ISSN 2618-1584 Vol. 5, Issue 6: 49-54, DOI: 10.25236/FMSR.2023.050609

The Classification of Heart Disease Based on Artificial Network

Yang Sun[1,a,*]

[1]*International Education College, Changchun University of Technology, Changchun, Jilin, 130000, China*
[a]*20191161@stu.ccut.edu.cn*
**Corresponding author*

Abstract: *Heart disease is a chronic disease which is not infectious but has high mortality in clinic. It is difficult to be accurately analyzed by traditional medical decision-making and diagnosis methods. With the emergence of a large number of clinical diagnosis, treatment and examination reports and network electronic medical record data, information technology provides a large number of data basis for interventional medical diagnosis and auxiliary medical pathological diagnosis. Based on plenty of data for clinical diagnosis and treatment for effective extraction and processing, machine learning algorithms can make accurate diagnosis for heart diseases, predict the probability of disease disease cases and patients, and combined with relevant professional knowledge, in the field of balancing potential data analysis and processing, in order to get better disease diagnosis, improve disease prevention, diagnosis and research status.*

Keywords: *Heart disease, Electronic medical records, Medical pathological diagnosis, Auxiliary medical diagnosis, Machine learning*

1. Introduction

In the social and economic progress and development to people's living conditions, but also brought a lot of new problems, since the 1990s, cardiovascular disease, heart disease blowout development, has rapidly become the main cause of death of Chinese residents. The situation of each patient is different. With the increase and accumulation of a large number of disease data, the relationship between diseases is becoming more and more complex. The medical diagnosis method still in the handicraft stage gradually cannot meet the needs of the present society, and the probability of doctors' misdiagnosis is also increasing gradually. Because of the complex structure of the heart, the risk factors can be a combination of conditions and not a single cause, requiring extremely specialized diagnosis. Generally speaking, doctors make judgments about patients based on their own diagnostic experience. Therefore, for the same patient, different qualified doctors may give different diagnosis results and treatment plan. This may miss the patient's best chance of diagnosis and may lead to irreversible results in the future. The above situation often occurs in regions with relatively backward medical resources. Doctors in small towns may not be able to make accurate judgments due to their lack of experience, which also leads to the embarrassing situation that it is hard to get a vote for medical resources in big cities. In order to avoid the misjudgment caused by the inexperience of attending physicians, we can use the classification algorithm of machine learning to give reference opinions and assist doctors in the judgment of diseases[1]. In most cases, a reliable classification algorithm can ensure that doctors make a fairly accurate judgment. In the application scenario of disease classification, machine learning can apply different classification algorithms for different diseases to obtain different classification results. Therefore, if machine learning classification technology is applied in the diagnosis of heart disease, it can not only assist doctors to make judgment to a certain extent, but also further enhance the application value of machine learning in other disciplines.

2. Literature review

Firstly, the relevant literatures on prediction and classification of heart disease published by foreign scholars since 2014 were investigated. Tripathy R.K. et al used the least square support vector machine model to detect the presence of cardiovascular disease in patients in 2014 [1]. In 2015, Rati Wongsathan

ISSN 2618-1584 Vol. 5, Issue 6: 49-54, DOI: 10.25236/FMSR.2023.050609

et al. used three types of neural networks, namely multi-layer perceptron neural network (MLP-NN), radial basis function neural network (RBF-NN) and generalized regression neural network (GR-NN), to test the effect of heart disease classification [2].In 2019, Heena Farheen Ansari et al. proposed an improved breadth K-nearest Neighbor algorithm (SKNN) based on the traditional K-nearest Neighbor algorithm, and the results showed that the algorithm had good performance in the detection and classification of heart disease [3].Khen Mahendra et al. proposed a heart segmentation method based on DenseNet's full convolutional neural network, and proposed a new double-loss function whose weighting scheme could combine the advantages of cross entropy and dice loss, thus bringing qualitative improvements to segmentation [4].At the same time, the excellent literatures on heart disease detection and classification published by domestic experts and scholars since 2014 were also carefully studied. For example. In 2014, Deepika S. and Jaisankar N. introduced in detail the network structure and algorithm description of three artificial neural network algorithms, applied extreme learning machine algorithm to the application test of heart disease detection and classification, and compared the experimental results with support vector machine and BP algorithm [6]. Zhuang Qiaohui proposed the application of improved random forest algorithm based on full convolutional network. XGBoost and fully connected network were used to classify heart disease respectively, and the classification effects of the two methods were compared [7].

3. Feature screening for the diagnosis and prediction of heart disease

In this experiment, we adopted five algorithms of conditional attribute selection and dimensionality reduction of heart disease data set. Then we established a classification model based on support vector machine (SVM) algorithm, and adopted 50% cross validation method to verify the reliability of the classification model. Here, we use sensitivity and specificity as the evaluation criteria for classification effect.

3.1. Heart disease data preprocessing

We first looked at whether we selected 24 condition attributes, including gender, type of chest pain, whether hypertension, fasting glucose level, resting electrocardiogram results exercise electrocardiogram results, use of digitalis, use of exercise electrocardiogram results on B blockers, use of nitrate exercise electrocardiogram (ECG) results were normal. Use of calcium channel blockers exercise ECG is normal, the use of diuretics exercise ECG is normal, if exercise can lead to angina, is low blood pressure, and 13 in the positive and negative attributes discretization of ST segment slope peak of sports, such as the rest of the 11 attribute discretization unfinished, is still beyond the scope of the Machine data form. In addition, there was a small amount of missing data in our data. Because the amount of missing data was not large, we adopted a simple average filling method. That is, calculate the average assignment value of each attribute at each level and the condition of heart disease, and then calculate the average fill number to the missing position of this attribute at this level.

3.2. Attribute reduction based on optimization algorithm

Optimization algorithm is often used to solve multi-objective optimization problems, which is used to solve the contradiction between each dimension when the objective dimension is larger than 3 dimensions [2]. In other words, optimization is often used for attribute reduction, which can usually get better reduction effect. This paper selects two optimization algorithms based on attribute reduction, genetic algorithm and attribute importance heuristic algorithm.

3.3. Overview of genetic algorithms

In essence, genetic algorithm is the solution space and solution process of random attempt problem. In the process of random attempt, new solutions are constantly created, and the worst part of the solution is gradually eliminated, while the optimal part of the solution is always retained. When it is no longer in the process of finding a new solution, that is, the solution space of the subset is stable, we can say that we have found a local optimal solution or even a global optimal solution. In the application of genetic algorithm, we first take the solution space of binary coding problem and every solution encoded in the space as individuals, and any number of coding individuals as individual genes, and these individuals constitute the whole population[3]. Then we can generate an initial population, based on the principle of evolution, the individual in the population selection, crossover and mutation operation, after the evolution

ISSN 2618-1584 Vol. 5, Issue 6: 49-54, DOI: 10.25236/FMSR.2023.050609

of some algebra, will eventually produce a stable population, that is, we are looking for the optimum population (140, we need to pay attention to, the available optimum population can be a global optimal solution, but also can be a local optimal solution.

4. Overview of classification algorithms

Machine learning, which started in 1952, has developed rapidly in recent years, and has been constantly combined with various fields, with great development and changes in all walks of life. As an important part of machine learning field, a lot of practices have been carried out and great success has been achieved. Machine learning can be divided into supervised learning and unsupervised learning, semi-supervised learning, reinforcement learning and deep learning. Supervised learning is usually given by learning training sets, which can classify new data by exploring its internal relationship and establishing a classification model, so as to realize the projection of new data. Training sets are usually used to study data that are complete and the results are clear. The main difference between unsupervised and supervised learning is that training sets are not labeled as results and do not specify the results that need to be classified. There is semi-supervised learning between the first two. Reinforcement learning is often used to solve the problem of constantly choosing better learning strategies in the interaction between ontology and environment in order to maximize the benefits..It mainly uses neural network model to process data, and induces abstract logic to establish classification model and realize classification[4].

4.1. Supporting vector machine algorithm

Support vector machine (SVM) is a binary classification model, which shows a good classification effect for small samples. The core idea is to use a data set to map to a high-dimensional space and find a hyperplane in it, so as to achieve data classification. For nonlinear problems, mapping is to find a kind of adaptive, transform the sample data into linear fraction data, and find the hyperplane that can be classified. For nonlinear problems, we usually take nonlinear transformation to map the training set data to the high-dimensional space, transform the nonlinear problem into a linear problem, and make the samples in the high-dimensional space linearly separable. That is, if the original dimension of the sample is finite, then there must be a higher dimensional space that can make the original sample linearly separable at higher dimensions.

4.2. Logistic regression algorithm

Logistic regression algorithm is applied to the analysis of the dependent variable is usually a kind of qualitative variables, is to adapt to the high frequency of a classification algorithm, are often used to solve the problem of binary classification and classification, practical strong, logically, regression algorithm is essentially based on linear regression algorithm, increased the Sigmoid function, namely the logic function[5]. Logistic regression algorithm does not need to consider the normality of independent variable types and data in application, and its coefficients can be perfectly explained. Compared with other classification algorithms, Logistic regression algorithm has unique advantages.

4.3. Naive Bayes algorithm

In addition to the two most widely used classification algorithms, naive Bayes algorithm is also widely used. Naive Bayes algorithm as a classical machine learning method, is one of the few classification algorithm design based on probability and statistics theory,59 years old, is usually used in 160 text classification problems, in this case sample size, naive Bayes classification algorithm can obtain good results. And the principle of the algorithm is very simple and easy to implement. Naive Bayes algorithm is based on the classification model of Bayes theorem (3-35) in probability theory. In the process of application, the first assumption is that the set of attributes are mutually independent, and the joint probability distribution between each attribute from the input to the output is studied, and then the input of the classifier that studies Bayes' theorem is used in order to obtain the posterior probability maximum output. We can find that the premise of naive Bayesian algorithm is to assume that the functional attributes are independent of each other, without considering the coupling effect between them, it is very simple and rough[6]. The application of this Bayesian theorem is called "simple" Bayesian algorithm. However, a large number of practices have proved that although the set is simple and rough, the naive Bayes classifier model is very effective in many fields.

ISSN 2618-1584 Vol. 5, Issue 6: 49-54, DOI: 10.25236/FMSR.2023.050609

5. Classification diagnosis and prediction of heart disease

The diagnosis and prediction of heart disease classification problems to be solved in this step are typical of four types of classification problems. Data were used to classify 460 patients with heart disease. According to the data processing results of the first quarter, we can know that according to the patient's risk of heart disease, we will classify the data of grade 4 heart disease, the number of data categories, item 178, Item 123, item 119 and item 40, and there is an imbalance between the data, so the data we selected and the data after completion, Form article 4×120 how to measure data, all levels of 120 numbers of heart disease data. After that, we will get 24 d of one-dimensional integrated conditional attributes and decision attributes, forming a 25 x480 dimensional data set. The conditional attributes were 24 d, and the determining attributes were one-dimensional, with 120 heart disease data at all levels. Then, five algorithms were used to filter and reduce the dimension of conditional attributes, and then the classification model was built[7]. Support vector machine (SVM) algorithm was used to verify the reliability of the classification model.We need to build a 4 x(4-1)/2-6 support vector machine (SVM) classification model. Finally, we also use sensitivity classification and specificity classification as the influence of classification criteria.

5.1. Attribute reduction based on brute force algorithm

The first step is to choose a solution. The solutions of a brute force algorithm with selected attributes are selected one by one in the solution space, and the conditional attributes and decision attributes contained in the solutions are combined into a new data set as labels. Step two, categorize. In the case of using the support vector Machine (SVM) classification algorithm to build four classifiers, the public needs to build 4 x(4-1/2-6) model of support vector machine (SVM) classification, select the classifier, the final classification results of the average classification results, and record the classification accuracy. The third step is to choose the optimal solution. Since we are not considered an empty set in the solution space, we need to perform 16,777,215 operations to end up with the same number of classification results. The minimum solution of the attribute under the condition of optimal classification effect is selected, and the corresponding solution is to find the optimal solution for us. The list contains the solution and conditional attribute to find the simplest attribute reduction result for us.

5.2. Attribute reduction based on genetic algorithm

Step 1, we will start with 24 d of conditional attributes in binary code. Most conditional attributes have a value range of 0 and 1 while we will code directly, but a small number of conditional attributes, such as chest pain 3 have 4 attribute types and their range is 1,2,3,4, so we will have binary code into two digits, i.e. 00,01, 10 and 11 of Step 2 which designs preliminary parameters. After several attempts, we finally designed the maximum number of iterations to be 500, set the initial number of individuals in the population to 100, randomly generate 100 individuals, and bring these individuals into the initial population as the first generation population. Step 3 is to design the fitness function of fitting attribute reduction and calculate the fitness value of each individual. Here, we will fit the function on the basis of conditional attributes, and conduct classification design according to the classification accuracy of each individual. Here, we use the support vector machine (SVM) algorithm as the classifier, construct six classifiers in four SVM classification models, and take the average classification effect as the fitness value. Step 4 is to judge whether the population's fitness value, stop condition or iteration times reach the maximum value for iteration. Step 5,is selecting actions, in which we take individuals from a population with high individual fitness and put them into the next population as one generation with high individual fitness. Step 6 is cross operations. In the selection of excellent individuals in the parent population, a part of the individual is randomly selected, its gene fragments are exchanged, new individuals are generated, and added to the new population. Step 7 is mutation. In a new generation of population, individual genes are randomly selected and manipulated. At this point, we create a new generation of people. Step 8, return to the process of step 3. Step 9: Output the optimal solution, translate the code, obtain the result of the simplest attribute reduction of the genetic algorithm, and end the operation of the genetic algorithm.

5.3. Comparison of classification results of multiple attribute reduction methods

After careful analysis of the classification results of five attribute reduction methods, it can be seen that the sensitivity and specificity of the optimal reduction classification results are 86.4% and 95.2%, which are significantly lower than the binary classification problem of heart disease diagnosis and

ISSN 2618-1584 Vol. 5, Issue 6: 49-54, DOI: 10.25236/FMSR.2023.050609

prediction under the classification effect of minimal attribute set. In the four classification problems of heart disease classification and diagnostic prediction, the reduced features of five attribute reduction algorithms were classified, and the classification effect was compared with pure judgment. Secondly, the sensitivity of the classical genetic algorithm and the heuristic search algorithm based on attribute importance to attribute reduction after classification is less than 80%, and the sensitivity of the genetic algorithm to attribute reduction classification is even more than 75%. However, after rough set fusion, the sensitivity and specificity of genetic algorithm solutions for classification are improved by 8.3% and 3.9% respectively, and the sensitivity and specificity of heuristic algorithm solutions for attribute importance of classification are improved by 5.2% and 2.3% respectively. This shows that in the field of classification and prediction of heart disease diagnosis, it is effective to improve the classification accuracy by integrating the optimal reduction of rough set optimization search. Finally, in this relatively complex classification problem, it can be seen that in feature selection, the heuristic search algorithm based on attribute importance has better performance and classification results than the genetic algorithm. But the final classification results are also very limited. In general, these two optimization algorithms are very useful in attribute reduction methods, and combined with rough set, will achieve better attribute reduction effect in a short operation time, and then get better classification effect.

6. Conclusion

With the gradual implementation of the poverty alleviation project in China, the life of the Chinese people has been greatly improved. In recent years, the urban middle class has been the largest expansion group in human history, and it is also the largest group of people in the pursuit of good health, and the pursuit of good health is almost unlimited. In order to meet the needs of the mainstream medical care in China, people put forward more requirements on the medical system, and the doctors' work pressure is becoming more and more heavy, and the doctor-patient relationship begins to appear tension and conflict. However, the healthcare industry remains stuck in the manufacturing stage, with doctors increasingly relying on one-on-one, face-to-face service for the diagnosis process, and doctors and patients increasingly relying on the clinical experience of doctors. Against this background, we must find ways to improve the efficiency of the health care industry and streamline medical services. However, the medical industry as a personalized industry, diagnosis of the disease varies from person to person, this is the medical industry for the industrialization of the difficulty of service. With the worldwide development of ai technologies such as big data and machine learning, we are gradually finding a way to solve this problem. This paper will attempt to solve the problem of heart disease classification and diagnosis in the field of prediction, hoping to find a reliable and accurate algorithm path to help doctors diagnose heart disease classification, reduce doctors' work pressure, improve the efficiency of heart disease diagnosis. In this paper, genetic algorithm and heuristic algorithm based on attribute importance are used to optimize feature selection algorithm for two kinds of data sets. Then, we combine genetic algorithm and heuristic search algorithm based on attribute importance fusion with rough set theory to propose a new attribute selection algorithm. After screening, the attribute set is divided into test set and training set, and the classification algorithm of support vector machine (SVM) and rough set are used to compare and analyze the effect of attribute selection[8]. In the end, we will reduce the optimal results as a new data set, in which we try to use support vector machine (SVM), logistic regression and naive Bayes and three machine learning classification algorithms, and select the results of the optimal algorithm as an adaptive classification model[9]. In the problem of binary classification for simple prediction of heart disease diagnosis, the attribute importance of genetic algorithm based on heuristic algorithm is better than that of heuristic algorithm. However, in order to predict these four classification problems, the support vector machine (SVM) algorithm is compared with logistic regression algorithm and naive Bayes algorithm, and the results show that the support vector machine algorithm is more suitable for predicting the two-value classification problems of heart disease diagnosis. However, the naive Bayes classification model is the best in four classification problems.

References

[1] Tripathy R.K., Sharma L.N., Dandapat S. (2014) A new way of quantifying diagnostic information from multilead electrocardiogram for cardiac disease classification. Healthcare Technology Letters, 1, 98-103.

[2] Wongsathan R., Pothong P. (2016) Heart disease classification using artificial neural networks. Heart Disease Research and Analysis, 8, 96-99.

[3] Ansari H.F., Namdeo V. (2019) An efficient SKNN based approach for heart disease classification.

ISSN 2618-1584 Vol. 5, Issue 6: 49-54, DOI: 10.25236/FMSR.2023.050609

International Journal of Advanced Technology and Engineering Exploration, 6, 101-106.
[4] Khened M., Alex V., Krishnamurthi G. (2018) Fully convolutional multi-scale residual densenets for Cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. Medical Image Analysis, 51, 21-45.
[5] Matin M.S. (2023) Heart disease classification based on ECG using machine learning models. Biomedical Signal Processing and Control, 84, 104796.
[6] Deepika S., Jaisankar N. (2023) Review on machine learning and deep learning-based heart disease classification and prediction. The Open Biomedical Engineering Journal, 17, 25-40.
[7] Zhuang Qiaohui. (2019) Research and application of improved random forest algorithm. Master Thesis of Huaqiao University, 1, 62. http://cdmd.cnki.com.cn/Article/CDMD-10385-1019625879.htm.
[8] Otoom A.F., Abdallah E.E., Kilani Y., et al. (2015) Effective diagnosis and monitoring of heart disease. International Journal of Software Engineering and Its Applications, 9, 143-156.
[9] Cao Jiajia, Yan Yuan, Chen Yi, et al. (2022) Application of SVM based on PSO optimization in cardiology classification. Journal of Dongguan University of Technology, 29, 50-56.