



# Effective diagnosis of heart disease through neural networks ensembles

Resul Das<sup>a,\*</sup>, Ibrahim Turkoglu<sup>b</sup>, Abdulkadir Sengur<sup>b</sup>

<sup>a</sup> Department of Informatics, Firat University, 23119 Elazig, Turkey

<sup>b</sup> Department of Electronics and Computer Science, Firat University, 23119 Elazig, Turkey

## ARTICLE INFO

### Keywords:

Heart disease  
SAS base software  
Neural networks  
Ensemble based model

## ABSTRACT

In the last decades, several tools and various methodologies have been proposed by the researchers for developing effective medical decision support systems. Moreover, new methodologies and new tools are continued to develop and represent day by day. Diagnosing of the heart disease is one of the important issue and many researchers investigated to develop intelligent medical decision support systems to improve the ability of the physicians. In this paper, we introduce a methodology which uses SAS base software 9.1.3 for diagnosing of the heart disease. A neural networks ensemble method is in the centre of the proposed system. This ensemble based methods creates new models by combining the posterior probabilities or the predicted values from multiple predecessor models. So, more effective models can be created. We performed experiments with the proposed tool. We obtained 89.01% classification accuracy from the experiments made on the data taken from Cleveland heart disease database. We also obtained 80.95% and 95.91% sensitivity and specificity values, respectively, in heart disease diagnosis.

© 2008 Elsevier Ltd. All rights reserved.

## 1. Introduction

Heart disease, which is usually called coronary artery disease (CAD), is a broad term that can refer to any condition that affects the heart (<http://yourtotalhealth.ivillage.com/heart-disease-fast-facts.html>, 2008). CAD is a chronic disease in which the coronary arteries gradually harden and narrow and the most common form of cardiovascular disease in the United States and the leading cause of heart attacks. Moreover, cardiovascular disease is the leading killer of American women, causing almost 500,000 deaths every year (<http://yourtotalhealth.ivillage.com/heart-disease-fast-facts.html>, 2008).

While many people with heart disease have symptoms such as chest pain and fatigue, as many as 50% have no symptoms until a heart attack occurs. According to the American heart association (AHA), CAD is the leading killer of American men and women, responsible for more than one of every five deaths in 2001 (<http://www.americanheart.org>, 2008). Many statistics show CAD as the leading cause of premature and permanent disability among American workers.

There are many risk factors associated with CAD. Some risk factors for CAD, such as your sex, age and family history, cannot be changed. Other risk factors for CAD that are related to lifestyle often can be changed (<http://yourtotalhealth.ivillage.com/heart-disease-fast-facts.html>, 2008). For example, smoking, high chole-

sterol, high blood pressure and physical inactivity are all risk factors for coronary artery disease that can be modified and in some cases eliminated with lifestyle changes and medication. Diabetes and obesity can sometimes be prevented when lifestyle changes are made early in life.

Having so many factors to analyze to diagnose the heart diseases, physicians generally make decisions by evaluating the current test results of the patients. The previous decisions made on other patients with the same condition are also examined by the physicians. These complex procedures are not easy when considering the number of factors that the physician has to evaluate. So, diagnosing the heart disease of a patient involves experience and highly skilled physicians.

Recent advances in the field of artificial intelligence have led to the emergence of expert systems for medical applications. Moreover, in the last few decades computational tools have been designed to improve the experiences and abilities of physicians for making decisions about their patients. Motivated by the need of such an expert system, in this study, we propose a method to efficiently diagnose the heart disease. The proposed system uses neural networks ensemble model. Ensemble based methods enable an increase in generalization performance by combining several individual neural networks train on the same task. We realized our proposal with SAS base software 9.1.3 (licence number: 291468). SAS enterprise miner streamlines the entire data mining process from data access to model assessment. It supports all necessary tasks within a single, integrated solution while providing the flexibility for efficient collaborations. We obtained 89.01% classification accuracy from the experiments made on the data taken from

\* Corresponding author.

E-mail address: [rdas@firat.edu.tr](mailto:rdas@firat.edu.tr) (R. Das).

Cleveland heart disease database. We also obtained 80.95% and 95.91% sensitivity and specificity values, respectively, in heart disease diagnosis.

The remaining of the paper is organized as follows. We present a brief overview on related work, dataset description, neural networks and ensemble based methods in the background section. Basic descriptions for this study can be found in this section. Methodology and implementation of the proposed system is described in Section 3. The components of the SAS base software are introduced briefly. The implementation constraints are also given in this section. Experimental results are discussed in Section 4. The SAS base software represents several statistical evaluation tests and different graphics for the users. Brief explanation can also be found in this section. In Section 5, we finally conclude our study.

## 2. Background

### 2.1. A brief overview on related works

So far a great variety of classification algorithms have been employed on Cleveland heart disease database and high classification accuracies have been reported in the last decades (<ftp://ftp.ics.uci.edu/pub/ma>, 2008; Cheung, 2001; Polat et al., 2005; Ozsen & Gunes, in press; <http://www.phys.uni.toruu>, 2008). Robert Detrano, who constructed the Cleveland heart disease database, used logistic regression algorithm and obtained 77.0% classification accuracy (<ftp://ftp.ics.uci.edu/pub/ma>, 2008). Newton Cheung utilized C4.5, Naïve Bayes, BNND and BNNF algorithms and reached the classification accuracies of 81.11%, 81.48%, 81.11% and 80.96%, respectively (Cheung, 2001). Polat et al. proposed a method that uses artificial immune system (AIS) and obtained 84.5% classification accuracy (Polat et al., 2005). Then, a similar method was used by Ozsen and Gunes and 87.0% classification accuracy was reported (Ozsen & Gunes, in press). Moreover, more results were reported by using ToolDiag, RA and WEKA tools in the following reference (<http://www.phys.uni.toruu>, 2008). As can be seen from the related reference the reached higher classification accuracy was 77% when these tools were used.

### 2.2. Database description

The heart disease database was taken from UCI machine learning repository (<ftp://ftp.ics.uci.edu/pub/ma>, 2008). The Cleveland heart disease data was obtained from V.A. Medical Center, Long Beach and Cleveland Clinic Foundation from Dr. Robert Detrano. The database contains 303 samples of which 297 are complete samples and six are samples with missing attributes. Originally, the database has 76 raw attributes. However, all the published experiments only refer to 13 of them, and are listed as follows:

- (1) age,
- (2) sex,
- (3) chest pain type (four values),
- (4) resting blood pressure,
- (5) serum cholestoral in mg/dl,
- (6) fasting blood sugar >120 mg/dl,
- (7) resting electrocardiographic results (values 0, 1 and 2),
- (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: 3 = normal; 6 = fixed defect and 7 = reversable defect.

### 2.3. Neural networks

Artificial neural networks were originally developed by researchers who were trying to mimic the neurophysiology of the human brain (Bishop, 1995). By combining many simple computing elements (neurons or units) into a highly interconnected system, a complex phenomenon such as intelligence was indented to produce. A schematic diagram for an artificial neuron model is shown in Fig. 1. Nowadays, neural network researchers have incorporated methods from statistics and numerical analysis into their networks. More specifically, feedforward neural networks are a class of flexible nonlinear regression, discriminant and data reduction models. By detecting complex nonlinear relationships in data, neural networks can help to make predictions about real-world problems (Sengur et al., 2007).

The neural network node provides a variety of feedforward networks that are commonly called backpropagation networks. Backpropagation refers to the method for computing the error gradient for a feedforward network, a straightforward application of the chain rule of elementary calculus (Hanbay et al., 2008). By extension, backpropagation refers to various training methods that use backpropagation to compute the gradient. By further extension, a backpropagation network is a feedforward network trained by any of various gradient descent techniques. There are numerous algorithms available for training neural network models; most of them can be viewed as a straightforward application of optimization theory and statistical estimation. Most of the algorithms used in training artificial neural networks are employing some form of gradient descent. This is done by simply taking the derivative of the cost function with respect to the network parameters and then changing those parameters in a gradient-related direction. The most popular of them is the back propagation algorithm, which has different variants. Standard backpropagation is a gradient descent algorithm. It is very difficult to know which training algorithm will be the fastest for a given problem, and the best one is usually chosen by trial and error. An ANN with a back propagation algorithm learns by changing the connection weights, and these changes are stored as knowledge.

Neural networks are trained by experience, when applied an unknown input to the network it can generalize from past experiences and product a new result. The output of the neuron net is given by the following equation:

$$y(t+1) = a\left(\sum_{j=1}^m w_{ij}x_j(t) - \theta_i\right)$$

$$f_1 \triangleq \text{net}_i = \sum_{j=1}^m w_{ij}x_j - \theta_i \quad (1)$$

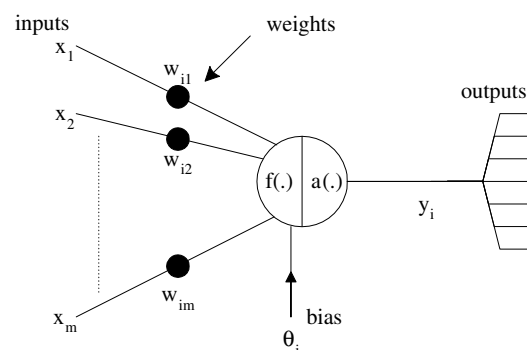


Fig. 1. Artificial neuron model (Hanbay et al., 2008).

where  $X = (X_1, X_2, \dots, X_m)$  represent the  $m$  input applied to the neuron,  $W_i$  represent the weights for input  $X_i$ ,  $\theta_i$  is a bias value and  $a(\cdot)$  is activation function. There are many kind of activation function. Usually nonlinear activation functions such as sigmoid, step are used.

#### 2.4. Ensemble based methods

The ensemble based methods creates new models by combining the posterior probabilities (for class targets) or the predicted values (for interval targets) from multiple predecessor models. The new model is then used to score new data. Thus, this new model obtains an increase in generalization performance by combining several individual models trained on the same task (<http://support.sas.com/d, 2008>). A schematic illustration of ensemble model can be seen in Fig. 2.

The creation of an ensemble approach is often divided into two steps; (1) generate individual members, and (2) appropriately combine individual members' outputs to constitute a new output. The basic method for forming ensemble based methods is to train each member model using random parameters. In the neural network context, these methods include techniques for training with different network topologies, different initial weights, different learning parameters, and learning different portions of the training set (Anastasiadis & Magoulas, 2006).

Moreover, methods for creating ensembles focus on creating classifiers that disagree on their decisions. In general terms, these methods alter the training process in an attempt to produce classifiers that will generate different classifications.

### 3. The proposed methodology and implementation

The proposed methodology, which is illustrated in Fig. 3, is implemented with the SAS base software 9.1.3. SAS enterprise

miner streamlines the entire data mining process from data access to model assessment. It supports all necessary tasks within a single, integrated solution while providing the flexibility for efficient collaborations (<http://support.sas.com/d, 2008>). SAS enterprise miner is designed for data miners, marketing analysts, database marketers, risk analysts, fraud investigators, business managers, engineers and scientists who play strategic roles in identifying and solving critical business or research issues (<http://support.sas.com/d, 2008>).

SAS base software 9.1.3 includes two different programs. These programs are called SAS enterprise guide 4.3 and SAS enterprise miner 5.2. While SAS enterprise guide 4.3 program was used for data preprocessing, SAS enterprise miner 5.2 program was used to analyze and recognize the heart disease by combining several neural networks with ensemble node. As can be seen from Fig. 3, the performed system is composed of five components. The brief description of each component is given in the following.

*Heart database* component holds the features that are used to characterize healthy persons and patients. As it was mentioned earlier, the database composed of 14 columns and 297 rows. About 13 columns indicate the features and one column indicates the class labels. Heart disease data set was transferred by using SAS enterprise guide 4.3 to analyze in SAS enterprise miner 5.2.

*Data partition* component was used to partition the input data into train and validation data sets. Partitioning provides mutually exclusive data sets. Two or more mutually exclusive data sets share no observations with each other. Partitioning the input data reduces the computation time of preliminary modeling runs.

*Variable selection* component was used in reducing the number of inputs by setting the status of the input variables that are not related to the target as rejected. Although rejected variables were passed to subsequent nodes in the process flow, these variables were not used as model inputs by a successor modeling node.

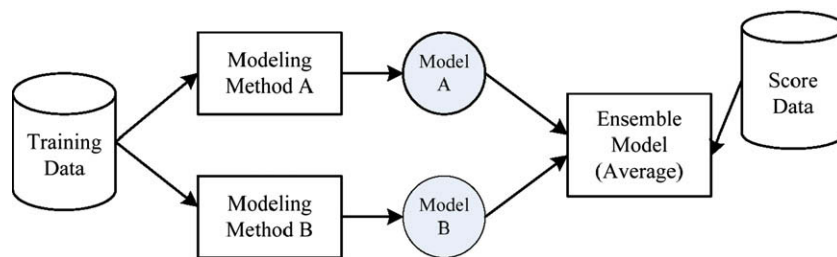


Fig. 2. A schematic illustration of ensemble model (<http://support.sas.com/d, 2008>).

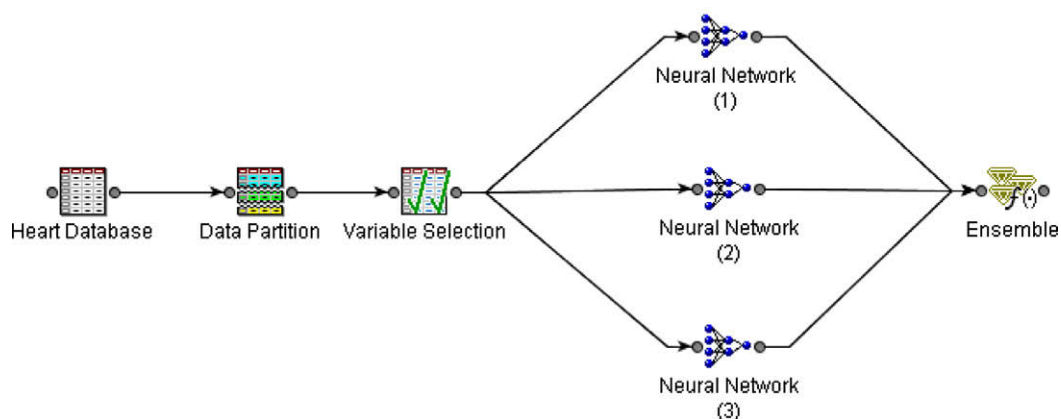


Fig. 3. Proposed system for heart disease recognition.

Neural networks block component was used to classify the feature space. Three independent neural networks models were used to construct this component. There are many types of neural networks architectures; however, multi-layer feedforward neural network is the most widely used for prediction. A multi-layer feedforward neural network typically has an input layer, an output layer and one or more hidden layers. In multi-layer feedforward networks, neurons are arranged in layers and there is a connection among the neurons of other layers. The inputs are applied to the input layer the output layer contributes to the output directly. Other layers between input and output layers are called hidden layers. Inputs are propagated in gradually modified form in the forward direction, finally reaching the output layer. The backpropagation learning algorithm has been used in the feedforward, single hidden layer neural network. The variants of the algorithm used in the study are the Levenberg–Marquardt (LM), scaled conjugate gradient (SCG) and Pola–Ribiere conjugate gradient (CGP) algorithms. A tangent sigmoid transfer function has been used for both the hidden layer and the output layer. We used 14 neurons in the hidden layer. The initial weights were chosen randomly.

Ensemble component was used to create new models by combining the posterior probabilities (for class targets) or the predicted values (for interval targets) from multiple predecessor models. The new model is then used to score new data.

#### 4. Experimental results and discussion

In this study, there were two diagnosis classes: healthy and a patient who is subject to possible heart disease. As it was noted earlier in the background section, several researchers proposed various methods for diagnosing the heart disease. The reported accuracies vary between 50% and 87%. The database contains 303 samples of which 297 are complete samples and six are samples with missing attributes. While 70% of the heart disease database was used for training the neural networks ensemble model, the rest of the heart disease database (30%) was used for validation of the proposed system. As we mentioned earlier, three different independent neural networks models were combined for constructing the ensemble model. We also increased the number of neural networks node in the ensemble model but no performance improvement was obtained.

Classification results of the system were displayed by using a confusion matrix. In a confusion matrix, each cell contains the raw number of exemplars classified for the corresponding combination of desired and actual network outputs. Table 1 gives the confusion matrix showing the classification results of this network.

SAS enterprise miner 5.2 represents several performance evaluation tests for the users. Here, we only examined these test methods briefly; meanwhile we represented our experimental results and the related displays and tables. Moreover a comparison table was constructed and can be found in this section. SAS enterprise miner 5.2 produces the following tables and plots in its results window: fit statistics table, classification chart, score rankings overlay chart, score rankings matrix chart, score distribution chart and ROC.

The fit statistics table in the ensemble results window displays the several statistics for the train and validates data sets. This table is composed of 5 columns and 13 rows, respectively. The columns of the table are called as the target, fit statistics, statistical labels, train and validation, respectively. The rows indicated the each statistical evaluation method. As can be seen from Table 2, the average profit which can be called as accuracy was 86.4% and 89.011% for training and validation data sets, respectively. Another observation from Table 2 is that, while 10 of the validation data set were wrongly classified, 81 of the validation data set were correctly classified.

**Table 1**  
Classification table

| Target | Outcome | Target percentage (%) | Outcome percentage (%) | Count | Total percentage (%) |
|--------|---------|-----------------------|------------------------|-------|----------------------|
| 0      | 0       | 85.4545               | 95.9184                | 47    | 51.6484              |
| 1      | 0       | 14.5455               | 19.0476                | 8     | 8.7912               |
| 0      | 1       | 5.5556                | 4.0816                 | 2     | 2.1978               |
| 1      | 1       | 94.4444               | 80.9524                | 34    | 37.3626              |

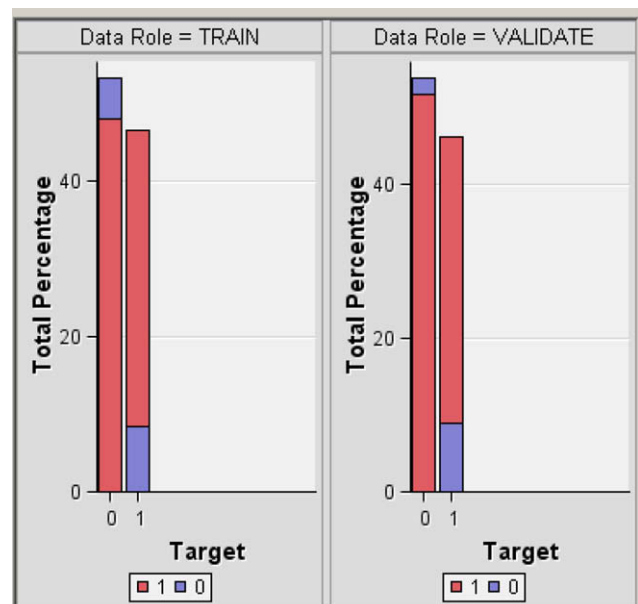
**Table 2**  
Fit statistics for ensemble model

| Target | Fit statistics | Statistics label                | Train    | Validation |
|--------|----------------|---------------------------------|----------|------------|
| sinif  | _ERR_          | Error function                  | 124.2403 | 78.6215    |
| sinif  | _SSE_          | Sum of squared errors           | 39.16356 | 19.27484   |
| sinif  | _MAX_          | Maximum absolute error          | 0.927191 | 0.999591   |
| sinif  | _DIV_          | Divisor for ASE                 | 412      | 182        |
| sinif  | _NOBS_         | Sum of frequencies              | 206      | 91         |
| sinif  | _WRONG_        | Number of wrong classifications | 28       | 10         |
| sinif  | _DISF_         | Frequency of classified cases   | 206      | 91         |
| sinif  | _MISC_         | Misclassification rate          | 0.135922 | 0.10989    |
| sinif  | _PROF_         | Total profit                    | 178      | 81         |
| sinif  | _APROF_        | Average profit                  | 0.864078 | 0.89011    |
| sinif  | _ASE_          | Average squared error           | 0.095057 | 0.105906   |
| sinif  | _RASE_         | Root average squared error      | 0.308313 | 0.325432   |
| sinif  | _AVERR_        | Average error function          | 0.301554 | 0.431986   |

Another evaluation display is the classification table chart which displays a stacked bar chart of the classification results for a categorical target variable. The following display shows an example of the classification table chart (Fig. 4).

The horizontal axis displays the target levels that observations actually belong to. The color of the stacked bars identifies the target levels that observations are classified into. The height of the stacked bars represents the percentage of total observations.

In the score rankings chart, several statistics for each group of observations are plotted on the vertical axis. For a binary target, all observations in the scored data set are sorted by the posterior probabilities of the event level in descending order. For a nominal



**Fig. 4.** Classification table chart.

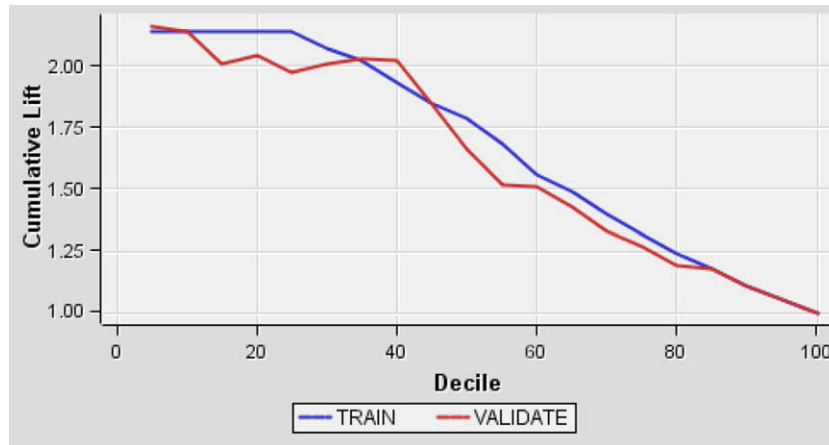


Fig. 5. Score ranking chart.

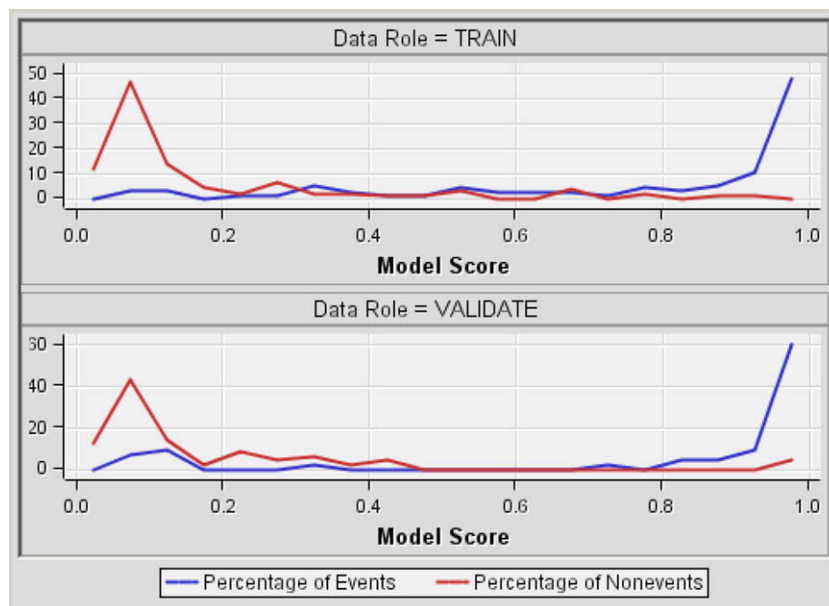


Fig. 6. The score distribution chart.

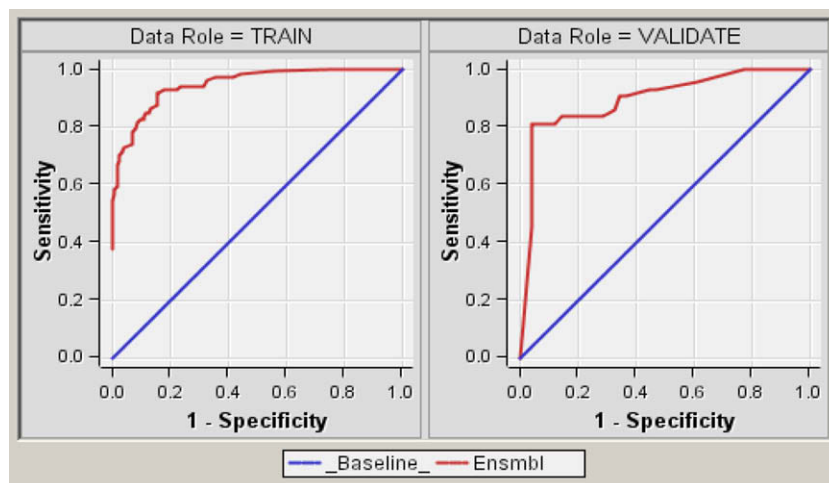


Fig. 7. ROC chart.



**Table 3**

Classification accuracies obtained with our proposed system and other classifiers from literature

| Author (year)                  | Method                      | Accuracy (%) |
|--------------------------------|-----------------------------|--------------|
| ToolDiag                       | IB1-4                       | 50.00        |
| WEKA, RA                       | InductH                     | 58.50        |
| ToolDiag, RA                   | RBF                         | 60.00        |
| WEKA, RA                       | FOIL                        | 64.00        |
| ToolDiag, RA                   | MLP + BP                    | 65.60        |
| WEKA, RA                       | T2                          | 68.10        |
| WEKA, RA                       | 1R                          | 71.40        |
| WEKA, RA                       | IB1c                        | 74.00        |
| WEKA, RA                       | K                           | 76.70        |
| Robert Detrano                 | Logistic regression         | 77.00        |
| Newton Cheung (2001)           | C4.5                        | 81.11        |
| Newton Cheung (2001)           | Naive Bayes                 | 81.48        |
| Newton Cheung (2001)           | BNND                        | 81.11        |
| Newton Cheung (2001)           | BNNF                        | 80.96        |
| WEKA, RA                       | Naive Bayes                 | 83.60        |
| Polat et al. (2005)            | AIRS                        | 84.50        |
| Polat et al. (2006)            | Fuzzy-AIRS-Knn based system | 87.00        |
| Our proposal – SAS base (2008) | Neural networks ensemble    | 89.01        |

or ordinal target, observations are sorted from highest expected profit to lowest expected profit. Then the sorted observations are grouped into deciles based on the group bin property and observations in a group are used to calculate the statistics that are plotted in deciles charts. By default, the horizontal axis of a score rankings chart displays the deciles (groups) of the observations. The vertical axis displays the following values, and their mean, minimum and maximum. The related figure can be seen in Fig. 5.

The score distribution chart plots the proportions of events, nonevents and other values on the vertical axis. The values on the horizontal axis represent the model score of a bin. The model score depends on the prediction of the target and the number of buckets used. For categorical targets, observations are grouped into bins, based on the posterior probabilities of the event level and the number of buckets. For interval targets, observations are grouped into bins, based on the actual predicted values of the target. The score distribution chart of a useful model shows a higher percentage of events for higher model score and a higher percentage of nonevents for lower model scores. Fig. 6 shows the score distribution chart for the heart disease detection system.

The receiver operating characteristic (ROC) chart is a graphical display that gives the measure of the predictive accuracy of a logistic model. The chart displays the sensitivity and specificity. Sensitivity is a measure of accuracy for predicting events that is equal to the true positive/total actual positive. Specificity is a measure of accuracy for predicting nonevents that is equal to the true negative/total actual negative of a classifier for a range of cutoffs. Fig. 7 displays the ROC chart for a binary target.

We compared our results with the previous results reported by earlier methods. Table 3 gives the classification accuracies of our method and the previous proposed methods.

## 5. Conclusions

Up to now, several studies have been reported focusing on heart disease diagnosis (<http://www.phys.uni.toruu, 2008>). These studies applied different methods to the given problem and achieved high classification accuracies using the dataset taken from UCI machine learning repository. In this study, SAS enterprise miner 5.2 was used to construct a neural networks ensemble based methodology for diagnosing of the heart disease. Experiments were conducted on the heart disease dataset to diagnose heart disease in a fully automatic manner. Three independent neural networks models were used to construct the ensemble model. The number of neural networks node in the ensemble model was also increased but no performance improvement was obtained. The experimental results gained 89.01% classification accuracy, 80.95% sensitivity and 95.91% specificity values for heart disease diagnosis.

SAS enterprise miner 5.2 supports all necessary tasks within a single, integrated solution while providing the flexibility for efficient collaborations. It also gives opportunities to the user to deal with various performance evaluation test methods. This allows the user to evaluate their system performance from many different points of views. Thus, SAS base software can be used in many machine intelligence applications.

## Acknowledgement

The authors gratefully acknowledge the financial support from the Scientific Research Projects Administration Unit of Firat University for this study performed under project with Grant No. 1526.

## References

- <<http://yourtotalhealth.ivillage.com/heart-disease-fast-facts.html>> (last accessed: February 13, 2008.).
- <<http://www.americanheart.org>> (last accessed: February 13, 2008).
- <ftp://ftp.ics.uci.edu/pub/machine-learning-databases> (last accessed: February 5, 2008).
- Cheung, N. (2001). *Machine learning techniques for medical analysis*. School of Information Technology and Electrical Engineering, B.Sc. Thesis, University of Queensland.
- Polat, K., Sahan, S., Kodaz, H., Günes, S. (2005). A new classification method to diagnosis heart disease: Supervised artificial immune system (AIRS). In *Proceedings of the Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN)*.
- Ozsen, S., Gunes, S. (in press). Attribute weighting via genetic algorithms for attribute weighted artificial immune system (AWAIS) and its application to heart disease and liver disorders problems. *Expert Systems with Applications*.
- <<http://www.phys.uni.toruu.pl/kmk/projects/datasets.html>> (last accessed: February 15, 2008).
- Bishop, Christopher M. (1995). *Neural networks for pattern recognition*. Oxford University Press.
- Sengur, A., Turkoglu, I., & Ince, M. C. (2007). Wavelet packet neural networks for texture classification. *Expert Systems with Applications*, 32(2).
- Hanbay, D., Turkoglu, I., & Demir, Y. (2008). An expert system based on wavelet decomposition and neural network for modeling Chua's circuit. *Expert Systems with Applications*, 34(4), 2278–2283.
- <<http://support.sas.com/documentation>> (last accessed: January 20, 2008).
- 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.