HDPS: Heart Disease Prediction System

AH Chen, SY Huang, PS Hong, CH Cheng, EJ Lin

Department of Medical Informatics, Tzu Chi University, Hualien City, Taiwan

Abstract

The diagnosis of heart disease in most cases depends on a complex combination of clinical and pathological data. Because of this complexity, there exists a significant amount of interest among clinical professionals and researchers regarding the efficient and accurate prediction of heart disease.

In this paper, we develop a heart disease predict system that can assist medical professionals in predicting heart disease status based on the clinical data of patients. Our approaches include three steps. Firstly, we select 13 important clinical features, i.e., age, sex, chest pain type, trestbps, cholesterol, fasting blood sugar, resting ecg, max heart rate, exercise induced angina, old peak, slope, number of vessels colored, and thal. Secondly, we develop an artificial neural network algorithm for classifying heart disease based on these clinical features. The accuracy of prediction is near 80%. Finally, we develop a user-friendly heart disease predict system (HDPS).

The HDPS system will be consisted of multiple features, including input clinical data section, ROC curve display section, and prediction performance display section (execute time, accuracy, sensitivity, specificity, and predict result).

Our approaches are effective in predicting the heart disease of a patient. The HDPS system developed in this study is a novel approach that can be used in the classification of heart disease.

1. Introduction

The diagnosis of heart disease in most cases depends on a complex combination of clinical and pathological data; this complexity leads to the excessive medical costs affecting the quality of the medical care [1]. According to the statistic data from WHO, one third population worldwide died from heart disease; heart disease is found to be the leading cause of death in developing countries by 2010. It shows one third American adult have one or more types of heart diseases based on American Heart Association report. Computational biology is often applied in the process of translating biological knowledge into clinical practice, as well as in the understanding of biological phenomena from the clinical data. The discovery of biomarkers in heart disease is one of the key contributions using computational biology. This process involves the development of a predictive model and the integration of different types of data and knowledge for diagnostic purposes. Furthermore, this process requires the design and combination of different methodologies from statistical analysis and data mining [2,3].

In the past decades, data mining have played an important role in heart disease research. To find the hidden medical information from the different expression between the healthy and the heart disease individuals in the existed clinical data is a noticeable and powerful approach in the study of heart disease classification. Heart disease classification provides the critical basis for the therapy of patients. Statistics and machine learning are two main approaches which have been applied to predict the status of heart disease based on the expression of the clinical data [4,5]. The features of the artificial neural network (ANN) [6], high accuracy and learning rate, make it worth trying as an algorithm to the prediction of heart disease. In this paper, we propose three steps to predict the heart disease status for presenting a more efficient and accurate heart disease prediction system.

2. Methods

2.1. Data sources

In this paper, we use the heart disease data from machine learning repository of UCI [7]. We have total 303 instances of which 164 instances belonged to the healthy and 139 instances belonged to the heart disease. 14 clinical features have been recorded for each instance.

2.2. Features description

Table 1 shows the 14 clinical features and their description. .

Table 1. Clinical features and their description

Clinical features	Description
Age	Instance age in years

Sex	Instance gender
Cp	Chest pain type
Trestbps (mmHg)	Resting blood pressure
Chol (mg/dl)	Serum cholesterol
Fbs	Fasting blood sugar
Restecg	Resting electrocardiographic results
Thalach	Maximum heart rate achieved
Exang	Exercise induced angina
Oldpeak	ST depression induced by exercise relative to rest
Slope	The slope of the peak exercise ST segment
Ca	Number of major vessels (0-3)colored by flourosopy
Thal	3 = normal; 6 = fixed defect; 7 = reversible defect
Num	Diagnosis of heart disease

In Table 1, there are 14 attributes used in this system, including 8 symbolic and 6 numeric: age (age in years), sex (male, female), Chest pain type (typical angina, atypical angina, non-angina pain, asymptomatic), Trestbps (resting blood pressure in mm Hg), cholesterol (serum cholesterol in mg/dl), fasting blood sugar < 120 mg/dl (true or false), resting electrocardiographic results (normal, having ST-T wave abnormality, showing probable or definite left ventricular hypertrophy by Estes' criteria), max heart rate, exercise induced angina (true or false), oldpeak (ST depression induced by exercise relative to rest), slope (up, flat, down), number of vessels colored by fluoroscopy (0-3), thal (normal, fixed defect, reversible defect), and class (healthy, with heart-disease).

2.3. Artificial neural network

In this paper, we use C as a tool to implement heart disease classification and prediction trained via ANN. Learning Vector Quantization (LVQ), a prototype-based supervised classification algorithm, is used in this study. LVQ can be understood as a special case of an artificial neural network. One benefit of LVQ is that it creates prototypes that are easy to interpret for experts in the respective application domain. LVQ applies a winner-take-all Hebbian learning approach that is a precursor to k-Nearest neighbor algorithm (KNN) and Self-organizing maps (SOM).

An LVQ system is represented by prototypes W=(w(i),...,w(n)) which are defined in the feature space of observed data. In winner-take-all training algorithms one determines, for each data point, the prototype which is closest to the input according to a given distance measure. The position of this so-called winner prototype is then adapted, i.e. the winner is moved closer if it correctly classifies the data point or moved away if it classifies the data point incorrectly. A key issue in LVQ is the choice of an appropriate measure of distance or similarity for training and classification.

The network contains three layers: the input layer, the hidden layer and the output layer, having 13 neurons, 6 neurons and 2 neurons respectively. Figure 1 displays the framework of the model.

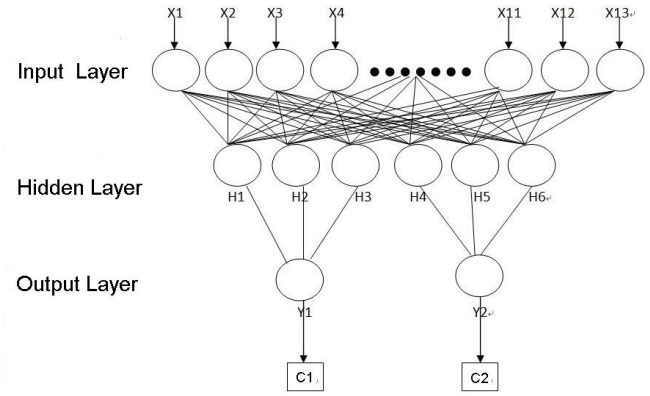

Figure 1. Framework of the ANN Model: The model contains three layers: the input layer, the hidden layer and the output layer, having 13 neurons, 6 neurons and 2 neurons respectively.

Figure 2 displays the computational steps in our codes. The clinical data are first separated into two equal parts randomly. One is used for training and the other is used for testing. Each feature is assigned randomly an initial weight. The calculated errors are used to adjust the weight of all features. The final weight of every feature is determined when the errors meet with the termination conditions. The testing data are then used to calculate the performance of this model. The process is repeated for 100 times. The output results include the average value of accuracy, specificity, and sensitivity. Finally, the ROC curve is calculated in order to check the goodness of the model.

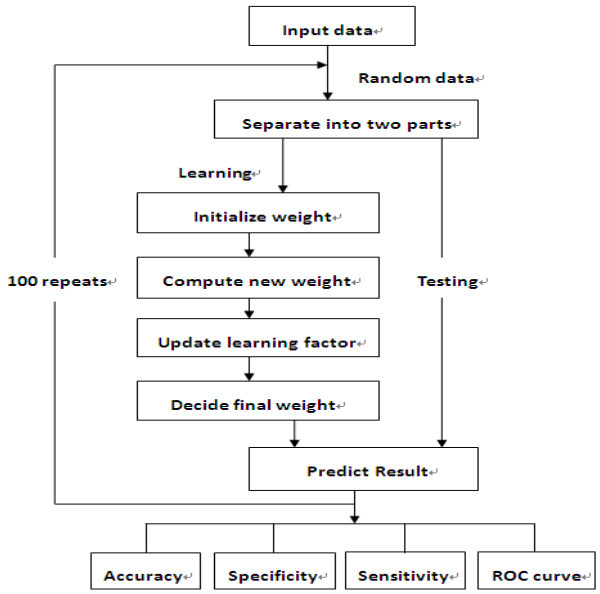

Figure 2. The computational steps of the LAV model: The method starts with 13 pieces of clinical data and proceeds to develop the LAV algorithm. After setting up the training models, the performance results can be calculated from the testing data.

3. Heart disease prediction system (HDPS)

In this paper, we develop a heart disease prediction system (HDPS) that can assist medical professionals in predicting heart disease status based on the clinical data of patients. The system generates prediction results using an artificial neural network (ANN) technique.

3.1. Flowchart of the HDPS

The system is developed in a C and C# environment. The HDPS converts codes written in C and integrates the results onto the C# interface. The design flowchart of HDPS is shown in Figure 3.

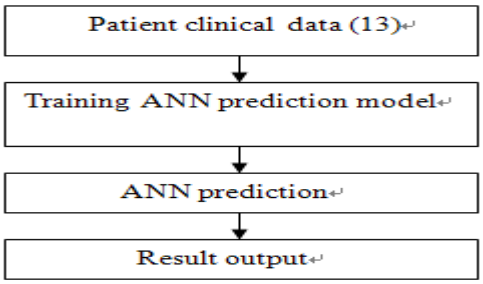

Figure 3. The design flowchart of the HDPS: It starts with 13 input clinical data and proceeds to develop ANN algorithm. After training model, it can generate the prediction results

3.2. The interface of the HDPS

The main functions of the HDPS interface shown as Figure 4 include:
(1) Input clinical data section: users input 13 pieces of clinical data.
(2) ROC curve display section: it displays the ROC curve of the HDPS.
(3) Prediction performance display section: it displays the HDPS prediction performance including accuracy, sensitivity, specificity and prediction result.
(4) Predict button: users click the button to get the result.
(5) Clear button: users click the button to clear the previous input
(6) Quit button: users click the button to leave this interface
(7) Prediction result: this text box shows the prediction result of the provided clinical data.

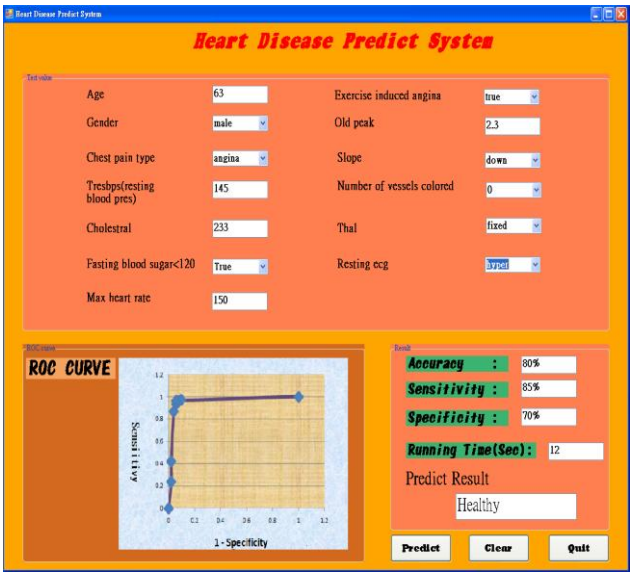

Figure 4. The HDPS interface: The HDPS interface consists of three panels: the Data input panel for the patient's clinical data, ROC curve display section, and prediction performance display section (execute time, accuracy, sensitivity, specificity, and predict result).

4. Results and discussion

We demonstrate here the usefulness of the proposed methodology to the clinical data of heart disease. Table 2 summarizes the results generated from our ANN algorithm.

Table 2. Classification performance of heart disease calculated from the 13 clinical features.

Accuracy (%)	Sensitivity (%)	Specificity (%)
80±5	85±5	70±5

The accuracy of classification is near 80% as well as 85% sensitivity and 70% specificity. To confirm the goodness of this model, a ROC curve is also displayed in Figure 4.

In the HDPS system, the users first need to input 13 pieces of clinical data. A default training model is stored in the database. After the users click the "Predict" button, the system will display the results. Figure 4 shows an example of a "Health" using the ANN algorithm.

5. Conclusion

In this paper, we develop a heart disease prediction system that can assist medical professionals in evaluating a patient's heart disease based on the clinical data of the patient. Our approaches include three steps. Firstly, we select 13 important clinical features, i.e., age, sex, chest pain type, trestbps, cholesterol, fasting blood sugar, resting ecg, max heart rate, exercise induced angina, old peak, slope, number of vessels colored, and thal. Secondly, we develop an artificial neural network algorithm for classifying heart disease based on these clinical features. The accuracy of prediction is near 80%. Finally, we develop a user-friendly heart disease predict system (HDPS) that generates prediction results using artificial neural network (ANN) techniques. The HDPS system is a computer-aided system developed from C and C# environment. Hopefully, this system can be used in the classification of heart disease.

Acknowledgements

The authors thank the National Science Council for its financial support for project NSC 100-2221-E-320-001 and Tzu Chi Foundation for its financial support for project TCRPP100011

References

[1] Wu R, Peters W, Morgan MW. The next generation clinical decision support: linking evidence to best practice. J Healthc Inf Manag, 2002;16:50-5.
[2] Thuraisingham BM. A Primer for Understanding and applying data mining. IT Professional 2000;1:28-31.
[3] Rajkumar A, Reena GS. Diagnosis of heart disease using datamining algorithm. Global Journal of Computer Science and Technology 2010;10:38-43.
[4] Anbarasi M, Anupriya E, Iyengar NCHSN. Enhanced prediction of heart Disease with feature subset selection using genetic algorithm. International Journal of Engineering Science and Technology 2010;2:5370-76.
[5] Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques. International Journal of Computer Science and Network Security 2008;8:343-50.
[6] Baxt WG. Application of artificial neural networks to clinical medicine. Lancet 1995;346:1135-38.
[7] UCI Machine Learning Repository [homepage on the Internet]. Arlington: The Association; 2006 [updated 1996 Dec 3; cited 2011 Feb 2]. Available from: http://archive.ics.uci.edu/ml/datasets/Heart+Disease

Address for correspondence.

Austin H Chen
Department of Medical Informatics, Tzu Chi University, No. 701, Sec. 3, Jhongyang Rd. Hualien City, Hualien County 97004, Taiwan
achen@mail.tcu.edu.tw