Accurate Prediction of Heart Disease Diagnosing Using Computation Method

[1]Hanumanthappa H, [2]Pundalik Chavan
[1]Assistant Professor, [2] Assistant Professor
[1]Computer Science & Engineering,
[2]Computer Science & Engineering, Bangalore, India

Abstract- In the last decades, the cardiovascular disease is major because, it leads to death. The data mining techniques can help us to predict the heart disease with reduced number of attributes these data mining techniques are used to physicians in medical fields. The aim of the study is used to predict the diagnosis of heart disease.

Index Terms-WAC,ANOVA,Logistic Regression,Levenberg–Marquardt, Scaled conjugate gradient, Pola-Ribiere conjugate gradient.

I.INTRODUCTION

1.1 Overview of heart

In the human body heart is one of the strong muscle. The human heart beats is 100000 times per day. The blood is pumped by heart for every heart beat is 70 times per minute. When we do exercise or in some excitement the heart beat increased as twice. Circulation process is done by the heart .During this process through the liver, some unused products can be removed. From 1970's many patients are suffering from cardiovascular disease.

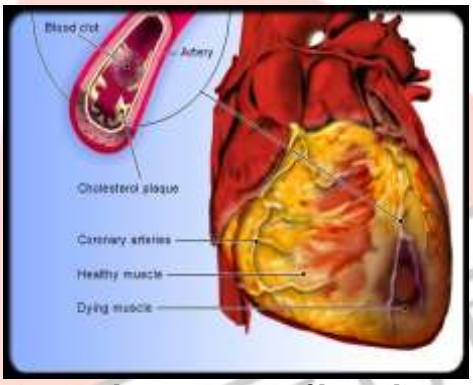

Figure 1 shows structure of heart disease

1.2 *Symptoms of cardiovascular disease*

- Chest pain.
- Pain in the neck or back.
- Physically Weakness, feeling sick to your stomach.
- Discomfort in the arms or shoulder.
- Dysphonia

1.3 Types of heart disease

***1.3.1 Coronary artery disease*:** The heart muscle supplies blood through blood vessels, during this process some diseases will occur to blood vessels. Some risk factors are diabetes, blood cholesterol and blood pressure is high.

***1.3.2 Cerebrovascular accident*:** Insufficient blood supply to the brain leads to brain attack. Major risk factors are physical inactive, unhealthy diet.

1.3.3 Congenital heart defect: the heart and blood vessels structures are defect present at birth caused by blood relationships between parents and genetics.example:holes in the heart, abnormal chambers and valves.

1.3.4 Rheumatic heart disease: the heart muscle and valves are damage from fever called rheumatic fever; this fever is caused by streptococcal bacteria.

1.3.5 Deep venous thrombosis (DVT): The formation of Blood clots in the leg veins. It leads to swelling, redness of legs. Later it will move to heart by knocing the viens. Some Risk factors are obesity, cancer.

The proposed methods may give more accuracy in prediction of heart disease than the existing systems and treatment may be done faster and effectively.

The aim of study is to Prediction of heart disease with reduced number of attributes. Physical sufferings of heart disease patients were more. Treatment and prediction was slow.

II.PREPARE YOUR PAPER BEFORE STYLING

II.I Data section

This dataset is extracted from Cleveland heart disease database. Some major factors for Cardiovascular disease, such as

Table 1 shows the different risk factors for heart disease.

Independent variables	Dependent variables
(1)Age, (2)Sex, (3)Chest pain type, (4)Resting blood pressure, (5)Serum cholestoroal in mg\dl, (6)Fasting blood sugar>120 mg\dl, (7)Restung electrocardiographic results, (8)Maximum heart rate achieved, (9)Exercise induced angina, (10)Old peak= ST depression induced by exercise relative to rest, (11)The slope of the peak exercise ST segment, (12)Number of major vessels ((0-3) colored By flourosopy, (13)thal.	(1) Healthy, (2) Patient.

In risk factors some factors can be changeable for example: blood cholesterol and if blood pressure is high means we can reduce it. Sex, age and family history are unchangeable risk factors.

Figure 2 shows Block diagram of heart disease

III. Existing Methods

III.I Artificial Neural Networks:

Artificial neural network can be invented by Dr.Robert Hecht-Nielsen. These ANN's are a system of interconnected nodes or neurons, this artificial neural network can compute the values from given inputs.

Architecture of neural network :

I. Feed-forward networks:

In this network the signals are travelling from input to output in only one direction. There is no
Back propagation method used in this method.

II. Feedback networks:

In feedback networks, the signals are travelling in both directions, it means from input to output and output to input until they reach certain condition. It uses back propagation method for solving the problems.

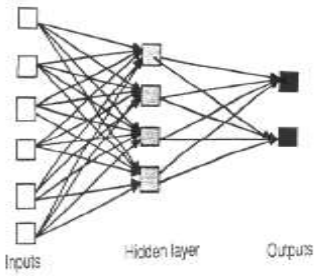

Fig 3 shows simple feed forward network

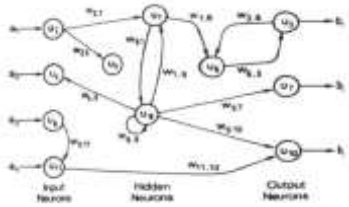

Fig 4 Shows complicated network

III. Network layers:

The artificial neural network having three layers or units. These all layers are interconnected each other from input layer to hidden layer and from hidden layer to output layer.

IV. Perceptrons:

Perceptrons are invented by Frank Rosen blatt, it uses weighted inputs with neurons and it will extract the features from input images. It will give idea of mammalian visual system. It can be used in pattern recognition

3.1.1 Applications :

I. Medicine field:

1. Diagnosis of heart disease
2. Instant Physician

II. Neural networks in business.

1. Marketing
2. Credit Evaluation

3.1.2 Advantages

1. ***Adaptive learning:***
   Neural networks accepts the data, whether it may be training data later it will learn how to do particular tasks for given data.
2. ***Self-Organisation:***
   In the learning time neural networks it creates one organisation for further usage.
3. ***Real Time Operation:*** ANN's having some hardware devices by using this capability it will compute real time operations.
4. ***Fault Tolerance via Redundant Information Coding:***
   If any network damage means it is having the capability to provide performance Degration.

3.2 *Levenberg-marquardt :*

The Levenberg-marquardt is used to solves the nonlinear least squares minimization problems

$$f(x) = \frac{1}{2}\sum_{j=1}^{m} r_j^2(x)$$

where

$x = (x_1, x_2, \ldots x_n)$ = vector

r_j = function from R^n to R.

r_j = residuals

Assumption =m >= n.

For non linear case :

$$\nabla f(x) = \sum_{j=1}^{m} r_j(x)\nabla r_j(x) = J(x)^T r(x)$$

$$\nabla^2 f(x) = J(x)^T J(x) + \sum_{j=1}^{m} r_j(x)\nabla^2 r_j(x)$$

3.2.1 Minimization methods

1. Newton	$H\,\delta x = -g$
2. Gauss-Newton	$J>J\,\delta x = -g$
3. Gradient descent	$\lambda\delta x = -g$

3.2.2 Levenberg-Marquardt algorithm

I.Gauss-newton approximation is bad in the negative curvature.

II. the steepest-descent approximation method is good in negative curvature problem.

III. this method varying approximation between gauss-newton and steepest-descent depend on how good.

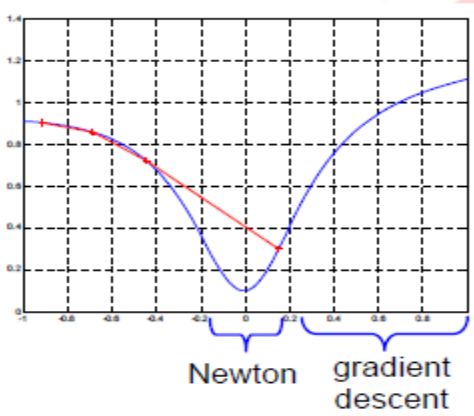

Figure 5 shows the variation between steepest-descent and Gauss-Newton steps

The method uses the modified Hessian $H(x, \lambda) = 2J>J+\lambda I$

When λ is small, H approximates the Gauss-Newton Hessian.

When λ is large, H is close to the identity, causing steepest-descent steps to be taken.

- $H(x, \lambda) = 2J>J+\lambda I$

1. Set $\lambda = 0.001$ (say)
2. Solve $\delta x = -H(x, \lambda)-1\ g$
3. If $f(xn+\delta x) > f(xn)$, increase λ ($\times 10$ say) and go to 2.
4. Otherwise, decrease λ ($\times 0.1$ say), let $xn+1 = xn+\delta x$, and go to 2.

3.3 Conjugate gradient methods

This method is used for minimizing quadratic function . THIS method is invented by Magnus Hestenes (1906-1991) and Eduard Stiefel (1909-1978).

3.3.1 Methods

I. *Scaled conjugate gradient :*

Optimization technique used by scaled conjugate algorithm. conjugate gradient methods can be used in numerical analysis. SCG uses the neural networks for second order information.

It uses O (N), N indicates the network and O Represents memory usage.

II.Polak-ribiere conjugate gradient:

This method is used to solve the nonlinear-least square minimization problems.

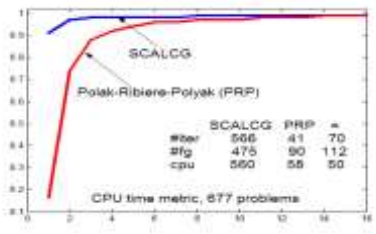

Fig 6 shows scaled and polak-ribiere conjugate garadient functions.

3.3.2 Advantages

1. Low storage method
2. CG only stores vector information
3. CG superlinear convergence for nice problems or when properly scaled
4. Great for solving QP sub problems

IV. Proposed methods:

4.1 Logistic regression:

Logistic regression is a type of regression used for classification model as well as prediction model. It is using binary variable (x) and used to predict the result of dependent variable(y).It is used to estimate parameters of qualitative response model. If the dependent variable is binary (i.e. only two variables) is called binary logistic regression. If the dependent variables is more than two variables is called multinomial logistic regression. It finds relation between the dependent variable and independent variables.

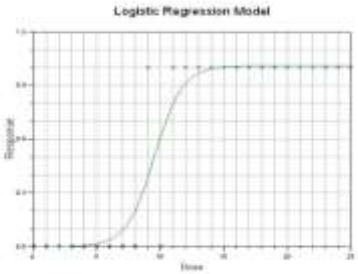

Figure 7 shows logistic regression model

Logistic function used in logistic regression takes values which lie between 0 and 1.

$$F(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}},$$

t is a linear combination of variables. Logistic function is written as:

$$F(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}.$$

Logistic function estimate result as probability of dependent variable either success or failure.

4.1.1. Binary logistic regression :

Binary logistic regression is used to predict the dependant variable by using independant variable.

Binary logistic regression equation.

y=b0+b1x1+b2x2+….bkxk=1

y=dependent variable

X=independent variable

B0=constant

b1=coefficient of variable x1

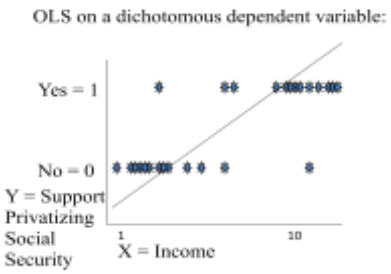

Figure 8 shows OLS on a dichotomous dependent variable

4.2 WAC (weighted associative classifier):

Weighted association rules can be used by weight associative classifier. The association rule can be Extracted from the data repository.

The steps are as follows:

1) For mining process the heart disease data can be pre-processed.
2) Every attribute is having its own weight. The weight ranges is 0 to 1.
   a) 0.9=high weight
   b) 0.1=low weight
3) The interesting pattern can be generated after pre-processing the data .

I. ***Attribute set weight*:** x is attribute weight. Average of attributes weights can be calculated.

II. ***Record weight/Tuple Weight:*** depending upon the weights of attributes tuple weight can be defined.

1.Attribute set weight :

$$w(x) = \frac{\sum_{i=1}^{n} \text{Weight of } a(i)}{\text{Number of attribute in } x}$$

2.Record weight :

$$w(x) = \frac{\sum_{i=1}^{n} \text{Weight of } a(i)}{\text{Number of attributes in record}}$$

4.3 ANOVA

Anova is invented by R.A. Fisher .the statistical test is provided by anova. It defines the comparing mean between two groups or variables.

4.3.1 Anova models :

i. Fixed Effects models

Applying one or more experiments to the subject to check whether the value of response variable changes is called fixed effect models.

ii. Random Effects models

The experiments to the subjects used are not fixed in random effect model. It considers the various levels of samples from large dataset.

iii. Mixed Effects models:

The experiment to the subject is both fixed and not fixed is called mixed effect model. The variance can be separated into two major components

1. ***Within groups*** –In groups individual differences occur
2. ***Between groups*** –To find differences' depend on treatment to one group is received.

I. One-factor completely randomized designs

Total SS = Treatment SS + Error SS

SS(Total) = SST + SSE

II. Randomized Block Designs

Total SS = Treatment SS + Block SS + Error SS

SS(Total) = SST + SSB + SSE

III. Two-FactorFactorial Experiments

- Total SS = Main effect SS Factor A + Main effect SS Factor B +AB Interaction SS + Error SS
- SS(Total) = SS(A) + SS (B) + SS (AB) + SSE

V. Discussion:

Heart disease can also named as silent killer disease.in order to cure the patients data mining techniques can be used. Our proposed method help to physicians to predict the heart disease with good accuracy.

VI. Conclusion:

The proposed method for using construction of prediction model for effective accuracy of heart disease. The 89.01% classification accuracy by using SAS base software 9.1.3 by using the proposed methods the accuracy may be increased to predict the diagnosis of heart disease and to evaluate the system using performance metric like specificity and sensitivity values.

References

[1]Anastasiadis, Aristoklis D., & Magoulas, George D. (2006). "Analysing the localisation sites of proteins through neural networks ensembles". Neural Computing and Applications, 15(3–4), 277–288.

[2]Mythili T.,Dev mukherji,nikita padalia,and abhiram naidu,"A Heart disease prediction model using SVM-decision trees-Logistic Regression(SDL)",International journal of computer applications,vol 68-no.16,April 2013

[3]S.J Gnanasoundhari,G.visalatchi,DR.M.balamurugan,"A survey on heart disease prediction system using data mining techniques",International journal of computer science and mobile applications,vol.2 Issue .2,Feb-2014

[4]Cheung,N.(2001).machine learning techniques for medical analysis.school of information technology and elestrical engineering ,B.sc . Thesis,university of queenland.

[5]Raalakshmi M,neelamegam ,Bharathi N and Shanmugasundaram kumaravel,"Effective diagnosis of heart disease using multilayer feed forward neural network using bac propogation algorithm and associative neural network",research journal of pharmaceutical,biological and chemical sciences,march-april-2014.