

Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm

Latha Parthiban and R.Subramanian

Abstract—Heart disease (HD) is a major cause of morbidity and mortality in the modern society. Medical diagnosis is an important but complicated task that should be performed accurately and efficiently and its automation would be very useful. All doctors are unfortunately not equally skilled in every sub specialty and they are in many places a scarce resource. A system for automated medical diagnosis would enhance medical care and reduce costs. In this paper, a new approach based on coactive neuro-fuzzy inference system (CANFIS) was presented for prediction of heart disease. The proposed CANFIS model combined the neural network adaptive capabilities and the fuzzy logic qualitative approach which is then integrated with genetic algorithm to diagnose the presence of the disease. The performances of the CANFIS model were evaluated in terms of training performances and classification accuracies and the results showed that the proposed CANFIS model has great potential in predicting the heart disease.

Keywords—CANFIS, Genetic Algorithms (GA), Heart disease, Membership Function (MF).

I. INTRODUCTION

A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Clinical decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. Wu, et al proposed that integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome [2].

Most hospitals today employ some sort of hospital information systems to manage their healthcare or patient data [3]. Unfortunately, these data are rarely used to support clinical decision making. The main objective of this research is to develop a prototype Intelligent Heart Disease Prediction System with CANFIS and genetic algorithm using historical heart disease databases to make intelligent clinical decisions which traditional decision support systems cannot.

Latha Parthiban is with the Department of Computer Science, Pondicherry University, Puducherry 605 014, India (e-mail: lathaparthiban@yahoo.com).

R.Subramanian is with the Department of Computer Science, Pondicherry University, Puducherry 605 014, India (e-mail: subbur@yahoo.com).

The cost of management of HD is a significant economic burden and so prevention of heart disease is very important step in the management. Prevention of HD can be approached in many ways including health promotion campaigns, specific protection strategies, life style modification programs, early detection and good control of risk factors and constant vigilance of emerging risk factors. This paper is organized as follows: Section 2 describes the database, Section 3 presents our proposed diagnostic system using CANFIS and genetic optimization. Section 4 presents the simulation results and the last section 5 our conclusions.

II. DESCRIPTION OF DATABASE

The publicly available Cleveland heart-disease database consists of 303 cases where the disorder is one of four types of heart-disease or its absence. There are fourteen variables as shown in Table I.

TABLE I
DESCRIPTION OF CLEVELAND HEART DISEASE DATABASE

| Observation | Description | Values |
|-------------|----------------|--|
| 1 | age | Age in years |
| 2 | sex | Sex of subject |
| 3 | CP | Chest pain |
| 4 | trestbps | Resting blood pressure |
| 5 | chol | Serum cholesterol |
| 6 | fb | Fasting blood sugar |
| 7 | restecg | Resting electrocardiographic result |
| 8 | thalach | Maximum heart rate achieved |
| 9 | exang | Exercise induced angina |
| 10 | oldpeak | ST depression induced by exercise relative to rest |
| 11 | slope | Slope of peak exercise ST segment |
| 12 | ca | Number major vessels colored by fluoroscopy |
| 13 | thal | Defect type |
| 14 | num (disorder) | Heart Disease |

III. DIAGONISTIC SYSTEM WITH CANFIS

The CANFIS model integrates adaptable fuzzy inputs with a modular neural network to rapidly and accurately approximate complex functions. Fuzzy inference systems are also valuable, as they combine the explanatory nature of rules (MFs) with the power of neural networks. These kinds of networks solve problems more efficiently than neural

networks when the underlying function to model is highly variable or locally extreme [4].

The fundamental component of CANFIS is a fuzzy axon, which applies membership functions to the inputs. The output of a fuzzy axon is computed using the following formula:

$$f(x, w) = \min \forall_i (MF(x_i, w_{ij})), \quad (1)$$

where i = input index, j = output index, x_i = input i , w_{ij} =weights (MF parameters) corresponding to the j th MF of input i and MF=membership function of the particular subclass of the fuzzy axon. This system can be viewed as a special three-layer feed forward neural network. The first layer represents input variables, the middle (hidden) layer represents fuzzy rules and the third layer represents output variables. The CANFIS architecture used in this study is shown in Fig 2.

A. CANFIS Architecture

Consider a CANFIS structure with n inputs and one output. For model initialize, a common rule set with n inputs and m IF-THEN rules as follows [5]

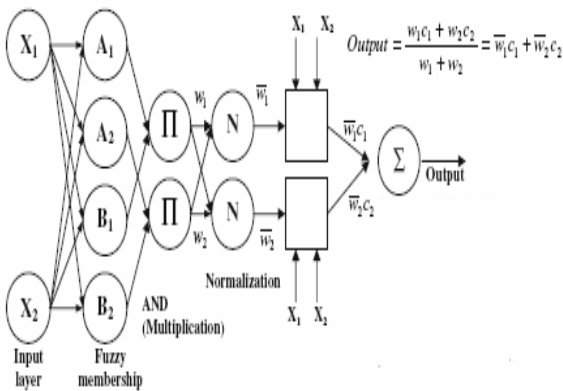


Fig. 2 A prototype two-input one-output CANFIS network and output calculation

- Rule 1: If z_1 is A_{11} and z_2 is A_{12} . . . and z_n is A_{1n}
then $u_1 = p_{11}z_1 + p_{12}z_2 + \dots + p_{1n}z_n + q_1$
- Rule 2: If z_1 is A_{21} and z_2 is A_{22} . . . and z_n is A_{2n}
then $u_2 = p_{21}z_1 + p_{22}z_2 + \dots + p_{2n}z_n + q_2$
- Rule m : If z_1 is A_{m1} and z_2 is A_{m2} . . . and z_n is A_{mn}
then $u_m = p_{m1}z_1 + p_{m2}z_2 + \dots + p_{mn}z_n + q_m$

Layers in CANFIS structure can be adaptive or fixed and their functions are:

Layer 1 (Premise Parameters): Every node in this layer is a complex-valued membership function (μ_{ij}) with a node function: $O_{1,ij} = \min(\mu_{A_{ij}}(z_i), \mu_{A_{ij}}(z_j))$ for $(1 \leq i \leq n, 1 \leq j \leq m)$. (2)

Each node in layer 1 is the membership grade of a fuzzy set (A_{ij}) and specifies the degree to which the given input belongs to one of the fuzzy sets.

Layer 2 (Firing Strength): Every node in this layer is product of all the incoming signals. This layer receives input in the form of the product of all the output pairs from the first layer: $O_{2,j} = w_j = \mu_{A_{i1}}(z_1)\mu_{A_{i2}}(z_2), \dots, \mu_{A_{in}}(z_n)$ for $(1 \leq i \leq m)$ (3)

Layer 3 (Normalized Firing Strength): Every node in this layer calculates rational firing strength:

$$O_{3,j} = \bar{w}_j = \frac{w_j}{\sum_{j=1}^m w_j} \quad \text{for } (1 \leq j \leq m). \quad (4)$$

Layer 4 (Consequence Parameters): Every node in this layer is multiplication of Normalized Firing Strength from the third layer and output of neural network:

$$O_{4,j} = \bar{w}_j u_j = \bar{w}_j (P_{j1}Z_1 + P_{j2}Z_2 + \dots + P_{jn}Z_n + q_j) \quad \text{for } (1 \leq j \leq m) \quad (5)$$

Layer 5 (Overall Output): The node here computes the output of CANFIS network:

$$O_{5,1} = \sum \bar{w}_j u_j \quad (6)$$

The bell fuzzy axon used in this study is a type of fuzzy axon that uses a bell-shaped curve as its membership function. Each MF takes three parameters stored in the weight vector of the bell fuzzy axon (Eq. 7):

$$MF(x, w) = \frac{1}{1 + \left| \frac{x - w_2}{w_b} \right|^{2m}} \quad (7)$$

where x =input and w =weight of the bell fuzzy axon. Fuzzy axons are valuable because their MF can be modified through back propagation during network training to expedite the convergence.

The second major component of CANFIS is a modular network that applies functional rules to the inputs. The number of modular networks matches the number of network outputs and processing elements in each network corresponding to the number of MFs. Two fuzzy structures are mainly used: the Tsukamoto model and the Sugeno (TSK) model. Finally, a combiner is used to apply the MF outputs to the modular network outputs. The combined outputs are then channeled through a final output layer, and the error is back propagated to both the MF and the modular network [6]

B. Genetic Optimization

In order to improve the learning of the CANFIS, quicker training and enhance its performance, we use genetic algorithms to search for the best number of MF for each input, and optimization of control parameters such as learning rate, and momentum coefficient. This approach also is useful to select the most relevant features of the training data which can produce a smaller and less complicated network, with the ability to generalize on freshly presented data, due to the removal of redundant variables.

The GA combines selection, crossover, and mutation operators with the goal of finding the best solution to a problem by searching until the specified criterion is met. The solution to a problem is called a chromosome, which is composed of a collection of genes. In hybrid neuro-fuzzy-genetic applications, genes are the CANFIS parameters to be optimized. The GA creates an initial population and then evaluates this population by training a network for each chromosome. It then evolves the population through multiple generations in the search for the best network parameters.

GAs cause the initial population to evolve towards a population that is expected to contain the best solution [7]. We

use the following reproduction evaluation cycle for each iteration-referred to as a generation. Chromosomes (individuals) from the current population are selected with a given probability; and copies of these chromosomes (individuals) are created. The selection of chromosomes is based on their fitness relative to the current population; that is, the stronger chromosomes will have a higher probability of being copied. The fitness is a function of the *CANFIS* model's response. Selected chromosomes are subjected to mutation and to crossover. Fig. 3 shows the *CANFIS*/genetic algorithm cycle for search of optimum parameters of the model.

These mathematical chromosomes could be operated upon by quasi-genetic processes of crossing over and mutation. To implement crossovers, chromosomes were randomly paired, and segments of paired chromosomes between two randomly determined breakpoints were swapped. Inversions could also be modeled, so that exchanged genetic material could be inverted before becoming incorporated into the recipient chromosome.

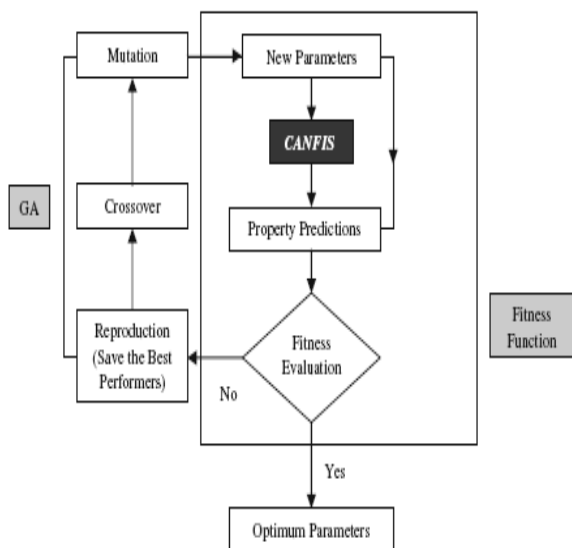


Fig. 3 The *CANFIS* /GA cycle for Optimization

Mutations were implemented by flipping a bit at a binary locus, so that a "0" bit was converted to a "1," or a "1" bit was converted to a "0." In this paper, for the optimization of the *CANFIS* model, GA used the serial method of binary type, roulette-wheel in the selection operator, tow-point crossover in the crossover operator, and boundary in the mutation operator. Automatic determination of the chromosomes length used to optimal search is one of the most important capabilities of the NeuroSolution software. Thus, all the chromosomes were automatically set in this software so that they consisted of the number of input neurons and membership functions, learning rate, and momentum.

IV. EXPERIMENTAL RESULTS

The simulations were realized by using NeuroSolution software. The distribution of datasets is given in Table II. The

training and testing datasets classification by *CANFIS* is given in Table III and IV respectively. The mean square error is only 0.000842.

TABLE II
DISTRIBUTION OF DATA

| Class | No HD | Type 1 HD | Type 2 HD | Type 3 HD | Type 4 HD |
|-------|-------|-----------|-----------|-----------|-----------|
| True | 81 | 27 | 17 | 17 | 6 |
| False | 0 | 0 | 1 | 0 | 0 |

TABLE III
TRAINING DATASET CLASSIFICATION

| Class | No HD | Type 1 HD | Type 2 HD | Type 3 HD | Type 4 HD |
|------------------|-------|-----------|-----------|-----------|-----------|
| Training dataset | 82 | 28 | 18 | 18 | 7 |
| Testing dataset | 82 | 27 | 18 | 17 | 6 |

TABLE IV
TESTING DATA CLASSIFICATION

| Class | No HD | Type 1 HD | Type 2 HD | Type 3 HD | Type 4 HD |
|-------|-------|-----------|-----------|-----------|-----------|
| True | 81 | 27 | 17 | 18 | 7 |
| False | 1 | 1 | 1 | 0 | 0 |

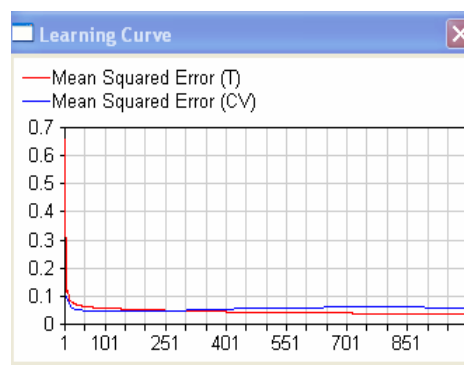


Fig. 4 Learning Curve for the cleaveland database

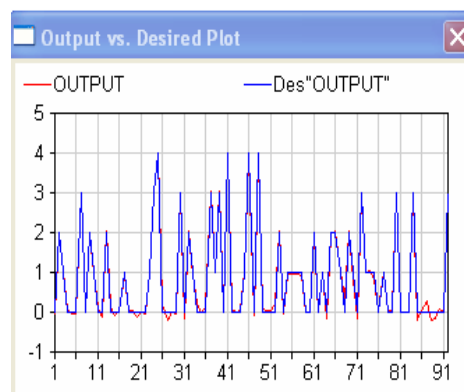


Fig. 5 Output Vs Desired Plot for the cleaveland database

V. CONCLUSION

From our studies, we have managed to achieve our research objectives. Available dataset of Heart disease from UCI Machine Learning Repository has been studied and preprocessed and cleaned out to prepare it for classification process. Coactive Neuro-fuzzy modeling was proposed as a dependable and robust method developed to identify a non-linear relationship and mapping between the different attributes. It has been shown that of GA is a very useful technique for auto-tuning of the CANFIS parameters and selection of optimal feature set. The fact is that computers cannot replace humans and by comparing the computer-aided detection results with the pathologic findings, doctors can learn more about the best way to evaluate areas that computer-aided detection highlights.

Latha Parthiban was born in Chennai, India, in 1973. She received the B.Eng. degree in Electronics and Communication Engineering from the University of Madras in 1994 and received her Masters from Birla Institute of Science and Technology in 1997. She is presently pursuing her Research in Department of Computer Science, Pondicherry University, India. Her fields of interest include Wavelets, image and Signal processing, , Computer Vision and Network Security .

Dr.R.Subramanian is a Professor in the Department of Computer Science, Pondicherry University, India. His research interest includes Numerical Analysis, Parallel Algorithms, Software Engineering,, Image Processing and Pattern Recognition

ACKNOWLEDGMENT

The heart-disease database is from the University of California, Irvine Repository of Machine Learning Databases and originates from R. Detrano at the Cleveland Clinic Foundation

REFERENCES

- [1] C.L.Blake, C.J.Mertz, "UCI Machine Learning Databases", <http://mllearn.ics.uci.edu/databases/heartdisease/2004>.
- [2] R.Wu, W.Peters, M.W.Morgan, "The Next Generation Clinical Decision Support: Linking Evidence to Best Practice", *Journal of Healthcare Information Management*. 16(4), pp. 50-55, 2002.
- [3] Mary K.Obenshain, "Application of Data Mining Techniques to Healthcare Data", *Infection Control and Hospital Epidemiology*, vol. 25, no.8, pp. 690-695, Aug. 2004.
- [4] G.Camps-Valls, L.Gomez-Chova, J.Calpe-Maravilla, J.D.Martin-Guerrero, E.Soria-Olivas, L.Alonso-Chorda, J.Moreno, "Robust support vector method for hyperspectral data classification and knowledge discovery." *Trans. Geosci. Rem. Sens.* vol.42, no.7, pp.1530-1542, July.2004.
- [5] R.B.Burrows, J.Montpetit, M.Faucher, J.Walmsley, "CART-NEURO-FUZZY statistical data modeling", 14th *Conference on Probability and Statistics in the Atmospheric Sciences*, pp. 160-167. Amer. Meteorological Society, MA.,1998.
- [6] JSR Jang, CT Sun, E Mizutani, *Neuro-Fuzzy and Soft-Computing- A Computational Approach to Learning and Machine Intelligence*. Prentice-Hall, Englewood Cliffs, NJ, 1997.
- [7] Taho Yang, Huan-Chang Lin and Meng-Lun Chen "Meta modeling approach in solving the machine parameters optimization problem using neural network and genetic algorithms: a case study". *Rob. Comp-Int. Manuf.* vol. 22, no.4, pp. 322- 331, Aug. 2006.
- [8] Stuart Russell, and Peter Norvig, "Artificial Intelligence: A Modem Approach". Prentice Hall, Englewood Cliffs, NJ,1995.
- [9] Sellappan Palaniappan, Rafiah Awang "Intelligent Heart Disease Prediction System Using Data Mining Techniques" International Conference on Computer Systems and Applications, April 2008. pp.108-115, AICCSA 2008. IEEE/ACS.
- [10] Magnus Stensmo and Terrence J. Sejnowski "Automated Medical Diagnosis based on Decision Theory and Learning from Cases " *World Congress on Neural Networks 1996 International Neural Network-society* pp. 1227-1231.