunan, China.Wenxue Tan, Lecturer,School of Computer and Electrical Engineering, Hunan University of Arts and Science, Changde, 415000, Hunan, China.
Correspondence LecJianbo Li;gbv2245374@163.com

Objectives: At present, China's population over 6o years old has been increasing year by year. As the problem of population aging is becoming more and more prominent, followed by a variety of health problems, which is studied.Methods: In this paper, Bayesian technology and decision tree technology are applied to the health assessment system of cardiovascular and cerebrovascular chronic diseases. Results:A health management system for chronic diseases is developed to enable health management experts to assess health according to physical examination data.Conclusion: The management of chronic cardiovascular and cerebrovascular diseases in physical examinees can reduce the incidence of chronic diseases and improve the quality of life of physical examinees.

Keywords:decision support; system model; geriatric health; multivariate analysis
Tob Regul Sci.™ 2021;7(5-2): 4914-4924
DOI: doi.org/10.18001/TRS.7.5.2.55

In recent years, the proportion of middle-aged and elderly people in China has increased yearby year and the aging process of population has been accelerating. With the increase of chronic non-communicable diseases, the national and family medical expenditure has increased dramatically. Health management emerged as the times require, especially in the health management industry for chronic diseases has developed rapidly [1]. Chronic non-communicable diseases (NCD) are the general name of cardiovascular and cerebrovascular diseases, diabetes, tumors and chronic obstructive pulmonary diseases, which have caused more than 60% of the world's deaths and are estimated to rise to 75% by 2030 (Tegn N et al.2016) [2]. In this paper, based on the physical examination data of 3078 people in a health management center in the past two years, the characteristics of cardiovascular and cerebrovascular diseases are mainly analyzed. The common coronary heart disease, hypertension, diabetes mellitus, hyperlipidemia, hyperglycemia, body mass index (BMI) and two living habits are analyzed. With the health examination system as the data support, the application and development of the health management system for chronic diseases are carried out. The functions of customer management, health files, health assessment, health intervention, chronic disease management and user

management modules, as well as the application interface for the health examination system are completed.

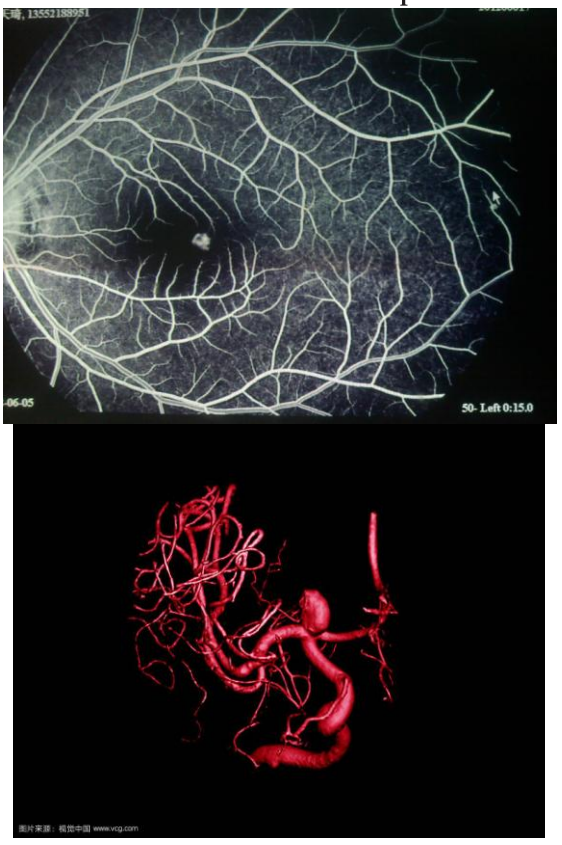
Chronic non-communicable diseases (NCD), a general term for cardiovascular and cerebrovascular diseases, diabetes, tumor, cancer and chronic obstructive pulmonary disease, have caused more than 60% of deaths worldwide and are estimated to rise to 75% by 2030[3]. Nowadays people are more and more keen on unhealthy diets such as "foreign fast food", coffee, takeout and so on. Staying up late has become the norm of many young people, even middle-aged people. These habits are closely related to the occurrence of chronic diseases[4]. To strangle chronic diseases in the cradle, the sooner we need to detect them and make active and effective health management, which is to prevent them early, to treat them quickly, and to make health management our daily necessities[5]. These data do surprise us and alarm us that we need to focus more on the prevention of chronic diseases[6]. Relevant departments in China have also paid more and more attention to and promulgated a set of laws and regulations, and also have carried out a series of medical system reform, so that everyone's health is guaranteed. Everyone can be cured of illness and serious illness can also be reimbursed, which establish a public medical service management system that makes everyone look up to the hospital and can afford to stay[7].

METHODS

Data Mining Technology

Data mining is a process of automatically discovering useful information in large database repositories. Data mining mainly uses some specific data mining algorithms to discover the relevant knowledge from the data within a certain range of operational efficiency, which determines the effect and efficiency of the whole process of knowledge discovery [8]. Data mining technology can solve the problems of data scalability, high dimensionality, heterogeneous data and complex data. Researchers from many different disciplines come together to develop effective tools for processing different data [9]. Data mining technology, which incorporates the methods of abstraction, estimation and hypothesis testing in statistics, and search algorithm in artificial intelligence, is a multi-disciplinary technology[10]. Data mining technology is particularly important in the era of big data and Internet +. Various spark, storm and other computing frameworks that follow the pace of the times are also emerging [11]. Data mining is applied to all walks of life and further promotes the development of artificial intelligence and pattern recognition. With the development of information technology, people's ability to generate data and store data has been greatly improved. People need newer and more effective means to mine large amounts of data and bring them into full play [12]. Knowledge discovery is the process of extracting relevant information from database by using data mining technology. Data mining is an important stage of knowledge discovery [13]. Data mining is a process of extracting useful information from massive, complex, ambiguous and random data. Data mining based on the original or existing experience to predict, which helps people make more favorable decisions [14]. The main processes of data mining include data cleaning, data integration, data selection, data transformation, data mining, model evaluation and knowledge representation [15]. Data mining technology is mainly to train the algorithm model of the data stored in the database so as to get the best prediction model. Then by using the model to predict the new data, the angiogram of the elderly based on the data technology is as follows:

Figure1
Vascular Shadow Map



On the one hand, data mining is the main process of training and optimizing prediction models. On the other hand, data mining technology can further enable enterprises to understand their own situation, which do a good job in advance of customer loss, life and death potential, chronic diseases and other predictions, prevent in the future. Mining technology mainly uses predictive mode. Predict future trends through past and latest data to help people assess the best business decisions and make health risk predictions for chronic diseases. Data mining technology can be classified according to different types of data, such as knowledge classification, database classification, mining methods classification, mining technology classification and so on. The main methods of data mining include decision tree, Bayesian technology and statistical analysis. Each method has its own suitable fields and conditions and its own advantages, so we should choose data mining technology according to the established conditions, because the choice of technology affect the final results. In this paper, decision tree, Bayesian technology, statistics and other classification methods are studied to select the data mining algorithm suitable for the health management system prediction model.

Routine Health Assessment Model Based on Naive Bias Classification

The process of health assessment is the process of analyzing and summarizing diseases. Naive Bayesian algorithm has the advantages of stable classification efficiency, simple calculation, and high classification accuracy and meets the purpose of this chapter. Therefore, naive Bayesian method is used to construct the chronic disease assessment and prediction model. The Bias theorem is calculated as follows. Calculate P (X) as the full probability formula:

$$P(X)=\sum_{i=1}^{n}P(Yi)P(X\mid Yi) \quad (1)$$

According to the data and conditions of cardiovascular and cerebrovascular disease evaluation, Since P (Y) is used in [0,1], the probability of cardio cerebrovascular disease in physical examination is shown in Table 1. The collection of 10 classes Y={y1, y2, y3,... y10}={coronary heart disease, hypertension, diabetes mellitus, smoking history, drinking history, increased triglycerides, total cholesterol, low density lipoprotein, hyperglycemia, body mass index} can be drawn. A prior probability of cardiovascular and cerebrovascular disease P (Y) is as follows:
P(Y1)=24,P(Y2)=6,P(Y3)=3,P(Y4)=3,P(Y5)=2,
P(Y6)=3,P(Y7)=3,P(Y8)=3,P(Y9)=2,P(Y10)=6.
Since P (Y) is used in [0,1], the probability of cardio cerebrovascular disease in physical examination is shown in Table 1.

Table1
Probability of Cardiovascular and Cerebrovascular Diseases

Name of the disease	The probability of illness is P (Y=1).	No probability of illness P (Y=0)
coronary heart disease	12/17	5/17
Hypertension	6/7	1/7
Diabetes	3/4	1/4
History of smoking	3/4	1/4
A history of drinking	2/3	1/3
Elevated total cholesterol	3/4	1/4
Increased triglyceride	3/4	1/4
Increased LDL	3/4	1/4
Hyperglycemia	2/3	1/3
Body mass index increased	6/7	1/7

It is very difficult to accurately estimate the posterior probability of every possible combination of attribute values. For the collected data of physical examination data, because of the large attribute data, a large training set is needed. By Bayesian theorem, the posterior probability is represented by the prior probability P (Y), the class conditional probability P (X | Y) and the evidence P (X), as shown in formula (2):

$$P(Y \mid X) = P(X \mid Y)P(Y) \mid P(X) \quad (2)$$

P (X) can be obtained from the full probability formula (1) and P (X) is always a constant, so only the class conditional probability P (X | Y) needs to be calculated. Set Z to represent the collected physical data set, which is independent of Y in given Z and X conditions.

$$P(X \mid Y, Z) = P(X \mid Z) \quad (3)$$

According to the physical data collected from 3078 people, the number of each disease is counted. The number of six diseases and two habits in the physical examination data is shown in Table 2.

Table2
Number of Seven Diseases and Two Habits in Physical Examination Data

Name of the disease	Number of people
coronary heart disease	322
Hypertension	353
Diabetes	139
History of smoking	247
A history of drinking	177
Elevated total cholesterol	802
Increased triglyceride	628
Increased LDL	368
Hyperglycemia	438
Body mass index increased	1275

Fast Health Assessment Model Based on Tree Decision Tree C4.5 Algorithm

In the actual application, many units of the physical examiner cannot meet all the criteria for cardiovascular and cerebrovascular diseases, which is because of the unit's economic situation or the physical examiner himself only choose some items for physical examination and other reasons. The health assessment model established by Bayesian technology must meet all the judgment conditions of cardiovascular and cerebrovascular diseases and cannot be null to ensure its accuracy. Therefore, the rapid health assessment model based on decision tree C4.5 is used to assess the health of health examinees, and the accuracy of the assessment results is improved. C4.5 algorithm chooses the highest information gain rate as the criterion and the attribute of the highest information gain rate as the splitting attribute. There are s data samples in set S. Si is in class Ci(i=1 , 2,… , m) number of samples. The entropy of ensemble S in Ci classification is as follows:

$$I(S1,S2,\ldots,Sm)=-\sum_{i=0}^{m} pi\log 2\,pi \quad (4)$$

$pi=\frac{Si}{S}$ represents the probability that any sample belongs to Ci class. Suppose that the attribute A has v different values {a1, a2..., av}, by attributes A, the dataset S can be divided into v subsets. Among them, Sj indicates that the value of attribute A in data set S is aj(j=1 , 2,… , v) itself that is sij indicating that the subset of Sj belongs to the Ci class.

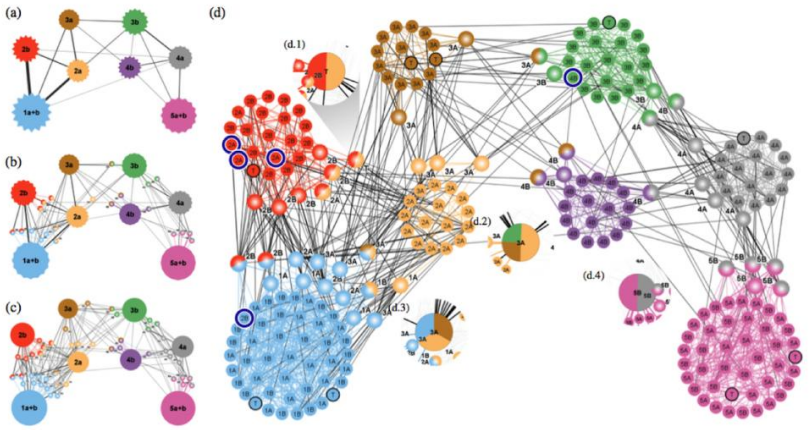
$$E(A)=\sum_{i=1}^{v}\frac{sij+s2j+\ldots+smj}{s}I(s1j,s2j,\ldots smj) \quad (5)$$

$\frac{sij+s2j+\ldots+smj}{s}$ represents the proportion of Sj subset in data S, corresponding to the given subset Sj:

$$I(s1j,s2j,\ldots smj)=-\sum_{i=1}^{m} Pij\log 2Pij \quad (6)$$

Among them, the probability of sample data class Ci in $Pij=\frac{smj}{sj}Sij/Sj$. In this paper, the data of physical examination in two units of a health management center from 2016 to 2017 are collected to 3652 samples. At the same time delete the past history and family history of the two conditions, only physical examination data, the final number of physical examination, 1242 men, 1836 women, the data has a certain representative. The basic data of the physical examiner comes from the personnel provided by the unit. The data is verified by the physical examination and there is no missing data, so no cleaning is needed. Although the physical examination data are added by specialists, there are still some physical examination personnel did not complete all the physical examination items. To clean up these data, the final physical examination data is 3078. The basic data such as the name, sex, age of the physical examiner have no effect on the establishment of the model, so these values can be ignored when constructing the decision tree. Disease determination conditions in physical examination data cannot distinguish the physical condition of the physical examination line, so it is necessary to increase the attributes of cardiovascular and cerebrovascular diseases. After analyzing and processing the data, the physical examination data table is obtained. The purpose of data generalization is to abstract a large amount of data, and then make it suitable for data mining, data statistical model construction. Sex does not need to be generalized as "male" or "female". The generalization of body mass index, blood pressure, blood lipid and blood sugar over normal range is "abnormal". The generalization in normal range is "normal". When one of the indexes is abnormal, it is generalized as "cardiovascular and cerebrovascular diseases" and is expressed by A. When it is normal, it is generalized to "no cardio cerebrovascular disease" by B. The classification of data is shown in Figure 2:

Figure2
Data Clustering Image

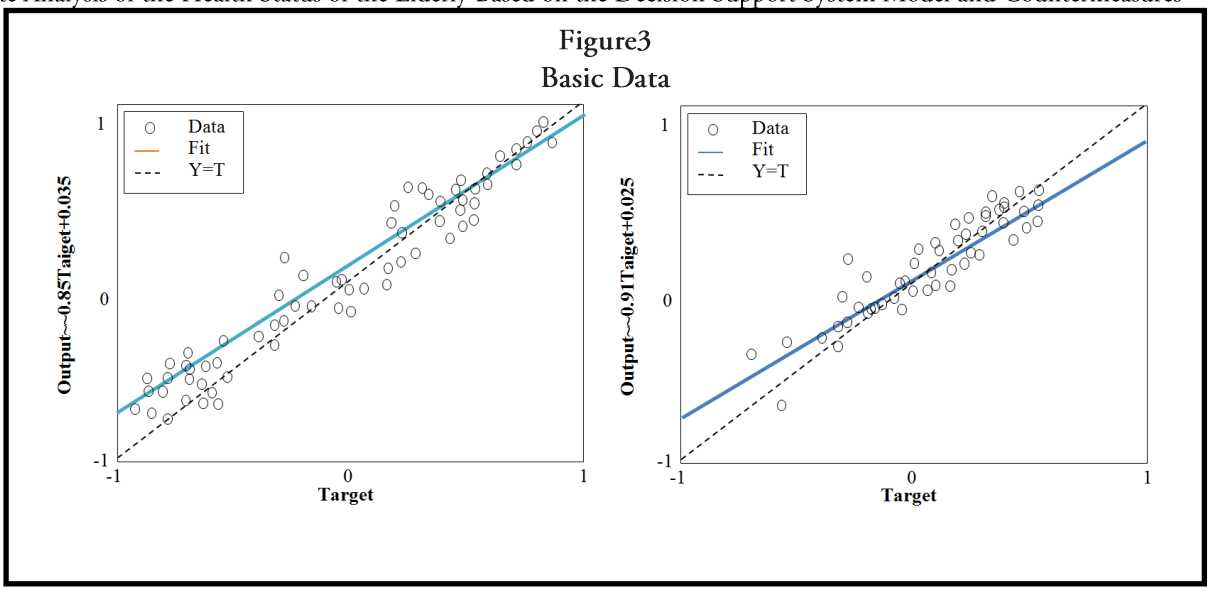


RESULTS

Data Presentation

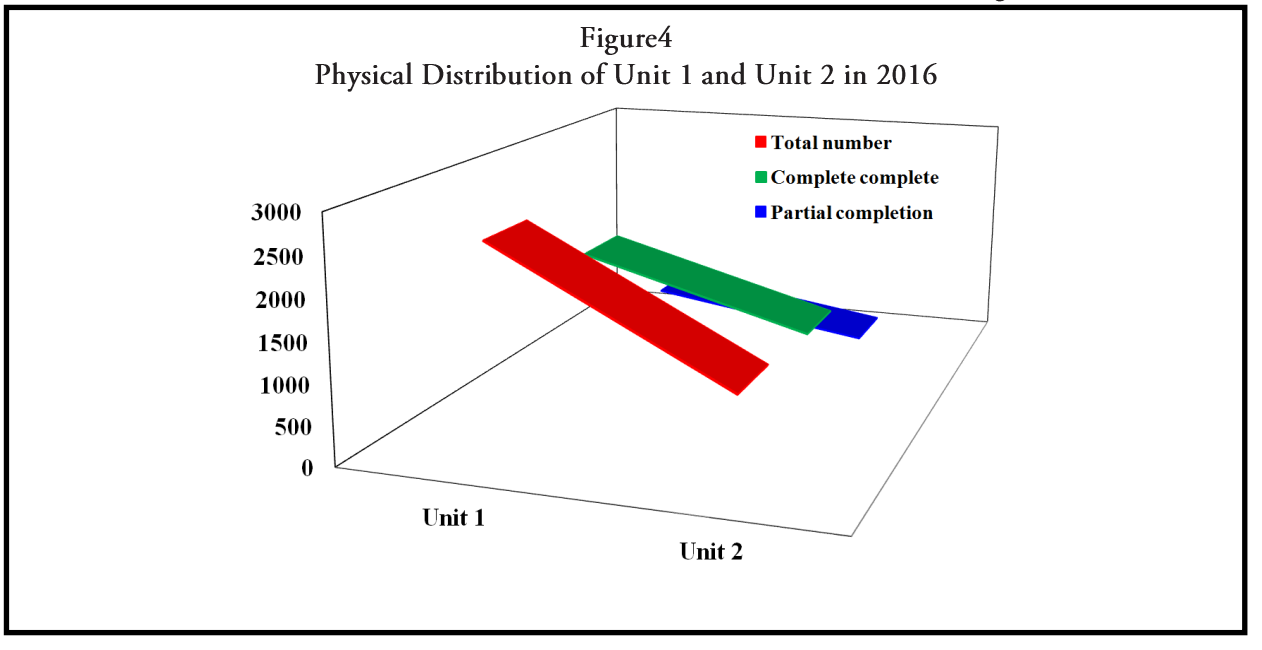
In this paper, relying on the health examination carried out by the health management center of a unit, the research on the health management system of cardiovascular and cerebrovascular chronic diseases is carried out on the basis of the health examination data of the past two years stored in the physical examination system database. Up to now, the health examination center has been transformed from a simple index examination, the continuation of the pattern of physical examination into a health examination, disease management as one of the management model. The center consists of senior physicians of the main inspection team, for each physical examiner issued general inspection recommendations, provide personalized, systematic, targeted health guidance and counseling services. For large companies and groups, the center also provides overall monitoring, analysis and evaluation. Comprehensive assessment of the health risks and disease incidence of the unit personnel provide health counseling and guidance and prevention of health risk factors. The chronic non-communicable diseases studied in this paper are mainly cardiovascular and cerebrovascular diseases. Here, six common chronic diseases and two living habits are analyzed and studied. Before the feature extraction and modeling, the data of physical examination are fully analyzed and the data fully understood. Reasonable feature extraction and modeling are done according to the classification of data sets. In this paper, the basic information of physical examination, past history, family history, physical examination indicators and other aspects of data exploration and processing. Comparing the medical data collected in 2016 and 2017, the number of medical examinations decreases in 2017 due to various reasons, such as turnover, death and so on. However, the number of people who have completed all the physical examination items is relatively flat, and the number of people who have partially completed the physical examination items decreases more, which shows that people pay more and more attention to the physical examination. The basic data are shown in Figure 3:

Figure3
Basic Data



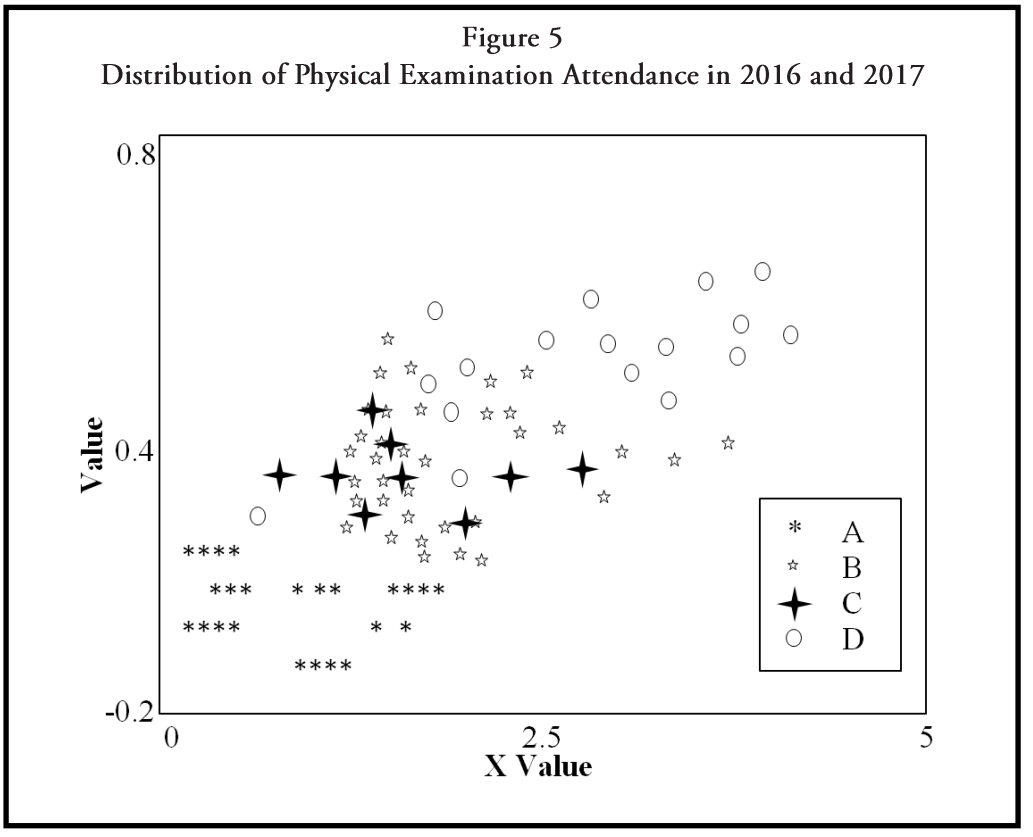
For reasons of confidentiality of the name of the unit and the basic information of the personnel, unit 1 and unit 2 are used to represent two different units. In 2016, a total of 2548 people were examined in Unit 1, 1904 completed the physical examination items, 644 partially completed, including 935 males and 1613 females. In 2016, 1104 people participated in the physical examination of Unit 2, 978 completed the physical examination items, and 125 partially completed 567 men and 573 women. In 2016, the physical distribution and sex distribution of units 1 and 2 are shown in Figure 4.

Figure4
Physical Distribution of Unit 1 and Unit 2 in 2016



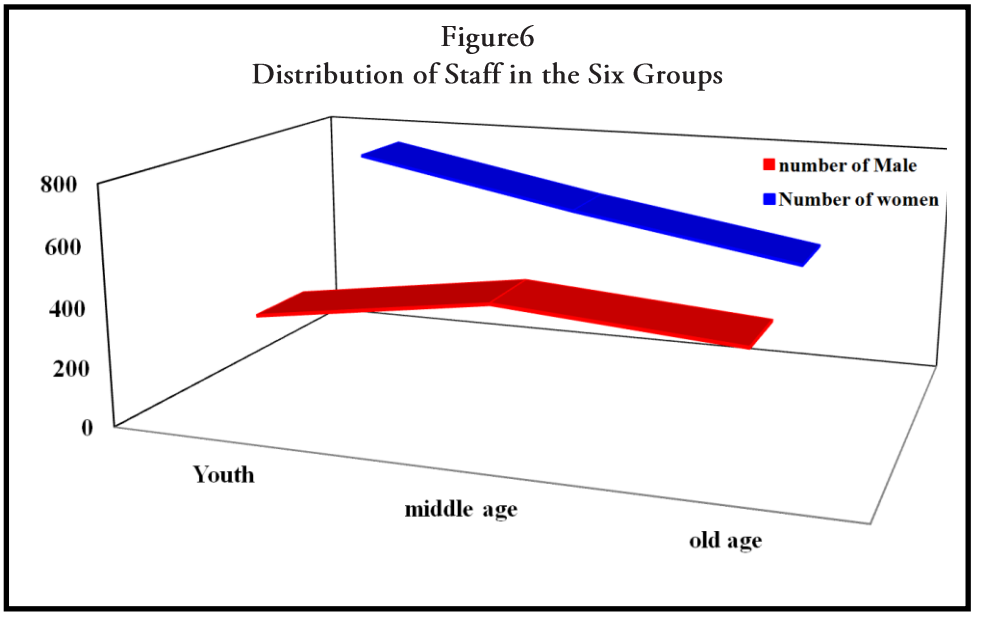
In 2017, 2100 people participated in the physical examination in Unit 1. 2072 people completed the physical examination project, 28 people partially completed, including 766 men and 1334 women. In 2017, 978 people partici pated in the physical examination in Unit 2. 916 people completed the physical examination and 62 people partially completed the physical examination, including 476 men and 502 women. In 2016 and 2017, a total of 3078 people participated in the physical examination at the

same time, and all completed the designated project, a total of 2882 people, partially completed 196 people. The specific distribution of physical examination attendance in two years is shown in Figure 5.



According to sex, 1242 males and 1836 females participate in the physical examination. There are great differences in sex between the physical examinees and the females. Consult relevant international authoritative medical journals and health management experts' suggestions according to the relevant information of physical examination personnel. The physical examination data are divided into youth group, middle-aged group and elderly group according to age, and then the above three groups are divided into six groups: male youth group, female youth group, male middle-aged group, female middle-aged group, male elderly group and female elderly group. The physical examination data in the physical examination group according to youth group, middle-aged group, the elderly group three groups for statistics. Among them, there are 1107 youth group, 1083 middle-aged group and 888 elderly groups. The youth group and the middle age group are at the same level, while the elderly group is relatively small. According to age and sex, the physical examination personnel are divided into six groups and the number of physical examination in each group is statistically analyzed, including 356 male youth group and 751 female youth group. There are 480 middle-aged men and 603 middle-aged women. There are 406 men in the elderly group and 482 in the female elderly group. The statistics on the number of staff in the six groups are shown in Figure 6.

Figure6
Distribution of Staff in the Six Groups



As can be seen from Figure 6, there are more women than men. Physical examination personnel for their own information considerations, there are changes in information, resulting in each ID corresponding to the information appear even many records.

In view of this situation, the user's information is sorted according to the ID card number or telephone number, and the latest information of each user is selected. Then match them according to ID and fill in the missing values in the latest record. The values are derived from the values in the previous record. After the missing value of the latest record is processed, the data is de-duplicated and the redundant records corresponding to each ID are deleted. Finally, the number of records taken by the physical examination personnel before and after re loading is shown in Table 3.

Table3
Basic Information of Physical Examination Personnel Before and After Re Loading

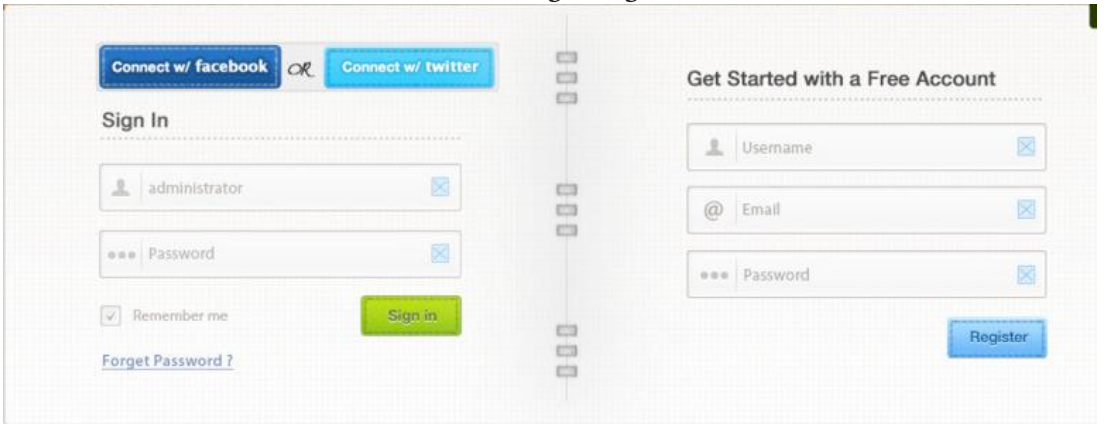
	Before repeating	After repeating the duplication
Basic information of physical examination personnel	3352	3087

As can be seen from table 3, the number of data to be re loaded is about 1.1 times that of the data after being re loaded. If not cleaned, the repetition rate of data is relatively high. Because the physical examination items are fixed and not compulsory, there are many choices for the physical examiner and they can choose some of the physical examination items for inspection, resulting in incomplete physical examination data. It is necessary to further deal with the missing values in the physical examination data. The conventional methods of missing value processing include deleting cases containing missing value, filling with special value, filling with average value and filling with median value. The method used in this paper is to delete the missing important data with the deletion of the missing value of the case processing method, for the other use of special value filling method. Outliers refer to the difference between the measured value and its average value, which is larger than the standard deviation of two times.

Analysis of Test Results

After a detailed analysis of the system requirements, it is necessary to select a reasonable development tool for development, because the choice of development tools for the system development efficiency, development costs, system stability and later maintenance are very important. The development environment of the system includes Struts2 and JDK7. Among them, JDK is the main function of the entire Java, including the use of Java environment, Java development tools and Java base class library. In MVC design mode, input, processing and output are processed separately, which includes three core components: Model, View and Controller, which deal with their own transactions respectively. A view is a page that the user operates. JSP technology is used to login and perform corresponding operations by completing the user name and password. After authorizing the system user, the user enters the system through the login page and operates. The user login page is shown in Figure 7.

Figure7
User Login Page



This chapter first introduces the development tools and operating environment of the system. Then the connection of database, the implementation of MVC mode and the display of important pages are introduced in detail, and the key code of the implementation process is listed. The system is developed based on MVC model and struts 2 frameworks, which has realized many functions of health management system, such as customer management, health files, health assessment, health intervention and chronic disease management.

DISCUSSION

In recent years, the proportion of middle-aged and elderly people in China has increased year by year and the aging process of population has been acceler ating. With the increase of chronic non-communicable diseases, the medical expenditure of the state and the family has increased dramatically, and health management has emerged as the times require. In this context, in this paper, a rapid health assessment model is built based on the Bayesian Theorem of conventional health assessment model and decision tree algorithm. Compared with several typical data mining algorithms, Naive Bayesian algorithm is selected to construct the conventional health assessment model and C4.5 algorithm in the decision tree is used to construct the rapid risk assessment model of chronic diseases. Naive Bias relies on the prior probability. Then the posterior probability is calculated by Bayesian theorem and the class with the maximum posterior probability is selected as the class to which the object belongs. C4.5 algorithm in decision tree is a classification

model based on policy choice, which can classify unknown data efficiently. A fast health assessment model based on C4.5 algorithm is established. Using different data sets to test, the test results have high accuracy.

Human Subjects Approval Statement

This paper did not include human subjects.

Conflict of Interest Disclosure Statement

None declared.

Acknowledgements

The study was supported by"Evaluation of Sports Skill Learning Based on Core Literacy in Primary and Middle Schools: National Social Science Foundation of Pedagogy Program In General, China (Grant No. BLA170222).

References

1. CanouiP F, Reinald N, Laurent M, et al. Geriatric assessment findings independently associated with clinical depression in 1092 older patients with cancer: the ELCAPA Cohort Study. *Psycho-oncology.* 2016;25(1):104.
2. Frantzidis C A, Gilou S, Billis A, et al. Future perspectives toward the early definition of a multivariate decision-support scheme employed in clinical decision making for senior citizens. *Healthcare Technology Letters.*2017;3(1):41-45.
3. Genes N, Kim M S, Thum F L, et al. Usability Evaluation of a Clinical Decision Support System for Geriatric ED Pain Treatment. *Applied Clinical Informatics,* 2016;07(01):128-142.
4. Grung B, Sandvik A M, Hjelle K, et al. Linking vitamin D status, executive functioning and self-perceived mental health in adolescents through multivariate analysis: A randomized double-blind placebo control trial. *Scand J Psychol.* 2017;58(2):123-130.
5. Gupta S K, Ansari F A, Nasr M, et al. Multivariate analysis and health risk assessment of heavy metal contents in foodstuffs of Durban, South Africa. *Environmental Monitoring & Assessment.* 2018;190(3):151.
6. Hai–Yan L I, Guo Y, Tian C, et al. A risk prediction score model for predicting occurrence of post-PCI vasovagal reflex syndrome: a single center study in Chinese population. *Journal of Geriatric Cardiology : JGC.* 2017;14(8):509-514.
7. Heidari E, Andiappan M, Banerjee A, et al. The oral health of individuals with dental phobia: a multivariate analysis of the Adult Dental Health Survey, 2009. *Br Dent J.*2017;222(8):595-604.
8. Kaori K, Yumi W, Kazutoshi N, et al. Modifiable Factors Associated with Cognitive Impairment in 1,143 Japanese Outpatients: The Project in Sado for Total Health (PROST). *Dementia & Geriatric Cognitive Disorders Extra.* 2016;6(2):341-349.
9. Kikkert L H J C, Vuillerme N, Campen J P V, et al. Gait characteristics and their discriminative power in geriatric patients with and without cognitive impairment. *Journal of Neuroengineering & Rehabilitation.* 2017;14(1):84.
10. Kokura Y, Maeda K, Wakabayashi H, et al. High Nutritional-Related Risk on Admission Predicts Less Improvement of Functional Independence Measure in Geriatric Stroke Patients: A Retrospective Cohort Study. *Journal of Stroke & Cerebrovascular Diseases the Official Journal of National Stroke Association.* 2016;25(6):1335-1341.
11. Kwj K, Timman R, Busschbach J J, et al. Multivariate Analysis of Health-related Quality of Life in Donors After Live Kidney Donation. *Transplantation Proceedings.* 2018;50(1):42.
12. Ozgur B O. The Awareness and Educational Status on Oral Health of Elite Athletes: A Cross-Sectional Study with Cluster Analysis. *Educational Research & Reviews,* 2016, 11(16):1521-1526.
13. Pérezzepeda M U, Gutiérrezrobledo L M. Calf circumference predicts mobility disability: A secondary analysis of the Mexican health and ageing study. *European Geriatric Medicine.* 2016;7(3):262-266.
14. Sullivan J L, Eisenstein R, Price T, et al. Implementation of the Geriatric Patient-Aligned Care Team Model in the Veterans Health Administration (VA). *Journal of the American Board of Family Medicine Jabfm.*2018;31(3):456.
15. Tegn N, Abdelnoor M, Aaberge L, et al. Invasive versus conservative strategy in patients aged 80 years or older with non-ST-elevation myocardial infarction or unstable angina pectoris (After Eighty study): an open-label randomised controlled trial. *Lancet.* 2016;387(10023):1057-1065.