Statistical RBF Network with Applications to an Expert System for Characterizing Diabetes Mellitus

Kyong-Sik Om, Hee-Chan Kim, Byoung-Goo Min, Chan-Soo Shin, and Hong-Kyu Lee

Abstract

The purposes of this study are to propose a network for the characterizing of the input data and to show how to design predictive neural network expert system which doesn't need previous knowledge base. We derived this network from the radial basis function networks(RBFN), and named it as a statistical RBFN. The proposed network can replace the statistical methods for analyzing dynamic relations between target disease and other parameters in medical studies. We compared statistical RBFN with the probabilistic neural network(PNN) and fuzzy logic(FL). And we testified our method in the diabetes prediction and compared our method with the well-known multilayer perceptron(MLP) neural network one, and showed good performance of our network. At last, we developed the diabetes prediction expert system based on the proposed statistical RBFN without previous knowledge base. Not only the applicability of the statistical RBFN to the characterizing of parameters related to diabetes and construction of the diabetes prediction expert system but also wide applicabilities has the proposed statistical RBFN to other similar problems.

I. Introduction

Diabetes mellitus is the most common endocrine disease and one of the major adult diseases in the industrialized countries. It is a dangerous disease because the diabetic patient is susceptible to a series of complications that cause morbidity and premature mortality. Therefore, prevention of the disease is more important and a lot of efforts have been made in characterizing the diabetes mellitus to develop an efficient method predicting the disease. We developed an expert system predicting the diabetes mellitus based on a set of examination data.

Expert system is a successful field of the studies of artificial intelligence. Professor Edward Feigenbaum of Stanford University defined: "An expert system is an intelligent computer program that uses knowledge and inference procedures to solve problems that are difficult enough to require significant human expertise for their solution....[8]." The emphasis of the studies about expert system has shifted from the hand-crafted system to the intelligent one. The former consisted of all the software components and the knowledge base were coded by the developers while the latter is based on the neural network or fuzzy logic [9][10][11][12][13]. It is a common agreement that the fuzzy logic approach is chosen when there exists *a priori* knowledge and the knowledge is represented by linguistic variables. On the other hand, neural network is a better solution when there does not exist *a priori* knowledge and the input data is in numerical form.

Recently, a radial basis function network(RBFN) has been studied intensively because of its attractive intrinsic advantages. There are many reports that RBFNs have very short learning time compared to other networks and have the potentials of on-line and real-time learning [1][2][3]. Also RBFNs have wide range of application fields including clustering [4], control [5], pattern recognition [6], optical flow segmentation [7], etc. In general, it is, however, required that the input data must be well-ranged in order to get the best performance. In other words, if there is little or no correlation

Manuscript received September 18, 1997; accepted March 27, 1998.
Kyong-Sik Om is with the Interdisciplinary Program-Major of Biomedical Engineering, Graduate School, Seoul National University.
Hee-Chan Kim, Byoung-Goo Min are with the Department of Biomedical Engineering, College of Medicine, Seoul National University.
Chan-Soo Shin and Hong-Kyu Lee are with the Department of Internal Medicine, College of Medicine, Seoul National University.

(or tendency) between the input and the output vectors at all, the neural network won't provide its best performance.

Our goal is to develop a disease-predicting expert system. The input data is a set of health examination in numerical form and we don't have *a priori* knowledge, so we chose a neural network approach. The purpose of this study is to propose a new neural network which is adequate especially to a medical expert system with a constraint of no requirement of *a priori* knowledge base. We derived this network from the RBFN, and named it the *statistical RBFN*. It stemmed from the fact that this network was applied to analyzing medical data, where various statistical methods are usually being used for analyzing relationships between input data and the target disease. We compared the characteristics of the statistical RBFN with an existing network of probabilistic neural network (PNN), and fuzzy logic (FL). The proposed network's performance was also evaluated with the data collected for predicting the diabetes mellitus. The overall performance of the proposed network was also compared with well-known multilayer perceptron (MLP).

This paper is organized as follows. First, in section II, the proposed statistical RBFNs are introduced. In section III, we compared the characteristics of the statistical RBFN with the PNN and FL. Application of the proposed network to prediction of diabetes mellitus is introduced in section IV. In section V, performance of the statistical RBFN in predicting diabetes mellitus is summarized in comparison with MLP. The construction of diabetes-predicting expert system is discussed in section VI. Finally, the conclusions are summarized in section VII.

II. Statistical RBF Networks

RBFNs were originally proposed as an interpolation method, and their properties as interpolants have been extensively studied [14]. It is now one of the main research fields in numerical analysis. RBFNs have been shown to have universal approximation ability by Hartman *et al.* [15] and Park and Sandberg [16][17]. Comparison of RBFNs and multilayer perceptrons(MLPs) is well summarized in [2, pp. 262-263]. As a big difference from the RBFNs, MLPs construct *global* approximation to nonlinear input-output mapping. Consequently, they have generalization capabilities in regions of the input space where little or no training data are available. On the contrary, RBFNs construct *local* approximations, with the result that these networks are capable of fast learning and reduced sensitivity to the order of presentation of training data. The core of RBFN is choosing an activation function F that has the following form [14]:

$$F(X)=\sum_{i=1}^{N} W_i\varphi(\|X-C_i\|)+w_0, \tag{1}$$

where $\{ \varphi(\|X-C_i\|) \mid i=1,2, \ldots, N\}$ is a set of N arbitrary (generally nonlinear) functions, known as RBF, and $\| \ \|$ denotes a norm that is usually taken to be Euclidean. The w_i is the weight connected between hidden layer and output layer. If the number of kernels (activation functions) is large enough, the bias term w_0 is not necessary. The known data points $C_i \in R^P$, $i=1,2, \ldots, N$ are taken to be the *centers* of the RBFNs. Fig. 1. shows the structure of RBFN without bias term w_0.

Theoretical investigations and practical results, however, seem to show that the type of nonlinearity $\varphi(\cdot)$ is not crucial to the performance of RBFNs [14]. Some of $\varphi(\cdot)$ are $\varphi(r) = r$ (linear), $\varphi(r) = r^3$ (cubic), $\varphi(r) = (r/\sigma)^2 \ln(r/\sigma)$ (thin-plate-spline function), $\varphi(r) = \exp(-r^2/(2\sigma^2))$ (Gaussian function), $\varphi(r) = \sqrt{r^2+1}$ (multiquadrics), and $\varphi(r) = 1/\sqrt{r^2+1}$ (inverse multiquadrics) [19][20][21][22]. If a function which is monotonically decreasing from it's center of peak such as Gaussian and inverse multiquadrics is used as an activation function, each hidden unit is assumed to have membership-like or statistical relationship to a target output. So in statistical analysis problems, these type of functions are usually used. For the sake of manipulational easiness, Gaussian functions are preferred than inverse multiquadrics. So the Gaussian was selected as an activation function of our new neural network. The theory for the Gaussian radial basis functions was well studied in [23].

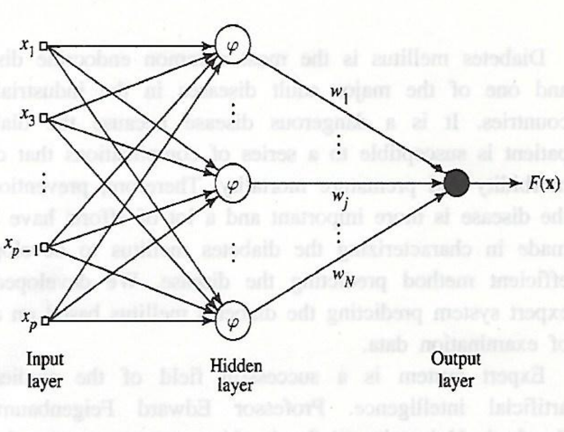


Fig. 1. Structure of the RBFN.

1. *The Structure of Statistical RBFNs*

Basic concept of the proposed statistical RBFN is based on the possibility theory[25,26]. It is because that the output of a disease-predicting network should be, in nature, possibility rather than probability that the examinee would be a patient in the future. Another point is that it is a typical two-class classification problem where two output states (patient or normal) are defined. In this problem, there exist two cases of possibility 0.5 which means that based on the input data, it is hard to say which output state is more possible than the other. One is when the input has equal tendency to both output states. The second case is that the input is in dead zone (no receptive field) which is far from the center of activation function in the Gaussian RBFN.

To solve this problem, the statistical RBFN is proposed. It has normalized output between +1 and -1 where the positive and negative output values represent possibility that the input has more tendency or correlation to one and the other output states, respectively. So the zero output value corresponds to the 0.5 possibility case when input does not belong to either of the output states. In this way, if the proposed network's output characteristics is proportional to the possibility, we can use this network in any statistical problem. The detailed structure and learning process of the proposed statistical RBFN are explained further below.

1) *The input units*: The input vector (input parameter, feature vectors, input data) must be normalized. In general, the range of [0, 1] is usually used. Normalization can be accomplished by two approaches. One is by help of the expert who knows the characteristics of input data well. The other is by analyzing the data histogram and select appropriate levels.

2) *The hidden units*: The hidden units connected to one of the output units must have balanced condition; half of them represent positive tendency and the remaining half shows negative tendency. This condition is necessary to get correct possibility output. If this condition is not satisfied, it will cause a biased output value. The reason for this can be explained by Fig. 2. As shown in Fig. 2-(a), if the sum of positive hidden units has the distribution of e^{-x^2} and that of negative units has $3e^{-(x-1)^2/4}$, the output always produce the negative value. But if we normalize these distributions to have the same area like Fig. 2-(b), we can get correct output, that is the output can be positive or negative like Fig. 2. (c). This situation is frequently happened in medical application where the input data usually has biased population. As an example, in our application of diabetes mellitus, the number of examinee who were not diabetics in 1993 but became patient in 1995 was only 63, but the number who stayed normal was about 1200.

3) *The centers*: Fixed centers selected at random sampling was used in our statistical RBFN. The learning strategies of RBFNs can be categorized into three groups as follows: *fixed centers* selected at random, *self-organized* selection of centers, and *supervised* selection of centers [2]. In many cases (with many hidden units), the method using fixed centers selected at random works well for statistical problem.

4) *The widths*: In an isotropic Gaussian function, its standard deviation is related to the width, the spread of the function from its center. Too narrow width may cause bad generalization capability and on the other hand, too wide width can produce overlapping of large area in neighborhood. It is, therefore, necessary to find an appropriate width value. In our proposed statistical RBFNs, this is the only one parameter to be determined in training procedure. The training algorithm will be explained in detail later.

5) *The weights*: Positive, negative, zero weights means the relation between the hidden and output units have positive, negative and zero correlation depending on the membership of φ, respectively. Typical values are +1, -1 as positive and negative weights, respectively.

6) *The output units*: The output value of each output unit is normalized as follow:

$$o_K = \frac{\sum_i^N W_i \varphi_i}{\sum_i^N \varphi_i} \tag{2}$$

Depending on the correlation between the input data and output state, the output takes positive or negative value. As explained before, zero output value produced by two major causes. One is from the dead zone, and the other from the equal positive and negative tendency. For an example in Fig. 2-(b) and (c), the points at $x = \pm 8$ are in dead zone, and $x = -1.5032$ and $x = 0.8366$ correspond to the equal positive and negative tendency case.

An example of our statistical RBFN is shown in Fig. 3 where three output case is represented. In this case, output nodes are independent, e.g., both output node A and B can produce positive values, then we can interpret it as the input belongs to A and B simultaneously with certain possibilities. The positive tendency between a hidden unit and an output unit is represented by a bold line and negative tendency by a thin line. Since each of three output is a separate two-class classification problem, the width parameters σ_1, σ_2, and σ_3 for A, B, and C differ from each other. Note that each output node consists of equal number of positive and negative tendency.

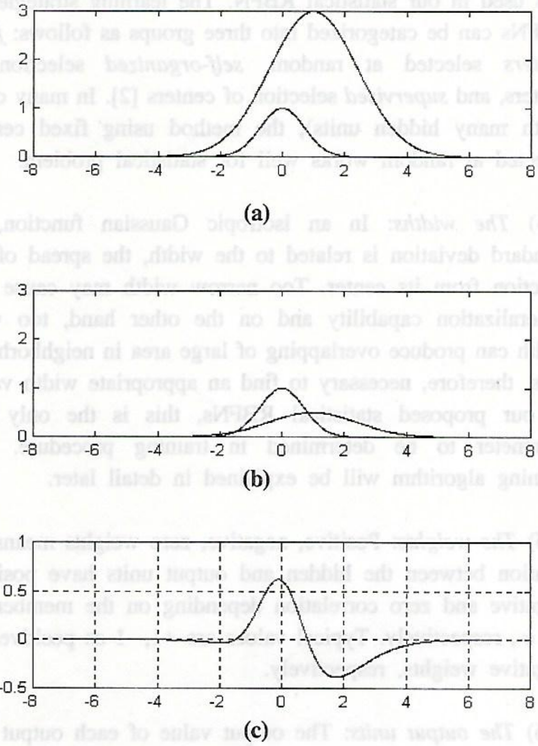


Fig. 2. (a) e^{-x^2} and $3e^{-(x-1)^2/4}$. (b) e^{-x^2} and $0.5e^{-(x-1)^2/4}$. (c) $e^{-x^2}-0.5e^{-(x-1)^2/4}$.

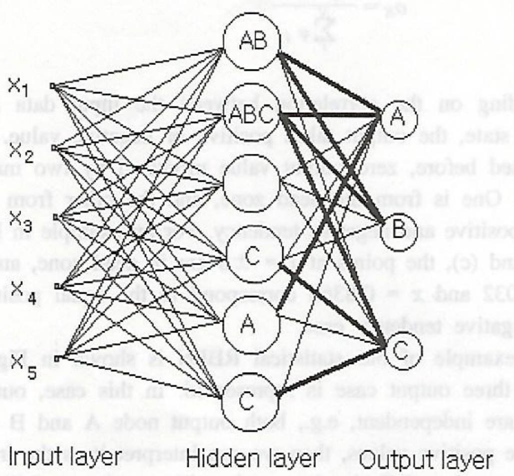


Fig. 3. Example of the statistical RBFN. Input and output dimension is five and three, respectively. Nodes in the hidden layer represent each memberships to the output layer such that the second hidden node (ABC) belongs to every output nodes and the empty node has no relationship to any output node. The bold and thin lines between hidden and output layer represent positive weight (typical value is +1) and negative weight (typical value is -1) , respectively.

2. *Training of the Network*

As explained before, in the proposed statistical RBFN, the only one parameter to be determined in training procedure is the width of the Gaussian. If width σ in the Gaussian RBFN is too small or large, the constructed statistical RBFN has low recognition ratio for the untrained data which means low generalization ability.

Steps for finding an optimal width parameter σ are given as follows :

Step 1. Initial width is set to a small value (σ_0).
Step 2. Calculate the classification ratio of input data (RAT(0)).
Step 3. Calculate the next σ (n) by adding a small value β to previous value $\sigma(n\text{-}1)$: $\sigma(n) = \sigma(n\text{-}1) + \beta$
Step 4. Calculate the overall classification ratio for RAT(n) with given width value.
Step 5. Does σ reach a final target σ value? If it is 'No', go to Step 3. If it is 'Yes', go to Step 6.
Step 6. Find the σ for the highest value of RAT(n).

This method is based on the maximum classification condition. It is also an fast-computing one-pass algorithm. Here, the input data used for netwok training must be different from those used for hidden layer formulation by locating centers of Gaussians in order to increase generalization capability of the network.

III. Comparisons of a Statistical RBFN with PNN and Fuzzy Logic

Possibility theory was introduced by Zadeh [25] in 1978 to model possible events based on the theory of fuzzy logic and fuzzy sets. Both possibility and probability are two techniques for representing and manipulating uncertain or imprecision. While the probabilities have to sum to 1 over the input domain range, the possibility values are not so restricted. High possibility does not necessarily imply high probability, nor does a low degree of probability imply a low degree of possibility. Nevertheless, if the possibility of an event becomes small, then its probability would tend to smaller too, however, the converse is not true. Furthermore, if an event is impossible, it is bound to be improbable. In conclusion, possibility provides a subjective appraisal whereas the probability does an objective one [26].

Proposed statistical RBFN and PNN [24] have different bases. The former is designed to produce possibility and the latter provides probability. To accomplish this object, PNN utilizes Parzen weighting function for the estimation of probability density function (pdf). Following considerations were made to provide the proposed network with possibility characteristics.

Firstly, the statistical RBFN uses membership-like, Gaussian, activation function for the hidden units. Secondly, the normalized output of Eq. (2) has the *possibility* characteristics rather than *probability* ones, e.g., if there exists a hidden unit which produces a very small positive value ε and the other kernels produce a zero one, the final output is $(-1\times0 + -1\times0 + -1\times0 + \ldots + 1\times0 + 1\times0 + 1\times\varepsilon) / (0 + 0 + 0 + \ldots + 0 + 0 + \varepsilon) = 1$ (very high). But in the probabilistic point of view, the output is ε (very low). This property enables us to characterize the input data subjectively. The mathematical derivations of the functional equivalence of RBFNs and fuzzy inference systems are reported in [27][28]. Our statistical RBFN is a special kind of RBFNs. So, as general RBFNs can carry out the characteristics of fuzzy inference systems, proposed statistical RBFN has the same characteristics of fuzzy inference systems. Actually, the normalized output of Eq. (2) corresponds to one of defuzzification methods, namely, *weighted combination method* (*simplified center of gravity method*) in fuzzy reasoning.

IV. Applications to Prediction of Diabetes Mellitus

1. Medical Background

The diagnosis of symptomatic diabetes is not difficult. It can be judged by the presence of diuresis(frequent and large amount of urine output) and hyperglycemia(state of high plasma glucose concentration). But the prediction of the possibility of a health examinee to become diabetic is very hard. It is also a meaningful study to find the relationship between the parameters taken in routine health examination and the possibility to be a diabetic patient.

Until now, there has been no report on the exhaustive and objective studies about these points. So, due to the lack of knowledge bases, it is impossible to make an diabetes predicting expert system with conventional fuzzy logic based methods. Also it is impossible to make an neural network based system since we do not know what factors are meaningful as the inputs of neural network. This problem leads us to think a new network and the statistical RBFN is able to cope with it.

Since we did not have *a priori* knowledge about what factors have correlation with diabetes, we considered all examination parameters. Therefore, the input data is very ill-posed in terms of conventional neural networks. Here, the 'very' means the deviation is very high and it is very possible that the input data might be meaningless for predicting diabetes. The input data was collected from the Yonchon district health survey, in Korea twice in 1993 and 1995. Total population was about 1200. The number of examinees who were not diabetics in 1993 but became patients in 1995 was 63. Examinee whose GLU2 (fasting plasma glucose level) is higher than 140 or GLU3 (plasma glucose level at two hours after meal) is higher than 200 was diagnosed to be a diabetic patient.

2. Description of Data

Total 14 health examination parameters were used in the proposed statistical RBFN and summarized in Table 1. The minimum and maximum values were obtained from the random sample data of 30 normal examinees. The minimum and maximum values were determined by investigating the histogram heuristically, and those shown in the Table 1 are not true minimum and maximum value. These values were used later as normalizing indices for the input of neural network, i.e., if one input value exceeds these extremes, it is set to be these values.

Table 1. Parameters under considerations.

Parameters	Meaning [unit]	Minimum value	Maximu Value
1. GOT	Glutamic-oxaloacetic transaminase	10	25
2. GPT	Glutamic-pyruvic transaminase	5	20
3. CHO	Cholesterol	120	170
4. TG	Density of neutral fat	80	250
5. HDL	Density of High Density Lipoprotein cholesterol	20	40
6. GLU2	Fasting plasma Glucose level	85	110
7. GLU3	Plasma Glucose level two hours after meal	70	120
8. AGE	Age [Year]	30	70
9. SBP	Systolic blood pressure [mmHg]	100	140
10. DBP	Diastolic blood pressure [mmHg]	65	85
11. HEI	Height [cm]	150	170
12. WEI	Weight [Kg]	55	70
13. BMI	Body mass index = Weight / (Hight2) [Kg/m2]	20.0	25.0
14. WHR	Waist to hip circumference ratio	0.80	0.90

V. Performance of the Statistical RBFN in Predicting Diabetes Mellitus

1. Results Comparison between the Proposed Statistical RBFN and MLP Network

The proposed statistical RBFN is constructed with 14 input units as parameters shown in Table 1 and 1 output unit. The input units represent the parameters to be analyzed, and the output unit represents the possibility value with which the specific examinee might be a diabetic patient or remained normal. As an activation function, the modified Gaussian kernel defined as follows was used;

$$\varphi(r) = \exp(-Kr^2). \tag{3}$$

where $1/(2\sigma^2)$ is simply replaced by K.

Now, choosing the width parameter in training procedure is changed to choosing K value. In Step 3 of width determining procedure, the variable β is determined as follows. In the range of $K < 0.1$, β is set 0.01, and in the range of $K > 0.1$, β is set 0.1. In Step 6, the optimal K was determined based on the classification results. The total number of input data sample was 126 which consisted of same number of 63 normal and patient cases to satisfy the constraint of same number of positive and negative tendency. The total 126 data samples were allocated into three procedures; 1) network construction by determining hidden nodes as the center position of Gaussain functions, 2) training the network by selecting the optimal width K, and 3) test the network's performance by evaluating the classification results. Table 2, 3 and 4 show the performance of the proposed network with different allocation of the input data samples. Allocation condition is represented by the number of the input data sample used for each proceudre 1), 2) and 3) as (62:32:32), (74:26:26) and (86:20:20) for Table 2, 3 and 4, respectively. These allocation conditions approximately correspond to (2:1:1), (3:1:1) and (4:1:1), respectively. In selecting optimal width value, it is recommended to choose K which provides nearly equally maximum classification score for both width determining data set and test data set. It is because there exists no correlation between the classification results for two different input data sets. Data1 and Data2 marked in the tables represent data set used in width determining procedure and testing data set, respectively. Every case shows that too wide or too narrow width parameter bring poor results as expected.

As a result, the input data allocation condition of (3:1) in Table 3 produce better results than other two conditions in terms of comparable classification score for both data sets of Data1 and Data2. So we can conclude that if we divide the given input data as the ratio of (3:1:1), and the 3/5 portion of data were used for hidden units and 1/5 for width parameter, then the classification score for both data set of unused in training and part of training will be comparable each other. Then, we selected the optimal value of K as 1.6 from Table. This decision was made since it is the midpoint value of K within the range of 0.9-2.3 where the classification score for Data1 was maximum in Table 3. The resultant classification score is summarized in Table 5 with the finalized conditions. One important point to note is that the classification score is not important in this problem. Final classification score of about 70% is very high when considering that it is very hard for the conventional neural networks to cope with these kinds of data. This fact can be verified by testing MLP network with the same data set.

For the comparison, we have also constructed the well-known MLP network. The MLP network has 14 input units, 74 hidden units, and 1 output unit. This corresponds to the condition of Table 3 in the statistical RBFN. The MLP network was set to have learning rate 0.1 and momentum term 0.7. The input parameters were also normalized between 0 and 1 as they were used in the statistical RBFN. The output was also normalized between 0 and 1, which represent normal and patient, respectively. The number of data used for network learning was 100 (50 for patient and 50 for normal). A statistical RBFN in Table 3 also uses 100 data for construction 74(37 patient + 37 normal) for hidden units 26(13 patients +13 normal) for choosing width parameter. After long period of iteration, MLP did not converge at all. After 30000 learning steps, we stopped the learning on purpose and testified the classification score. Then the output of the MLP for the learning data were either 0.499 or 0.501. This means that it's output oscillates and the learning process is meaningless. It is an expected result because there are almost no correlation between input data and output. The recognition ratio of the MLP for the untrained data - 13 data for patients and 13 data for normal persons - was also 50[%]. It again confirms the fact that the input data are very much ill-posed and these are meaningless parameters to predict output.

Table 2. Number of correct classification with number of data for (Hidden units : Width parameters) = (62:32) $\simeq$ (2:1)

K	Data 1 (32)	Data 2 (32)	K	Data 1 (32)	Data 2 (32)	K	Data 1 (32)	Data 2 (32)	K	Data 1 (32)	Data 2 (32)
0.01	16	15	0.1	25	23	2.1	25	21	4.1	25	17
0.02	17	18	0.2	26	21	2.2	25	21	4.2	24	17
0.03	22	18	0.3	26	21	2.3	25	21	4.3	24	17
0.04	25	19	0.4	27	21	2.4	25	21	4.4	24	17
0.05	25	20	0.5	27	21	2.5	25	20	4.5	24	16
0.06	25	21	0.6	27	21	2.6	25	20	4.6	24	16
0.07	25	21	0.7	27	21	2.7	25	18	4.7	24	16
0.08	25	22	0.8	27	21	2.8	25	18	4.8	24	16
0.09	25	22	0.9	27	22	2.9	25	18	4.9	24	16
			1.0	27	22	3.0	25	18	5.0	24	16
			1.1	27	22	3.1	25	18	5.1	24	16
			1.2	27	22	3.2	25	18	5.2	24	16
			1.3	27	22	3.3	25	17	5.3	24	16
			1.4	26	22	3.4	25	17	5.4	24	16
			1.5	26	22	3.5	25	17	5.5	24	16
			1.6	26	21	3.6	25	17	5.6	24	16
			1.7	26	21	3.7	25	17	5.7	24	16
			1.8	26	21	3.8	25	17	5.8	24	16
			1.9	25	21	3.9	25	17	5.9	24	16
			2.0	25	21	4.0	25	17	6.0	24	16

Table 3. Number of correct classification with number of data for (Hidden units : Width parameters) = (74:26) ≃ (3:1)

K	Data 1 (32)	Data 2 (32)	K	Data 1 (32)	Data 2 (32)	K	Data 1 (32)	Data 2 (32)	K	Data 1 (32)	Data 2 (32)
0.01	13	12	0.1	19	18	2.1	20	18	4.1	17	16
0.02	14	13	0.2	20	18	2.2	20	18	4.2	17	16
0.03	17	15	0.3	19	18	2.3	20	18	4.3	17	16
0.04	19	16	0.4	19	18	2.4	19	18	4.4	17	16
0.05	19	17	0.5	19	18	2.5	19	18	4.5	17	16
0.06	19	18	0.6	19	18	2.6	19	18	4.6	17	16
0.07	19	18	0.7	19	18	2.7	19	17	4.7	17	16
0.08	19	18	0.8	19	18	2.8	19	17	4.8	17	15
0.09	19	18	0.9	20	19	2.9	19	17	4.9	17	15
			1.0	20	19	3.0	19	17	5.0	17	15
			1.1	20	19	3.1	19	17	5.1	17	15
			1.2	20	19	3.2	19	17	5.2	17	15
			1.3	20	18	3.3	18	16	5.3	17	15
			1.4	20	18	3.4	17	16	5.4	17	15
			1.5	20	18	3.5	17	16	5.5	17	15
			1.6	20	18	3.6	17	16	5.6	17	15
			1.7	20	18	3.7	17	16	5.7	17	15
			1.8	20	18	3.8	17	16	5.8	17	15
			1.9	20	18	3.9	17	16	5.9	17	15
			2.0	20	18	4.0	17	16	6.0	17	15

Table 4. Number of correct classification with number of data for (Hidden units : Width parameters) = (86: 20) ≃ (4:1)

K	Data 1 (32)	Data 2 (32)	K	Data 1 (32)	Data 2 (32)	K	Data 1 (32)	Data 2 (32)	K	Data 1 (32)	Data 2 (32)
0.01	10	9	0.1	12	15	2.1	12	15	4.1	10	13
0.02	9	13	0.2	12	16	2.2	12	15	4.2	10	13
0.03	10	14	0.3	12	16	2.3	12	15	4.3	10	13
0.04	11	14	0.4	12	16	2.4	12	15	4.4	10	13
0.05	11	14	0.5	12	16	2.5	12	15	4.5	10	13
0.06	11	15	0.6	12	16	2.6	12	15	4.6	10	13
0.07	12	14	0.7	12	16	2.7	12	15	4.7	10	13
0.08	12	14	0.8	12	16	2.8	12	15	4.8	10	12
0.09	12	15	0.9	12	16	2.9	12	15	4.9	10	12
			1.0	12	16	3.0	12	15	5.0	10	12
			1.1	12	16	3.1	12	15	5.1	10	12
			1.2	12	16	3.2	12	15	5.2	10	12
			1.3	12	16	3.3	12	15	5.3	10	12
			1.4	12	16	3.4	12	14	5.4	10	12
			1.5	12	15	3.5	12	14	5.5	10	12
			1.6	12	15	3.6	12	14	5.6	10	12
			1.7	12	15	3.7	12	14	5.7	10	12
			1.8	12	15	3.8	12	14	5.8	10	12
			1.9	12	15	3.9	12	14	5.9	10	12
			2.0	12	15	4.0	12	14	6.0	10	12

Table 5. Classification ratio of the proposed statistical RBFN with K=1.6 in Table 3

Input Data Set		Classification percentage of statistical RBFN [%]
Data 1	Patients	9 / 13 = 69.2
	Normal Persons	11 / 13 = 84.6
	Total Persons	20 / 26 = 76.9
Data 2	Patients	10 / 13 = 76.9
	Normal Persons	8 / 13 = 61.5
	Total Persons	18 / 26 = 69.2
Average total persons		38 / 52 = 73.1

2. Ordering of Input Parameters in Sequence of High Correlation to Diabetes Mellitus

In this subsection, we examined the relations between 14 health examination parameters and diabetes mellitus. This was performed as a kind of simulation study by manipulating the input parameters to provide hypothetical situations. It is very interesting and important study for characterizing the disease. First of all, an imaginary normal input parameter set was established. We selected maximum and minimum values of each input parameter as two values equally spaced from the center of the normal group's histogram. Then all input parameters were normalized with these extreme values. Therefore, by setting all input parameter value as 0.5, a hypothetical normal data set was achieved. When all input values were 0.5(hypothetical normal examinee), the output of the proposed statistical RBFN with Table 3 condition and K=1.6 was -0.2134. This result proves again the plausibility of the established network since normal input data should bring negative output.

Next, we manipulated the input parameter set as each parameter was set to 1 and -1, maximum and minimum value, respectively, one by one while all other 13 parameters were kept unchanged as 0.5. The results are summarized in Table 6. Positive output represents that the corresponding input parameter change caused the network output shifted toward patient side and negative output means more shift to normal side from -0.2134, a hypothetical normal output point. In Table 7, the most influential 9 input parameters to diabetes are listed in diminishing order. This results were used in the next section for the construction of an expert system for diabetes prediction.

Table 6. Network output with manipulated input parameter set.

Parameters	Max case	Min case
1. GOT	0.0613	-0.0479
2. GPT	0.0048	-0.0049
3. CHO	-0.0609	-0.0336
4. TG	0.0758	-0.0672
5. HDL	0.0664	-0.0583
6. GLU2	0.1040	-0.0926
7. GLU3	0.0657	-0.0577
8. AGE	0.1200	-0.1853
9. SBP	0.0875	-0.1605
10. DBP	0.1262	-0.2014
11. HEI	-0.0263	-0.0541
12. WEI	-0.0521	-0.0366
13. BMI	-0.0560	-0.0393
14. WHR	0.1009	-0.1873

Table 7. Ordering of paremeters based on Table 6.

Rank number	Parameter	Correlation value	Rank number	Parameter	Correlation value	Rank number	Parameter	Correlation value
1	DBP	0.1262	4	WHR	0.1009	7	HDL	0.0664
2	AGE	0.1200	5	SBP	0.0872	8	GLU3	0.0657
3	GLU2	0.1009	6	TG	0.0758	9	GOT	0.0613

3. *Characteristics of the Proposed Statistical RBFNs*

In this subsection, we summarize the characteristics of our proposed statistical RBFNs.

1) The learning procedure requires only one parameter adjustment. It is the width of Gaussian an activation function used in the hidden units. It is an one-pass algorithm.

2) The output unit produces normalized output ranging from -1 to +1 not from 0 to +1.

3) The number of hidden units with weight +1 must equal to that of hidden units with weight -1.

4) Number of data used for (hidden units : width parameter) $\approx$ (3 : 1) is a reasonable ratio for the optimal construction of the proposed statistical RBFN. This requires further examinations.

5) In the statistical RBFN, output nodes are independent, e.g., if both output node A and B produce positive value no matter which is bigger, it means that the input belongs to A and B simultaneously.

6) Our statistical RBFN is based on the possibility theory rather than probability theory.

VI. Construction of an Expert System for Diabetes Prediction

An expert system to predict diabetes mellitus is a class of disease predicting medical expert system. We construct this expert system by limiting the input parameters of the statistical RBFN based on the result in previous section. It was accomplished by selecting only 9 input parameter listed in Table 7. We call it 'the enhanced statistical RBFN' as compared to the standard statistical RBFN. The rationale of this selection and elimination of input parameters is that other parameters than shown in Table 7 do not contribute to change the normal state to patient at all. Another step added in expert system construction is a decision threshold *DT* for the final decision of disease based on the network output value. The role of *DT* is to give an interface zone of 'Unknown' rather than a sharp edge between 'Patient' and 'Normal'. As a performance evaluation of the proposed expert system, the diabetes prediction score of the statistical RBFN with full input parameters and enhanced statistical RBFN was compared at 6 different threshold levels. The tested results are summarized in Table 8. Both cases bring nearly same prediction ratio (RP (Relative Prediction ratio) = (number of correctly predicted cases) / (number of total cases - number of predicted as unknown cases), AP (Absolute Prediction ratio) = (number of correctly predicted cases) / (number of total cases). But as the *DT* increases, the RP and AP of the enhanced statistical RBFN increases more rapidly and decreases more slowly, respectively, than the standard statistical RBFN. This implies that we can get the higher performance and short calculation time due to small number of input parameters with the enhanced statistical RBFN. This result also shows that the range of 0.1 to 0.2 is appropriate for *DT*.

Table 8. Performance results of the constructed expert system (RP : Relative prediction, AP : Absolute prediction, R : Correct prediction, U : Unknown, *DT* : Decision threshold)

DT	Full parameters					
	Patients		Normal		RP [%]	AP [%]
	R	U	R	U		
0	20	0	18	0	38 / 52 = 73.1	38 / 52 = 73.1
0.05	18	2	17	1	35 / 49 = 71.4	35 / 5 = 67.3
0.10	18	3	17	2	35 / 47 = 74.5	35 / 52 = 67.3
0.15	16	5	16	4	32 / 43 = 74.4	32 / 52 = 61.5
0.20	15	6	16	5	31 / 41 = 75.6	31 / 52 = 59.6
0.25	15	6	15	6	30 / 40 = 75.0	30 / 52 = 57.7

DT	Correlated parameters					
	Patients		Normal		RP [%]	AP [%]
	R	U	R	U		
0	18	0	18	0	36 / 52 = 69.2	36 / 52 = 69.2
0.05	18	2	18	1	36 / 49 = 73.5	36 / 52 = 69.2
0.10	18	3	17	2	35 / 47 = 74.5	35 / 52 = 67.3
0.15	18	4	17	3	35 / 45 = 77.8	35 / 52 = 67.3
0.20	18	4	17	4	35 / 44 = 79.5	35 / 52 = 67.3
0.25	15	7	15	6	30 / 39 = 76.9	30 / 52 = 57.7

VII. Conclusions

The main purpose of this study is to propose a new neural network to characterize diabetes mellitus by analyzing the relationship between the input health examination data and the disease. Based on the proposed neural network, it was also attempted to show how to design predictive neural network based expert system which doesn't require previous knowledge base. We derived this network from the RBFN, and named it as the statistical RBFN. The proposed network can replace the conventional statistical methods for analyzing relationship between a target disease and health examination parameters. Since it is based on the possibility theory, the proposed statistical RBFN is expected to be applied to almost all statistical problems.

We compared the characteristics of the statistical RBFN with the PNN and the FL. The performance of the proposed network was verified by comparing diabetes prediction capabilities with the well-known MLP neural network. The input data was collected through a large scaled health survey in Yonchon district area performed twice in 1993 and 1995. Since there exists no remarkable correlation between input health examination data and the disease, the result using conventional network (MLP) was poor as expected, but the proposed network produced relatively good classification score. We also ranked the input health examination parameter according to the influence on the diabetes mellitus. This was performed as a kind of simulation study by manipulating the input parameters to provide hypothetical situations. It may be interpreted as a list of the risk factor of diabetes mellitus, which can be utilized in prevention of the disease.

Finally, we developed the diabetes prediction expert system by enhancing the proposed statistical RBFN. Enhancement was performed in two aspects. Firstly, out of 14 health examination parameters, only 9 which shows higher correlation to the disease were used. A decision threshold *DT* was also incorporated to finally predict the disease. With these two preparations, we designed an expert system with improved results. It is important because any previous knowledge base was not used.

Future researches must be concentrated on the handling of linguistic data rather than the numerical one. We think this is possible because the proposed statistical RBFN has close relationship with fuzzy logic system. Also, a method of finding vector-type width parameter of Gaussian activation function, which is expected to enhance the performance in the hidden units must be studied.

Acknowledgment

This study was supported by a grant (HMP-95-G-2-24) of the '95 Highly Advanced National Project on the Development of Biomedical Engineering and Technology, Ministry of Health and Welfare, R.O.K. and a grant number: 02-97-220 from the Seoul National University Hospital Research Fund.

References

[1] J. Moody and C. Darken, "Fast learning in networks of locally-tuned processing units," *Neural Computa.*, vol. 1, no. 2, pp. 281-294, 1989.

[2] S. Haykin, *Neural Networks*, Macmillan College Publishing Company, Inc. 1994.

[3] M. Heiss and S. Kampl, "Multiplication-free radial basis function network," *IEEE Trans. Neural Networks*, vol. 7. No. 6, pp. 1461-1464, Sept. 1996.

[4] S. V. Chakravarthy and J. Ghosh, "Scale-based clustering using the radial basis function network," *IEEE Trans. Neural Networks*, vol. 7. No. 5, pp. 1250-1261, Sept. 1996.

[5] L. Behera, M. Gopal, and S. Chaudhury, "On adaptive trajectory tracking of a robot manipulator using inversion of its neural emulator," *IEEE Trans. Neural Networks*, vol. 7. No. 6, pp. 1401-1414, Nov. 1996.

[6] M. Rosenblum Y. Yacoob, and L. S. Davis, "Human expression recognition from motion using a radial basis function network architecture," *IEEE Trans. Neural Networks*, vol. 7. No. 5, pp. 1121-1138, Sept. 1996.

[7] A. G. Bors and I. Pitas, "Median radial basis function neural network," *IEEE Trans. Neural Networks*, vol. 7. No. 6, pp. 1351-1364, Nov. 1996.

[8] A. Barr and E. A. Feigenbaum, *The Handbook of Artificial Intelligence*, vol. I, Villiam Kaufmann.

[9] P. L. Choong, C. J. S. deSilva, H. J. S. Dawkins, and G. F. Sterrett, "Entropy maximization networks: An application to breast cancer prognosis," *IEEE Trans.*

Neural Networks, vol. 7, no. 3, pp. 568-577, May. 1996.

[10] L. M. Fu, "Learning capacity and sample complexity on expert networks," *IEEE Trans. Neural Networks*, vol. 7, no. 6, pp. 1517-1520, Nov. 1996.

[11] A. Kandel, *Fuzzy Expert Systems*, CRC press, 1994.

[12] D. Cayrac, D. Dubois, and H. Prade, "Handling uncertainty with possibility theory and fuzzy sets in a satellite fault diagnosis application," *IEEE Trans. Fuzzy Systems*, vol. 4, no. 3, pp. 251-269, Aug. 1996.

[13] C. Rodriguez, S. Rementeria, J. I. Martin, A. Lafuente, J. Muguerza, and J. Perez, "A modular neural network approach to fault diagnosis," *IEEE Trans. Neural Networks*, vol. 7, no. 2, pp. 326-340, March, 1996.

[14] M. J. D. Powell, "The theory of radial basis function approximation in 1990," *Advances in Numerical Analysis*, vol. 3, W. Light, Ed. Oxford : Clarendon, pp. 105-210, 1992.

[15] E. J. Hartman, J. D. Keeler, and J. M. Kowalski, "Layered neural networks with Gaussian hidden units as universal approximation," *Neural Computa.*, vol. 2, pp. 210-215, 1990.

[16] J. Park and I. W. Sandberg, "Universal approximation using radial basis function Networks," *Neural Computa.*, vol. 3, pp. 246-257, 1991.

[17] J. Park and I. W. Sandberg, "Approximation and radial basis function networks," *Neural Computa.*, vol. 5. pp. 305-316, 1993.

[18] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 1996.

[19] B. A. Whitehead, T. D. Choate, "Cooperative-competitive genetic evolution of radial basis function centers and widths for time series prediction," *IEEE Trans. Neural Networks*, vol. 7, no. 4, pp.869-880, July, 1996.

[20] D. S. Broomhead and D. Lowe, "Multivariable function interpolation and adaptive networks," *Complex Syst.*, vol. 2, pp. 321-355, 1988.

[21] T. Musavi, W. Ahmed, K. H. Chan, K. B. Faris, and D. M. Hummels, "On the training of radial basis function classifier," *Neural Networks*, vol. 5, pp. 595-603, 1992.

[22] Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Networks*, vol. 2, no. 2, pp. 302-309, 1991.

[23] T. Poggio, and R. Girosi, "Networks for approximation and learning," *Proceedings of the IEEE*, Vol. 78, No. 9, pp. 1481-1497, Sept. 1990.

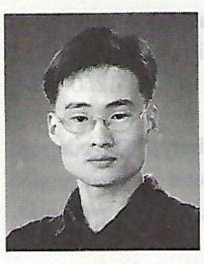
[24] F. Specht, "Probabilistic neural networks," *Neural Networks*, vol. 3, pp. 109-118, 1992.

[25] L. A. Zadeh, "Fuzzy sets as basis for theory of possibility," *Fuzzy Sets and Systems*, vol. 1, pp. 3-28, 1978.

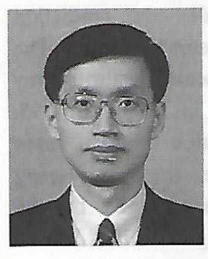
[26] M. Schneider, A. Kandel, G. Langholz, and G. Chew, *Fuzzy Expert System Tools*, John Wiley and Sons, 1996.

[27] K. J. Hunt, R. Haas, and R. M. Smith, "Extending the functional equivalence of radial basis function networks and fuzzy inference systems," *IEEE Trans. Neural Networks*, vol. 7, no. 3, pp. 776-781, May 1996.

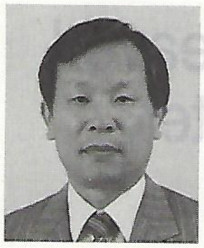
[28] J. S. R. Jang and C. T. Sun, "Functional equivalence between radial basis function networks and fuzzy inference systems," *IEEE Trans. Neural Networks*, vol. 4, no. 1, pp. 156-159, Jan. 1993.



Kyong-Sik Om was born in Korea, in 1971. He received the B.S. and M.S. degrees in Electronic Engineering from Inha University, Inchon, Korea, in 1994 and 1996, respectively. Since 1996, he has working toward the Ph.D. degree in Biomedical Engineering at Seoul National University, Seoul, Korea. His interest fields are total artificial heart, biomedical signal processing systems, rehabilitation engineering, alternative medicine, fuzzy sets and fuzzy logic, neural networks, genetic algorithm, computer vision, speech recognition, emotional engineering, expert system, nonlinear signal processing (order statistic filter, fuzzy filter, ...) digital image processing, walking, painting, reading, traveling, listening of musin, poem, etc. He is a student member of IEEE, IEEK, KOSOMBE, ICASE, and KFIS.



Hee-Chan Kim was born in Korea, in 1959. He received the B.S. and M.S. degrees in Electronics Engineering in 1982 and 1984, respectively, and Ph.D. degree in Control and Measurement Engineering in 1989 from Seoul National University, Seoul, Korea. Since 1991 when he joined as a faculty member, he is an assistant professor in Department of Biomedical Engineering, College of Medicine, Seoul National University, Seoul, Korea. His research interest is application of electronics to medical and biological engineering.



Byoung-Goo Min was born in Korea, in 1943. He received the B.S. degree in Electrical Engineering from Seoul National University, Seoul, Korea, in 1965, and M.S. and Ph.D. degrees from Rutgers University, USA, in 1970 and 1972, respectively,. Since 1979, he has been a chairman and professor in Dept. of Biomedical Engineering, Seoul National University Hospital, Seoul, Korea. His research interest is artificial heart.

Hong-Kyu Lee was born in Korea, in 1944. He received the M.D. degree in 1968 and M.S. and Ph.D. degrees in Medicine in 1971 and 1978, respectively, all from Seoul National University, Seoul, Korea. Since 1976 when he joined as a faculty member, he is a chairman and professor and in Department of Internal Medicine-Endocrinology Division, College of Medicine, Seoul National University, Seoul, Korea. His research interest includes computer application in medicine.

Chan-Soo Shin was born in Korea, in 1962. He received the M.D. degree in 1987 and M.S. and Ph.D. degrees in Medicine in 1990 and 1998, respectively, all from Seoul National University, Seoul, Korea. Since 1995 when he joined as a faculty member, he is an full-time lecturer in Department of Internal Medicine-Endocrinology Division, College of Medicine, Seoul National University, Seoul, Korea. His research interest includes computer application in medicine.