PERGAMON

Pattern Recognition 36 (2003) 61–68

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Classification of heart rate data using artificial neural network and fuzzy equivalence relation

U. Rajendra Acharya[a], P. Subbanna Bhat[b], S.S. Iyengar[c,*], Ashok Rao[d], Sumeet Dua[c]

[a]*Ngee Ann Polytechnic, 535 Clementi Road, Singapore, 599 489, Singapore*
[b]*Karnataka Regional Engineering College, Surathkal, Srinivasnagar, India 574 157*
[c]*Department of Computer Science, Louisiana State University, 298 Coates Hall, Baton Rouge, LA 70808, USA*
[d]*CEDT, IISc, Bangalore, India*

Received 13 August 2001; accepted 25 February 2002

Abstract

The electrocardiogram is a representative signal containing information about the condition of the heart. The shape and size of the P-QRS-T wave, the time intervals between its various peaks, etc. may contain useful information about the nature of disease afflicting the heart. However, these subtle details cannot be directly monitored by the human observer. Besides, since bio-signals are highly subjective, the symptoms may appear at random in the time scale. Therefore, the signal parameters, extracted and analysed using computers, are highly useful in diagnostics. This paper deals with the classification of certain diseases using artificial neural network (ANN) and fuzzy equivalence relations. The heart rate variability is used as the base signal from which certain parameters are extracted and presented to the ANN for classification. The same data is also used for fuzzy equivalence classifier. The feedforward architecture ANN classifier is seen to be correct in about 85% of the test cases, and the fuzzy classifier yields correct classification in over 90% of the cases. © 2002 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.

Keywords: Heart rate; Pattern recognition; ECG; Neural network; Fuzzy equivalence; Disease classification

1. Introduction

Electrocardiography deals with the electrical activity of the heart. Monitored by placing sensors at the limb extremities of the subject, electrocardiogram (ECG) is a record of the origin and propagation of the electric potential through cardiac muscles. It is considered a representative signal of cardiac physiology, useful in diagnosing cardiac disorders.

The state of cardiac health is generally reflected in the shape of ECG waveform and heart rate. It may contain important pointers to the nature of diseases afflicting the heart. However, bio-signals being non-stationary signals, this reflection may occur at random in the time scale. (That is, the disease symptoms may not show up all the time, but would manifest at certain irregular intervals during the day.) Therefore, for effective diagnostics, the study of ECG pattern and heart rate variability signal (instantaneous heart rate against time axis) may have to be carried out over several hours. Thus, the volume of the data being enormous, the study is tedious and time consuming. Naturally, the possibility of the analyst missing (or misreading) vital information is high. Therefore, computer-based analysis and classification of diseases can be very helpful in diagnostics.

The present paper makes use of heart rate variability (HRV) as the base signal for analysis and classification of diseases. The heart rate is evaluated by measuring the time interval between the successive *R*-peaks (*R*–*R* interval) of the ECG waveform. It is known that almost all the useful frequency components in ECG signal falls below 40 Hz [1], and therefore sampled at the rate of 200 samples/s. The heart rate, plotted against the time scale provides the HRV signal,

* Corresponding author. Tel.: +1-225-578-1252; fax: +1-225-388-1465.
E-mail address: iyengar@bit.csc.lsu.edu (S.S. Iyengar).

PII: S0031-3203(02)00063-8

from which certain parameters are extracted for classification [2,3].

2. Neural network classifier

Artificial neural networks (ANN) are biologically inspired networks—inspired by the human brain in its organization of neurons and decision making process—which are useful in application areas such as pattern recognition, classification, etc. [4]. The decision making process of the ANN is more holistic, based on the aggregate of entire input patterns, whereas the conventional computer has to wade through the processing of individual data elements to arrive at a conclusion.

The neural networks derive their power due to their massively parallel structure, and an ability to learn from experience. They can be used for fairly accurate classification of input data into categories, provided they are previously trained to do so. The accuracy of the classification depends on the efficacy of training, which in turn depends upon the rigor and depth of the training. The knowledge gained by the learning experience is stored in the form of connection weights, which are used to make decisions on fresh input.

Three issues need to be settled in designing an ANN for a specific application: (i) topology of the network; (ii) training algorithm and (iii) neuron activation function. A network may have several 'layers' of neurons and the overall architecture may either be feedback or feedforward structure. If the task is merely to distinguish linearly separable classes, a single layer perceptron classifier is quite adequate. If the class separation boundaries can be piecewise linear approximated, then a two layer perceptron classifier needs to be used. If the class boundaries are more complex, a three layer *feedforward* neural network, with sigmoid activation function is more suitable [5,6]. The most important reason in favour of such a network is that the sigmoid function $f(x)$ is differentiable for all values of x, which allows the use of the powerful *backpropogation* learning algorithm (BPA) [7]. In the present case, the nature of class boundaries is not clearly known, and therefore, the three layer network with sigmoid activation function is being used as classifier (Fig 1).

The BPA is a supervised learning algorithm, which aims at reducing the overall system error to a minimum. The connection weights are randomly assigned at the beginning; and progressively modified to reduce the overall mean square system error. The weight updating starts with the output layer, and progresses backwards. The weight update aims at maximizing the rate of error reduction, and hence it is termed as 'gradient descent' algorithm [8]. The weight increment is done in 'small' steps; the step size is chosen heuristically, as there is no definite rule for its selection. In the present case, a learning constant $\eta = 0.9$ (which controls the step size), is chosen by trial and error.

It is desirable that the training data set be large in size, and also uniformly spread throughout the class domains. In the absence of a large training data set, the available data may be used iteratively, until the error function is reduced to an optimum level. For quick and effective training, data is fed from all classes in a routine sequence, so that the right message about the class boundaries is communicated to the ANN.

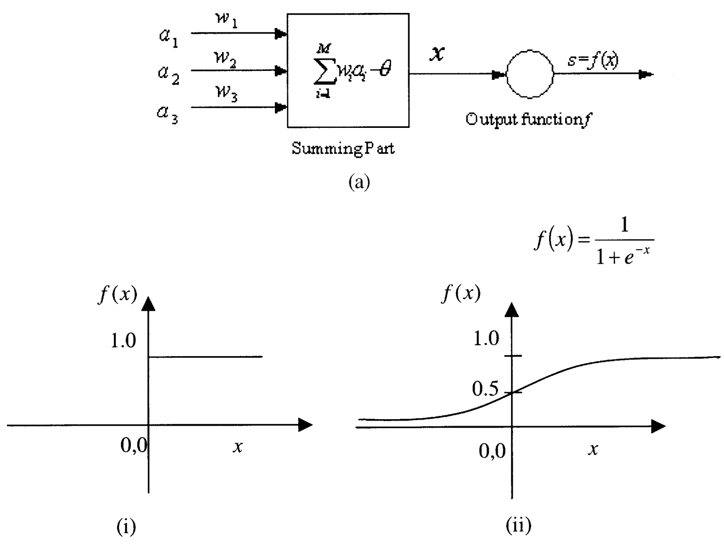

Fig. 1. (a) Model of an artificial neuron (processing unit). (b) Neuron activation functions: (i) unipolar binary functions; and (ii) unipolar sigmoid function.

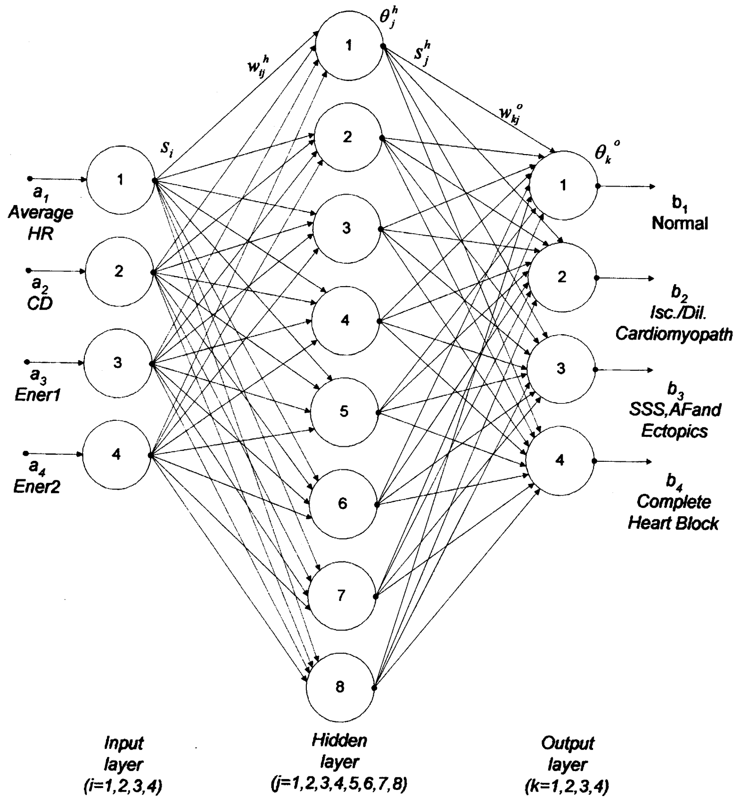

Fig. 2. Three-layer feedforward neural network classifier.

The ANN used for classification is shown in Fig. 2. The input layer consists of nodes to accept data, and the subsequent layers process the data using the activation function. The output layer has four neurons, giving rise to an output domain of 16 possible classes. However, the network is trained to identify only four classes given by decoded binary outputs [0001, 0010, 0100, 1000]. The outputs of the hidden layer (s_j^h) and output layer (b_k) are evaluated using Eqs. (1) and (2)

$$s_j^h = f\left(\sum_{i=1}^{4} w_{ji}^h s_i - \theta_j^h\right), \tag{1}$$

$$b_k = f\left(\sum_{j=1}^{4} w_{kj}^o s_j^h - \theta_k^o\right), \tag{2}$$

where w_{ji}^h and w_{kj}^o are the connection weights and θ_j^h and θ_k^o are the bias terms, respectively.

The error vectors of hidden layer (e_j) and output layer (e_k) are calculated using Eqs. (3) and (4), respectively:

$$e_k = b_k(1 - b_k)(d_k - b_k), \tag{3}$$

$$e_j = s_j^h(1 - s_j^h)\sum_{k=1}^{4} w_{kj}e_k, \tag{4}$$

where d_k is the desired output.

The weight update equations of the output and hidden layers are given below:

$$w_{kj}(new) = w_{kj} + \eta s_j^h e_k, \tag{5}$$

$$w_{ji}(new) = w_{ji} + \eta s_i e_j, \tag{6}$$

$$\theta_k^o(new) = \theta_k^o + \eta e_k, \tag{7}$$

$$\theta_j^h(new) = \theta_j^h + \eta e_j. \tag{8}$$

3. Disease classification using ANN

For the purpose of this study, the cardiac disorders are classified into four categories namely:

(i) Ischemic/dilated cardiomyopathy,
(ii) Complete heart block,
(iii) Sick sinus syndrome, atrial fibrillation (AF), ectopics,
(iv) Normal.

Table 1
Range of input parameters of ANN classification model

Class	HR (bmp) (Average)	Ener 1	Ener 2	CD
Normal	50–100	0.06–0.35	< 0.8	$\geqslant 0.33$
Isc/dil. cardiomyopathy	60–120	0.06–0.60	> 0.8	$\geqslant 0.80$
Complete heart block	30–40	0.05–0.32	> 0.4	$\geqslant 0.80$
Ectopics, AF & SSS	50–100	0.20–2.05	> 0.3	$\geqslant 0.20$

The ANN classifier is fed by four parameters derived from the heart rate signal:

(1) Average heart rate: Though the heart rate is a non-stationary signal, the range of heart rate for various disease categories are seen to be different, the average heart rate can serve as a parameter of classification (Table 1). The average is evaluated for 10 min interval. Secondly, the frequency of heart rate variation for various diseases are seen to be different. The power spectrum of heart rate variability signal shows a marked concentration of energy in different frequency bands [9–11]. Therefore, the ratio of energy content in different frequency bands can be used as parameters of classification. In the present case, two input signals are derived by evaluating the ratio of energy content in two separate frequency bands:
(2) Ener 1 = [energy content in the band (33.3–100 Hz)]/[energy content in the band (0–33.3 Hz)]
(3) Ener 2 = [energy content in the band (66.7–100 Hz)]/[energy content in the band (0–66.7 Hz)]
(4) Correlation dimension factor: Heart rate signal being a non-stationary signal, important insight can be gained from a phase-space plot obtained by representing heartrate $x(k)$ in X-axis and *delayed heartrate* $x(k+m)$ in Y-axis [12–14]. A technique for estimating the embedding dimension of the phase-space pattern was proposed [15,16]. In the present work an embedding dimension of 5 was chosen.

When the heartrate is steady and unchanging, the phase-space plot reduces to a point, but otherwise, the trajectory spreads out to give some patterns on the screen. An example of the phase-space plot of normal heart rate is shown in Fig. 3. The pattern that emerges can be interpreted for finer details—such as whether the heart rate is periodic, chaotic, or random, etc. A correlation dimension factor is defined to obtain a quantitative measure of the nature of trajectory, and the ranges of CD factor for various heart diseases are identified. The CD factor is defined as

$$CD = \lim_{r \to 0} \frac{\log C(r)}{\log(r)}, \tag{9}$$

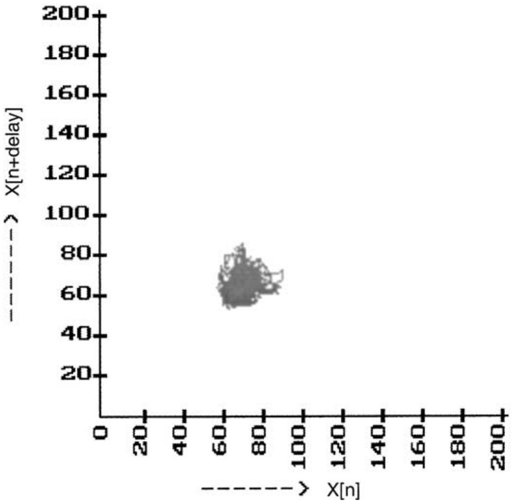

Fig. 3. Phase space plot of normal heart rate (CD = 0.46).

where the correlation integral $C(r)$ is given by

$$C(r) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1, i \neq j}^{N} \Theta(r - |x_i - x_j|), \tag{10}$$

where x_i, x_j is the points of the trajectory in the phase space, N is the number of data points in phase space, r the radial distance around each reference point x_i, and Θ is the Heaviside function.

The range of CD for different classes of diseases is shown in Table 1.

For the purpose of training and testing the classifier, a data base of 342 patient samples is divided into two sets—a training set of 276 arbitrarily chosen samples and a test set of 66 samples (Table 1). The training consisted of 10,000 iterations.

During the *training* phase, each output of the ANN is an analog value in the range 0 → 1.0, whereas the 'desired' output is either 0 or 1.0. During the *recall* phase, the output signal is approximated to binary levels by comparing it with threshold value of 0.5.

4. Fuzzy equivalence relation classifier

A more efficient classifier is developed using fuzzy equivalence relation. The process of classification involves obtaining a *fuzzy relation* matrix for each class of data, and then comparing a fresh input with each group for classification [17].

The fuzzy *equivalence relation* requires the properties of reflexivity, symmetry and transitivity, be satisfied. If it satisfies only the first two—that is, reflexivity and symmetry properties—it is termed as fuzzy *compatible* relation. Though it is usually difficult to identify an equivalence relation directly, it is possible to identify a compatible

Table 2
Training and testing data set

Class	No. of data set used for training	No. of data set used for testing	of correct of correct classification (10,000 iterations) (%)
Normal	90	20	85
Isc/dil. cardiomyopathy	90	20	85
Complete heart block	12	06	84
Ectopics, AF & SSS	84	20	90

relation in terms of an appropriate '*distance function*' of the Minkowski class. The general expression used for the *distance function* (Minkowski class) is given below

$$R(x_i, x_j) = 1 - \delta \left(\sum_{l=1}^{n} |x_{il} - x_{jl}|^q \right)^{1/q}, \tag{11}$$

where n is the total dimensionality of the input data point, l the dimensionality index of the input data $(1, 2, \ldots, n)$, p the size of the input data set, i, j the input index $i, j \in [1..p]$, q the distance function parameter, and δ the normalizing factor to ensure the resultant $R(x_i, x_j) \in [0, 1]$.

n represents the total dimensions of the data, each of which dimension refer to the components of the input data. For example, from Table 1, the input data (HRV signal) is represented by four components [HR(Average), Ener 1, Ener 2, CD], hence $n = 4$, here.

The Minkowski relation can be evaluated for integer values of q; for $q = 1$, the 'distance function' happens to be the Hamming distance; for $q = 2$, it is the Euclidean distance, etc. The normalizing factor δ is taken as the inverse of the largest distance among the data pairs.

As indicated above, for our purposes, the input data (HRV signal) is represented using the four parameters used for ANN classification earlier (Table 1). Thus the data has four components ($n=4$). The size of the training data set (defined by p_k $k \in [1..4]$) is different for each class i (second column of Table 2).

In the present case, the Euclidean *distance function* of Minkowski class ($q=2$) is used as the basis to define mutual relation among the input data belonging to a particular class. Thus, Eq. (11) reduces to

$$R(x_i, x_j) = 1 - \delta \left(\sum_{l=1}^{4} |x_{il} - x_{jl}|^2 \right)^{1/2},$$

where the symbols have their usual meaning.

The result of the above evaluation can be listed in the form of a symmetrical $p_k x p_k$ matrix, which satisfies both

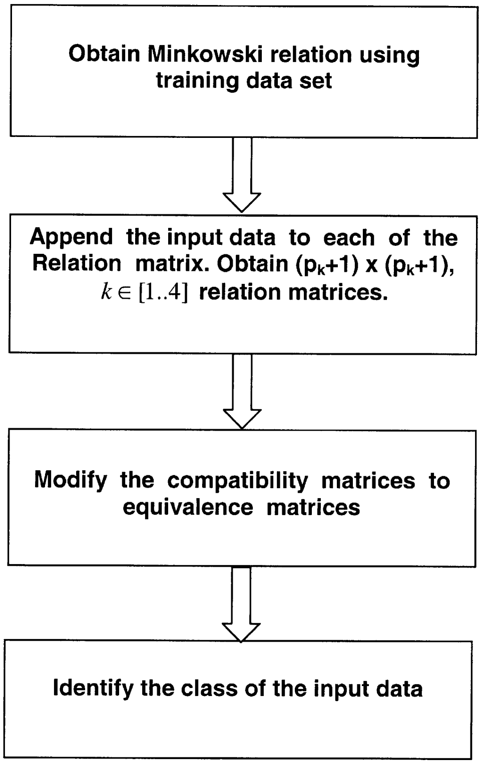

Fig. 4. Flowdiagram for prediction using fuzzy equivalence relation.

reflexivity and symmetry conditions. Thus it defines a compatibility relation, but not necessarily the equivalence relation. Therefore, the relation matrix is further processed to obtain a transitive closure (TC), equivalence relation. The processing algorithm is described in the following discussion. Firstly, few definitions are in order.

For a relation R, we write $R(u, v)$ if the pair $\langle u, v \rangle$ is a member of the set. The TC of a relation R, often written R^* is the smallest set such that if $R(u, v)$ and $R(v, w)$, then $R(u, w)$. For example, the 'ancestor relation' is the TC of the 'parent relation', 'dominates' is the TC of the 'immediately dominates' relation, the 'greater than' relation $u > v$ is the TC of the 'successor relation' $\mathrm{succ}(u, v)$ (which relates integers to the next greater integer $\mathrm{succ}(u, v)$).

We also talk about the *reflexive* TC if $R^*(u, u)$ holds generally.

Given a relation R, its TC R^* can be determined as follows. R is transitive iff $(a, b) \in R \Lambda (b, c) \in R \Lambda (a, c) \in R$. We may add elements to a relation R and create a new relation that is the TC of R. However, the procedure requires an iterative process. We find the TC by examining every pair of elements of R where the second element of the first pair matches the first element of the second pair.

That is $(a, b) \in R$ and $(b, c) \in R$.

Algorithm *Fuzzy_Classifier*
Begin

1. Initialization
    1.1 Read classified data from am input file
        $[\text{input\_data}]_{i,j}[\text{classes}]_k$ <- Buffer
            i: 1 to m (number of data)
            j: 1 to n (number of attributes)
            k: 1 to i; where there are '0' classes
    1.2 Read unclassified data from test file
        [unclassified]i,j <- Buffer
            i: 1 to p(number of data)
            j: 1 to n(number of attributes)
2. Preprocessing
    2.1 Append unclassified_data onto input_data
        $[\text{input\_data}]_{i,j}$ <- $[\text{input\_data}]_{i,j} \Downarrow [\text{unclassified\_data}]_{k,j}$
            i: m+1 to m+p
            k:1 to p
            j: 1 to n
    2.2 Normalize the data matrix attribute wise selecting the maximum in each attribute

$$[\text{input\_data}]_{i,j} \leftarrow \frac{[input\_data]_{i,j}}{\max\{[input\_data]\}_j}$$

            i: 1 to (m+p)
            j: 1 to n
3. Compute the Fuzzy Equivalence Relation
    3.1 Find the compatibility relation between the data using distance function of Minkowski class

$$R_c(x_i, x_j) = 1 - \delta(\sum_{l=1}^{n}(x_{il} - x_{jl})^2))^{1/2}$$

            i,j: 1 to m+p
            where $\delta = 1/n^{1/2}$
    3.2 Find the transitive closure of R_c
4. Classification
    Beginning from row m+1 search for the maximum membership degree until column m and store the corresponding index in matrix max_membership
        $[\text{max\_membership}]_i = \max[R_t]_{i,j}$;
            i:m+1 to m+p
            j:1 to m
    corresponding to each unclassified data the class to which it belongs is given by [classes][max_membership]; i:1 to p

End

Fig. 5. Algorithm *fuzzy_classifier* for classification using fuzzy equivalence relation.

Table 3
Results of fuzzy equivalence classification[a]

Class	No. of data set set used for training	No. of data set set used for testing	Percentage of Percentage of correct classification (%)
Normal	90	20	95
Isc/dil. cardiomyopathy	90	20	90
Complete heart block	12	06	100
Ectopics, AF & SSS	84	20	95

[a]The implementation was experimented on a variety of data sets and results presented here represent the average performance.

Transitivity requires that (a, c) must also be an element of R. If it is not, then we must add it to the new relation that we are building into the TC. Let us call the new relation R'. (Initially $R' = R$ and when the process of adding edges is over $R' = R^*$.) After we have examined all such pairs of members of R and added the required edges to R' where needed we must then begin the same process again.

The resultant R^* is the TC of R.

After computing the TC and having satisfied the properties of reflexivity, symmetry and transitivity, the fuzzy equivalence relation matrix so obtained can now be used for classification of fresh data. The flowdiagram to do so is depicted in Fig. 4.

The formal algorithm is depicted in Fig. 5. The results of the classification are listed in Table 3.

5. Conclusion

Both the neural network classifier and the fuzzy equivalence relation are developed as diagnostic tools to aid the physician. The tools do not yield results with 100% accuracy. The accuracy of the tools depend on several factors, such as the size and quality of the training set, the rigor of the training imparted, and also parameters chosen to represent the input. The results listed in Tables 2 and 3, indicate that the classifiers are effective to the tune of about 85–95% accuracy.

Acknowledgements

This project is in part supported by the LEOFS-Board of Regents funding on BCVC program.

References

[1] M. Merri, D.C. Farden, J.G. Mottley, et al., Sampling frequency of the electrocardiogram for spectral analysis of the heart rate variability, IEEE Trans. Biomed. Eng. 37 (1990) 99–106.

[2] R. Wariar, C. Eswaran, Integer coefficient bandpass filter for the simultaneous removal of baseline wander, 50 and 100 Hz interference from the ECG, Med. Biol. Eng. Comput. (1991) 333–336.

[3] J. Pan, W.J. Tompkins, Real time QRS detector algorithm, IEEE Trans. Biomed. Eng. 32 (3) (1985) 230–323.

[4] R.P. Lippman, Pattern classification using neural networks, IEEE Commun. Mag. (1989) 47–64.

[5] S.V. Kartalopoulos, Understanding Neural Networks and Fuzzy Logic-Basic Concepts and Applications, Prentice-Hall, New Delhi, 2000.

[6] S. Haykin, Neural Networks: A Comprehensive Foundation, MacMillan College Publishing Company, New York, 1995.

[7] P. Patterson, S. Draper, A neural net representation of experienced and non-experienced users during manual wheelchair propulsion, J. Rehabil. Res. Dev. 35 (1998) 43–51.

[8] B. Yegnanarayana, Artificial Neural Networks, Prentice-Hall, New Delhi, India, 1999.

[9] A. Glenn Myers, G.J. Martin, N.M. Magid, P.S. Barnett, J.W. Schaad, J.S. Weiss, M. Lesch, D.H. Singer, Power spectral analysis of sudden cardiac death: comparison to other methods, IEEE Trans. Biomed. Eng. BME-33 (12) (1986) 1149–1156.

[10] L. Philip Saul, Y. Arai, R.D. Berger, L.S. Lilly, W.S. Colucci, R.J. Cohen, Assessment of autonomic regulation in chronic congestive heart failure by heart rate spectral analysis, Am. J. Cardiol. 61 (1988) 1292–1299.

[11] M.R. Sheila, A.L. Goldberg, R. Ruthazer, J. Mietus, L.A. Lipsitz, Spectral analysis of heart rate dynamics in elderly persons with postprandial hypotension, Am. J. Cardiol. 69 (1992) 201–205.

[12] A.M. Fraser, H.L. Swinney, Independent coordinates for strange attractors from mutual information, Phys. Rev. A 33 (1986) 1134–1140.

[13] A.M. Fraser, Information and entropy in strange attractors, IEEE Trans. Inform. Theory 35 (1989) 245–262.

[14] N. Iyengar, C.K. Peng, A.L. Goldberger, L.A. Lipsitz, Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics, Am. J. Physiol. 271 (1996) 1078–1084.

[15] P. Grossberger, I. Procassia, Measuring the strangeness of strange attractors, Physica D 9 (1983) 189–208.

[16] M. Ding, E. Grebogi, E. Ott, T. Sauer, J.A. Yorke, Estimating correlation dimension from a chaotic time series: when a plateau occur? Physica D 69 (1993) 404–424.

[17] G. Klir, B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice-Hall, New Delhi, India, 1995.

About the Author—U. RAJENDRA ACHARYA obtained his M. Tech in Bio-Medical Engineering in 1997 from Manipal Institute Technology, Manipal and Ph.D from Karnataka Regional Engineering College, Surathkal, India in 2001. He is a Associate Professor, in Bio-Medical Engineering department in Manipal Institute of Technology, Manipal, India. Presently, he is working as a visiting faculty in Ngee Ann Polytechnic, Singapore.

His current interests are Visualization, Bio-signal processing and Image Processing.

About the Author—P. SUBBANNA BHAT obtained his B.E and M. Tech. degrees in Electronics and Communication Engg. from the Karnataka Regional Engg College, Surathkal, India, in 1974 and 1977. He obtained his Ph.D. in the area of Power Electronics from Indian Institute of Technology, Kanpur in 1984. Ever since, he is with the faculty of Electronics and Communication Engg., at KREC Surathkal.

His current interests are Control Theory, Signal Processing, Fuzzy Logic and Neural Networks.

About the Author—S.S. IYENGAR is the professor and head of Department of Computer Science at Louisiana State University, Baton Rouge, LA. He has written over 250 research papers in leading journals and conferences. He has also authored over 5 textbooks. He is an IEEE Fellow, Fellow of the ACM (Association of Computing Machinery) and Fellow of the American Association of Advancement of Science (AAAS). He is also the winner "Research Master Award" at Louisiana State University.

His current interests are Data Mining, Bioinformatics, Multi-sensor fusion, Signal Processing and Image processing.

About the Author—ASHOK RAO obtained his M.Tech and Ph.D degrees from IISc, Bangalore and Indian Institute of Technology, Bombay in 1984 and 1991, respectively. Presently, he is the head, Network Project at IISc Bangalore, which focuses on the HR development in Southern India through Undergraduate Electronics Engineering Education. This is funded by SDC (Swiss Agency for Development and Co-operation). His area of Interests is Signal and Image Processing, Pattern Recognition, Neural Networks, Applied Linear Algebra, DSP Algorithms and Architecture, SPV, Technical Education and Appropriate Technology.

About the Author—SUMEET DUA obtained his MS in Systems Science and Ph.D. in Computer Science from Louisiana State University, Baton Rouge, LA in 2000 and 2002, respectively. Presently, he is a Postdoctoral Researcher at Biological Computing and Visualization Center, Department of Ophthalmology, Louisiana State University Health Services Center. He has published in the area of data mining, knowledge discovery, global optimization and bioinformatics. He is a full member of Sigma Xi, New York Academy of Sciences, IEEE Computer Society, ACM and The association for research in vision and ophthalmology. His areas of interest include Data Mining, Bioinformatics, Signal Processing, and Multimedia Indexing.