

Interval type-2 fuzzy logic system for diagnosis coronary artery disease

Adha Mashur Sajiah*, Noor Akhmad Setiawan, Oyas Wahyunggoro

*Department of Electrical Engineering and Information Technology, Universitas Gadjah Mada
Jalan Grafika No.2 Bulak Sumur, Yogyakarta 55281, Indonesia*

Article history:

Received: 30 October 2016 / Received in revised form: 25 November 2016 / Accepted: 27 November 2016

Abstract

Coronary artery disease (CAD) is a disease that has been the deadliest disease in Indonesia. The ratio of cardiologists over potential patients is not appropriate either. Intelligent system which can help doctors or patients for cheap and efficient diagnosing CAD is needed. Medical record data, acquisition of cardiologist knowledge and computing technology can be utilized for developing fuzzy logic based intelligent system. Type-1 fuzzy logic system (T1 FLS) has been widely used in various fields. T1 FS has limitation in representing and modelling uncertainty and minimize the impact. Whereas, type-2 fuzzy set (T2 FS) was also introduced as fuzzy set that can model uncertainty more sophisticated. T2 FLS does have a higher degree of freedom when modeling uncertainty but it is quite difficult to make the membership function. An interval T2 FS is a T2 FS in which the membership grade on third dimension is the same everywhere so it is simpler than T2 FS. This paper aims to clarify the better capability of IT2 FLS over T1 FLS on the development of CAD diagnosis system. Rules and membership function were formulated with the help of fuzzy c-means. This study illustrated the causes of CAD risk factors, fuzzification, type reduction and defuzzification. The resulted system was tested with percentage split method (50%-50%) to produce training data and testing data. This test is performed ten times with random seed to separate the data set. The resulted system generates an average of 73.78% accuracy, 71.94% sensitivity and 76.52% specificity.

Keywords: Coronary Artery Disease; Fuzzy Logic System; Interval Type-2 Fuzzy Set; Fuzzy C Means.

1. Introduction

Coronary artery disease (CAD) is one of the biggest cause of death in the world in the last decade [1]. CAD became pandemic because developing countries experiencing the epidemiological transition of diseases from hunger to degenerative diseases. Moreover, CAD tends to afflict younger population and thus can reduce productivity and employment [2]. Not least in Indonesia the data in 2012 show the disease is the top cause of death in Indonesia [3].

CAD is a disease that is caused by the accumulation of plaques in the coronary arteries. The coronary arteries supply the heart with a charge of oxygen-rich blood. Plaque is made up of fat, cholesterol, calcium and other elements. This condition is called atherosclerosis [4].

The main problem is the Indonesian people come to the doctor when the experienced symptoms are more severe CAD. If the patients can be more aware of heart health and can check their heart health with early symptom stages it will help people with CAD.

In the previous case it is necessary to develop a computer-based classification system of CAD. Artificial intelligence system become a solution that is appropriate to diagnose early, fast and cheap. Intelligent system that can be selected is using fuzzy logic system.

Type-1 fuzzy logic system (T1 FLS) has been widely used

in various fields for more than three decades [5]. L. Zadeh introduced the fuzzy sets of type-1 (T1 FS) in 1965. While T1 FS is meant to represent uncertainty, studies have shown T1 FS has limitations in representing and modeling uncertainties and minimize its effect. This is because T1 FS was certainly in the sense that the value of membership is crisp value [6]. Then 10 years after introducing the T1 FS, 1975, L. Zadeh also introduced type-2 fuzzy sets (T2 FS) as a fuzzy set which can be more sophisticated modeling uncertainty. In 1998, Mendel and Karnik provide a complete theory of the type-2 fuzzy logic system (FLS T2) [5]. T2 FS is set to T1 FS-shaped membership functions. Fuzzy logic with a higher order (air-order 2) has in recent years become popular applied in the case of pattern recognition, classification and clustering [7].

Interval type-2 fuzzy sets (IT2 FS) is a special type of T2 FS, FS T2 is the type most widely used due to a lower computational cost than the original FS T2. T2 FLS is a fuzzy logic system in which at least one T2 FS. Fuzzy logic system requires a type reducer to turn T2 to T1 FS because the output of the fuzzy inference engine is T2 FS.

Various studies have shown that results of T2 FLS are better than results of T1 FLS. Tutorials and articles that more and more of the T2 FLS resulted in the amount of research on the T2 FLS is also growing. Although the cost of computing T2 FLS is greater than T1 FLS but advances in computing capability hardware also higher resolve this issue.

T2 FLS has many parameters that need to be considered. Choosing the membership function, choosing the type reducer

* Corresponding author. Tel.: +62-85756851122.

Email: adha.m.sajiah@gmail.com.

or choosing defuzzifier. This is a potential of T2 FLS.

CAD diagnosis automation system with fuzzy logic system has been widely performed. But the fuzzy sets used are still using T1 FS.

Based on the current development of CAD diagnosis with fuzzy logic system, the objective of this research is to clarify the potential capability of IT2 FLS over T1 FLS on CAD diagnosis. With the capability of IT2 FLS, the better performance system could be obtained.

2. Materials and Methods

In this section the data set for training and testing model and proposed methodology are described. The proposed methodology here consists of rules generation, membership function design and interval type 2 fuzzy logic system design.

2.1. Materials

In the development of this system is used data set Statlog (Heart) from the UCI Machine Learning Repository. This data set is similar to the CAD database that already exists in the database repository UCI Heart Disease data set but with a slightly different form. Some differences eg statlog data set has a fewer number of data sets and no any attributes are missing.

Statlog data set consists of 270 observations of patient data. Attributes of these data sets related to the results of a physical examination, laboratory diagnosis and stress testing, all of which amounted to 13 attributes and one attribute information CAD diagnosis. The more detailed data set is shown in Table 1.

2.2. Methods

To build the proposed fuzzy inference system needed fuzzifier, fuzzy rules, fuzzy inference engine, type reducer and defuzzifier. Components of the system according to Fig. 1 which shows the basic components IT2 FLS. To meet these components, we do the following steps.

2.2.1. Rules Generation

There are many rules generation method based on input-output data. The method used in this paper is fuzzy c-means clustering. The number of rules generated from this method is the same as the number of cluster defined for separated data group. So for each attribute (clinical feature) c^i will be grouped to c_1^i and c_2^i class.

Because there are two classes and group of c_1^i as antecedents will establish rule 1 (absent) and group of c_2^i as antecedents will establish rule 2 (present), so there is only two rules ($N_r = 2$). The number of antecedent of each rule is as much as the number of clinical features diagnosis of CAD that is 13 ($N_p = 13$). Model of the rule output fuzzy c-means clustering are as follows:

$$R^n: \text{IF } x_1 \text{ is } \tilde{X}_1^n \text{ and } x_2 \text{ is } \tilde{X}_2^n \text{ and } \dots \text{ and } \tilde{X}_p^n \text{ then } y \text{ is } Y^n$$

where $n = 1, 2 (N_r)$, \tilde{X}_p^n ($p = 1, 2, \dots, N_p$) is IT2 FS for IT2 FLS. Whereas $Y^n = [\underline{y}^n, \bar{y}^n]$ is an interval, simple TSK model.

Table 1. Data set statlog attribute

Attribute	Unit	Range
Age	Year	29-77
Sex	male(1), female(0)	0, 1
Chest Pain Type	typical angina (1), atypical angina (2), non-anginal pain (3), asymptomatic (4)	1, 2, 3, 4
Resting Blood Pressure	mmHg	94-200
Serum Cholesterol	Mg/dl	126-564
Fasting Blood Sugar	fasting blood sugar > 120 mg/dl (1) else (0)	0, 1
Resting Electrocardiographic Results	normal (0), having ST-T wave abnormality (1), LV hypertrophy (2)	0, 1, 2
Maximum heart rate achieved	bpm	71-202
Exercise induced angina	yes (1), no (0)	0, 1
Oldpeak	mm	0-6.2
The slope of the peak exercise ST segment	upsloping (1), flat (2), downsloping (3)	1, 2, 3
Number of major vessels	-	0-3
Thal	normal (3), fixed defect (6), reversible defect (7)	3, 6, 7

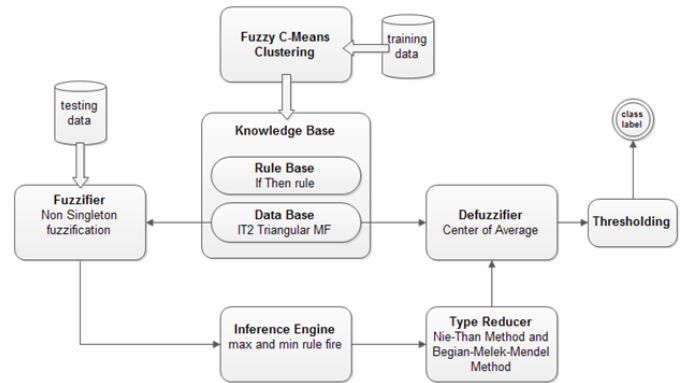


Fig. 1. Structure of IT2 fuzzy logic system CAD classifier

2.2.2. Membership Function Design

Once the rule base is made then we define the data base or membership functions as the base for fuzzification of and defuzzification to crisp input. The use of IT2 FS is a key differentiator in IT2 FLS so designing IT2 FS with determination maximum footprint of uncertainty (FOU) will affect the performance of the classification. At this stage the first we design membership functions T1 FS to be used as baseline T1 MF. Furthermore, based on the T1 baseline, FOU IT2 is determined being presentation randomness and uncertainty of MF data. Here are the details used design strategy.

1) Designing Baseline T1 MF with Fuzzy C-Means Clustering

Fuzzy c-means clustering (FCM) forms a cluster with two attributes, namely {mean, standard deviation}. Membership function that is used in this case is triangular MF. This MF has three parameters $\{l_i, c_i, r_i\}$. To determine these three parameters, we used components of the mean and standard deviation of output FCM as shown in the following equation

$$waist_i = \sigma_i \times \sqrt{-2 \ln 0.5} \quad (1)$$

$$l_i = \mu_i - 2waist_i \quad (2)$$

$$c_i = \mu_i \quad (3)$$

$$r_i = \mu_i + 2waist_i \quad (4)$$

where μ_i is mean, σ_i is standard deviation of a cluster and $waist_i$ is the helper terms for distance between apex triangle and its two points on the foot of triangle by horizontal axis.

2) Designing Footprint of Uncertainty (FOU)

Based on T1 MF described in the preceding stage FOU is designed with hoping to model the uncertainty of such randomness and ambiguities in the data and determining the appropriate membership function.

According to Tan et al [7], a cluster with more scattered data are considered more uncertain, therefore a wider FOU should be used in MF antecedent in the class. The paper used total euclidean distance of data points to its cluster (d_j) then normalized to determine FOU.

From the above it is intuitively using the standard deviation (FCM output) as the equivalent property with the d_j is qualified as a determinant feature of FOU. In this study, there are two areas of FOU determination: a. LMF Height (H_j) which represents the uncertainty in the center of the cluster. b. UMF and LMF end points that represent the uncertainty of fuzzy set limits. For case (a) it is assumed that there is no doubt about the position of the center of the cluster so that there is no difference in height between UMF and LMF. As for the deviation of the position of the end point UMF and LMF, the case of (b) as described previously using the standard deviation of the cluster. UMF and LMF point can be determined by the following equation.

$$l_{i,UMF} = l_i - 0.5\sigma_i \quad (5)$$

$$c_{i,UMF} = c_i \quad (6)$$

$$r_{i,UMF} = r_i + 0.5\sigma_i \quad (7)$$

$$l_{i,LMF} = l_i + 0.5\sigma_i \quad (8)$$

$$c_{i,LMF} = c_i \quad (9)$$

$$r_{i,LMF} = r_i - 0.5\sigma_i \quad (10)$$

where $l_{i,UMF}$ is UMF left point, $c_{i,UMF}$ is UMF center point, $r_{i,UMF}$ is UMF right point, $l_{i,LMF}$ is LMF left point, $c_{i,LMF}$ is LMF center point, $r_{i,LMF}$ is LMF right point, l_i is T1 MF left point, c_i is T1 MF center point, r_i is T1 MF right point. UMF and LMF design can be seen on Fig. 2.

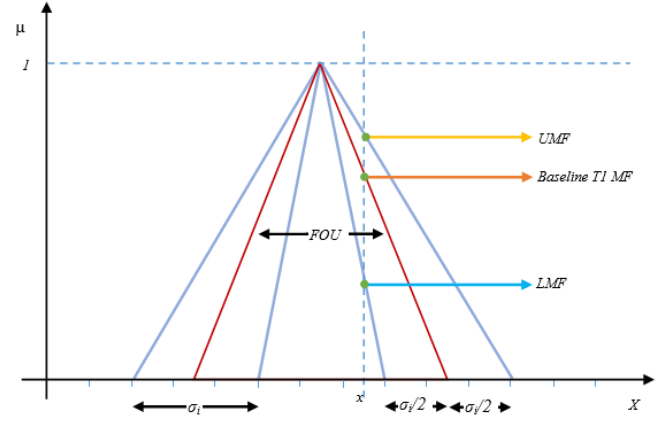


Fig. 2. Design of FOU IT2 MF

2.2.3. Interval Type 2 Fuzzy Logic System Design

1) Fuzzifier

Fuzzifier module in charge of mapping numerical vector $x = (x_1, x_2, \dots, x_p)^T \in X_1 \times X_2 \times \dots \times X_p \equiv X$ into T2 FS \tilde{A}_x in X [8].

2) Rules

Rule base became the basis of doing inference. IT2 rule base in accordance with the steps in 2.2.1. With p inputs (linguistic variable) $x_1 \in X_1, \dots, x_p \in X_p$ and one output $y \in Y$. This system is one kind of multiple input single output system (MISO).

The rules that are established are:

1. IF (age IS ageCluster1) AND (sex IS sexCluster1) AND (chestPainType IS chestPainTypeCluster1) AND (restBP IS restBPCluster1) AND (serumChol IS serumCholCluster1) AND (fbs IS fbsCluster1) AND (restECG IS restECGCluster1) AND (heartRate IS heartRateCluster1) AND (exerciseAngina IS exerciseAnginaCluster1) AND (oldPeak IS oldPeakCluster1) AND (slope IS slopeCluster1) AND (majorVesselNumber IS majorVesselNumberCluster1) AND (thal IS thalCluster1) THEN (prediction IS absent)
2. IF (age IS ageCluster2) AND (sex IS sexCluster2) AND (chestPainType IS chestPainTypeCluster2) AND (restBP IS restBPCluster2) AND (serumChol IS serumCholCluster2) AND (fbs IS fbsCluster2) AND (restECG IS restECGCluster2) AND (heartRate IS heartRateCluster2) AND (exerciseAngina IS exerciseAnginaCluster2) AND (oldPeak IS oldPeakCluster2) AND (slope IS slopeCluster2) AND (majorVesselNumber IS majorVesselNumberCluster2) AND (thal IS thalCluster2) THEN (prediction IS present)

3) Inference

Inference engine combine rules and do mapping from given input to desired output. This mapping is in T2 FS domain. In this studi TSK inference model is used.

In this module firing interval is computed. The firing interval of n^{th} rule, $F^n(x)$ can be computed as

$$F^n(x') = [\mu_{\underline{x}_1^n}(x'_1) \times \dots \times \mu_{\underline{x}_l^n}(x'_l), \mu_{\overline{x}_l^n}(x'_l) \times \dots \times \mu_{\overline{x}_1^n}(x'_1)] \equiv [\underline{f}^n, \overline{f}^n], n = 1, \dots, N \quad (11)$$

4) Type Reducer

Because output of inference engine is IT2 FS so we need a module called type reducer that has task to transform IT2 FS to T1 FS. There are many methods available. The most commonly used one is the center-of-sets type-reducer.

$$Y_{cos}(x') = \bigcup_{y^n \in F^n(x')} \frac{\sum_{n=1}^N f^n y^n}{\sum_{n=1}^N f^n} = [y_l, y_r] \quad (12)$$

with

$$y_l = \min_{k \in [1, N-1]} \frac{\sum_{n=1}^k \underline{f}^n y^n + \sum_{n=k+1}^N \underline{f}^n y^n}{\sum_{n=1}^k \underline{f}^n + \sum_{n=k+1}^N \underline{f}^n} \equiv \frac{\sum_{n=1}^L \underline{f}^n y^n + \sum_{n=L+1}^N \underline{f}^n y^n}{\sum_{n=1}^L \underline{f}^n + \sum_{n=L+1}^N \underline{f}^n} \quad (13)$$

$$y_r = \max_{k \in [1, N-1]} \frac{\sum_{n=1}^k \overline{f}^n y^n + \sum_{n=k+1}^N \overline{f}^n y^n}{\sum_{n=1}^k \overline{f}^n + \sum_{n=k+1}^N \overline{f}^n} \equiv \frac{\sum_{n=1}^R \overline{f}^n y^n + \sum_{n=R+1}^N \overline{f}^n y^n}{\sum_{n=1}^R \overline{f}^n + \sum_{n=R+1}^N \overline{f}^n} \quad (14)$$

where the switch points L and R are determined by

$$\underline{y}^L \leq y_l \leq \underline{y}^{L+1} \quad (15)$$

$$\overline{y}^R \leq y_r \leq \overline{y}^{R+1} \quad (16)$$

and $\{\underline{y}^n\}$ and $\{\overline{y}^n\}$ have been sorted in ascending order, respectively. y_l and y_r can be computed using the Karnik-Mendel (KM) algorithms.

5) Defuzzifier

Defuzzifier block performs simple computation by transform interval value to one crisp values. Defuzzified output is computed as

$$y(x) = \frac{y_l + y_r}{2} \quad (17)$$

Simple Eq. (17) is used because it yields the equilibrium value between two points y_l and y_r .

6) Thresholding

Because the output of defuzzifier is continuous crisp values, we need a function to convert the value to the class label prediction of CAD. CAD present class is marked by 1 and CAD absent class is marked by 0. Eq. (18) is used to determine the class label.

$$C(x) = \begin{cases} 0, & x \leq \theta \\ 1, & x > \theta \end{cases} \quad (18)$$

where θ is a threshold value. The threshold value that is used is 0.5 because it is the average value between class label 0 (absent) and 1 (present).

3. Results and Discussion

The proposed fuzzy logic system was implemented by the

Matlab programming. To test the system performance, we tested the system against statlog data set. Statlog data set contains 270 observation results was divided by split percentage method (50%-50%) into a training data set and testing data set. 50% of data set is used as data training (for modelling) and 50% of data set is used as data testing (validation). Separation was done randomly.

Testing was conducted 10 times so then 10 times also did for separation training data set and testing data set, then based on this training data set came FLS and counted each performance.

For the performance evaluation of the present system we have used the measures of specificity, sensitivity and accuracy. They are defined as follows:

1. Specificity measures percentage of normal patients classified correctly by model. It is determined as

$$specificity = \frac{TN}{FP+TN} \quad (19)$$

2. Sensitivity measures percentage of abnormal patients classified correctly by the model. It is determined as

$$sensitivity = \frac{TP}{TP+FN} \quad (20)$$

3. Accuracy measures percentage of correctly classified patients. It is determined as

$$accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (21)$$

where TP (true positive) denotes the number of abnormal patients correctly classified; TN (true negative) denotes the number of normal patients correctly classified; FP (false positive) denotes the number of healthy patients wrongly classified as abnormal patients; FN (false negative) denotes the number of abnormal patients wrongly classified as normal patients.

Result of ten measurements displayed in Table 2. Examples of the third test, sensitivity of system is 81.25% it means the system can detect 81.25% of patients with CAD and skip 18.75% of CAD patients. The system has a specificity of 75.86%, in other words around 66 of the 87 people with negative results are really negative and 21 positive people are diagnosed when the disease does not have CAD.

Demonstrated system performance is also quite good with an average rate of 73.78% accuracy, 71.94% sensitivity and 76.52% specificity using only 2 rules from FCM clustering. Rule used all of attributes of risk factor (13) for CAD. Our proposed system is then compared with three other systems that have been proposed before, they are: weighted rules FLS by P.K. Anooj [9], Rough Set FLS by Setiawan et al. [10] and Decision Tree FLS by Pal et al. [4].

Table 3 shows the comparative results of the four different systems. Although it is not fair to compare between them because each of them uses different testing method (number of the data validation, split method, used data set etc.), at least we can see the overall picture of the resulted system. From the table we can see our proposed system can outperform weighted FLS on sensitivity, specificity and accuracy. For rough set FLS although our performance is lower than it, we can still have a

chance to refine our system with adding more rules because rough set FLS using more rule sets. We also need to train our system with more data as rough set FLS do. Decision tree FLS have shown that knowledge with orderly maintained make itself the best method here. It does not only have many rules but it also has control over the rules with meta rules. We can counterfeit the decision tree and meta-rules approach to increase of our FLS performance.

Table 2. IT2 fuzzy logic system performance

Test	TP	FP	TN	FN	Sensitivity (%)	Specificity (%)	Accuracy (%)
1	47	13	57	18	72.31	81.43	77.04
2	43	17	55	20	68.25	76.39	72.59
3	39	21	66	9	81.25	75.86	77.78
4	32	28	61	14	69.57	68.54	68.89
5	47	13	51	24	66.20	79.69	72.59
6	37	23	58	17	68.52	71.60	70.37
7	44	16	51	24	64.71	76.12	70.37
8	47	13	53	22	68.12	80.30	74.07
9	36	24	72	3	92.31	75.00	80.00
10	47	13	53	22	68.12	80.30	74.07

Table 3. FLS performance comparison

Test	Sensitivity (%)	Specificity (%)	Accuracy (%)
Weighted Rules FLS	45.22	68.75	57.85
Rough Set FLS	81.00	85.00	83.00
Decision Tree FLS	95.85	83.33	84.20
Proposed FLS	71.94	76.52	73.78

Overall, opportunities to improve our system performance is still wide open either by adding rules, improve the membership function with a membership optimization (using genetic algorithm, particle swarm optimisation etc.), rule management and others [11,12]. Testing of the system is just applied to one data set, the system still needs to be tested against multiple data sets other than heart statlog data set.

4. Conclusion

In this study the structure of Interval Type-2 Fuzzy Logic System for the diagnosis of coronary artery disease has been described. The system has been developed for the early detection of CAD. The rules were automatically extracted with fuzzy c-means clustering and all of the clinical attributes were used to define risk of CAD patients. Interval type-2 fuzzy logic system has big opportunity to improve the resulted system more.

References

1. A. S. Go, D. Mozaffarian, V. L. Roger, E. J. Benjamin, J. D. Berry, M. J. Blaha, et al., *Executive summary: heart disease and stroke statistics-2014 update: a report from the American Heart Association*, Circulation. 129 (2014) 399–410.
2. World Health Organization, *Noncommunicable Diseases (NCD) Country Profiles: Indonesia*. 2014.
3. A. R. Omran, *Changing patterns of health and disease during the process of national development*. In: Health, Illness, and Medicine, G. L. Albrecht and P. C. Higgins, eds. Chicago: Rand McNally, 1979.
4. D. Pal, K. M. Mandana, S. Pal, D. Sarkar and C. Chakraborty, *Fuzzy expert system approach for coronary artery disease screening using clinical parameters*, Knowledge-Based Systems. 36 (2012) 162–174.
5. I. Maalej, C. Rekik, D. Ben, H. Abid and N. Derbel, *Interval Type-2 Takagi-Sugeno-Kang Fuzzy Logic Approach for Three-Tank System Modeling*, Industrial Electronics (ISIE), IEEE 23rd Int. Sympos, Istanbul, Turkey, 2014, pp. 144–149.
6. D. Wu, *A Brief Tutorial on Interval Type-2 Fuzzy Sets and Systems*. 2013.
7. P. Melin and O. Castillo, *A review on type-2 fuzzy logic applications in clustering, classification and pattern recognition*, Appl. Soft Comput. 21 (2014) 568–577.
8. W. W. Tan, C. L. Foo and T. W. Chua, *Type-2 Fuzzy System for ECG Arrhythmic Classification*, Fuzzy System Conference, IEEE International, 2007, pp. 1–6.
9. P. K. Anooj, *Clinical decision support system : Risk level prediction of heart disease using weighted fuzzy rules*, J. King Saud Univ. - Comput. Inf. Sci. 24 (2012) 27–40.
10. N. A. Setiawan, P. A. Venkatachalam and A. F. M. Hani, *Diagnosis of Coronary Artery Disease Using Artificial Intelligence Based Decision Support System*, Man-Machine Systems (ICoMMS), Int. Conf., Penang, Malaysia, 2009, pp. 1–5.
11. S. Muthukaruppan and M. J. Er, *A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease*, Expert Syst. Appl. 39 (2012) 11657–11665.
12. J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*, Upper Saddle River, NJ: Prentice Hall, 2001.