

# Evaluating the Severity of Dementia of Alzheimer's Disease by a Characteristic-Point-Based Fuzzy Inference System

Tang-Kai Yin and Nan-Tsing Chiu

## Abstract

**Cognitive ability screening instrument (CASI) is a cross-cultural mental screening test to evaluate the severity of dementia of Alzheimer's disease. In this research, a characteristic-point-based fuzzy inference system (CPFIS) is proposed to predict the CASI scores from the single-photon emission computed tomography (SPECT) volumes. Experiment results showed that CPFIS could provide a rough estimation of CASI scores that the P value for the hypothesis that the average of absolute errors of CPFIS is less than the average of absolute differences from the mean values was 0.0447. For comparison, a weighted support vector regression (SVR) machines was also tried. The performance of CPFIS was slightly better than SVR in this study (P=0.068).**

**Keywords:** *Alzheimer's disease, CASI, Computer-aided diagnosis, Fuzzy inference system, SPECT.*

## 1. Introduction

Alzheimer's disease is a chronic degenerative disease of the central nervous system. The symptoms of Alzheimer's disease are progressive loss of memory, cognitive function, and ultimately death. Clinically early detection of Alzheimer's disease is helpful in taking care of the patients. Fuzzy methods have been applied in the research of Alzheimer's disease. Kobashi et al. use fuzzy inferences to automatically segment the cerebrospinal fluid and the lateral ventricles from brain MR (magnetic resonance) images [1]. Sato et al. apply fuzzy K-means to perform image diagnosis with respect to frontal lobe atrophy [2]. Wang et al. use fuzzy measure in the regional rates of brain atrophy to early diagnose Alzheimer's disease. [3]

The nuclear imaging methods, single-photon emission computed tomography (SPECT) is a useful tool in analyzing hypoperfusion in patients with Alzheimer's disease. Most SPECT studies measure regional cerebral blood flow. Common regional abnormalities for Alzheimer's disease are symmetric or asymmetric bilateral temporal or parietal hypoperfusion [4].

Cognitive ability screening instrument (CASI) is a cross-cultural mental screening test including long-term memory (RMM), short-term memory (RCM), attention (ATT), concentration/mental manipulation (MEN), orientation (ORI), abstraction and judgment (ABS), language (LAN), visual construction (DRA), and list-generation fluency (ANM) [5]. In evaluating the severity of dementia, the sensitivity and specificity of CASI are higher than those of another screening test, MMSE (Mini-Mental State Examination).

Ushijima et al. employed SPECT to study the relations between the cognitive functions and regional cerebral blood flow [6]. Their research showed that the decline in regional cerebral blood flow in the parietal cortex and hippocampus reflected disorientation, and the most significant cortex affecting scores on each section of the MMSE were found to be the anterior temporal cortex for registration, the frontal cortex for attention and calculation, the medial temporal cortex for recall, and the posterior temporal cortex for language. Eberling et al. also demonstrated that MMSE was related to the regional cerebral blood flow in the frontal and parietal lobes [7]. Wolfe et al. used the cerebral blood flow of SPECT in the temporal lobe to predict the stage of dementia of Alzheimer's disease [8].

In this research, a characteristic-point-based fuzzy inference system (CPFIS) is proposed to predict the CASI scores from the SPECT volumes. The aim of the CPFIS is not only satisfactory precision performance, but also to employ as few purely linguistic fuzzy rules as possible by using a minimization-based systematic training method. Characteristic points (CPs) are defined as the few data points among the original training data which, when they are directly mapped to fuzzy rules and thus form the entire rule base, allow the underlying system to be effectively modeled. Three minimization-based algorithms in a sequence are proposed to train the CPFIS: a gradient-projection method [9], a Gauss-Jordan-elimination-based column

---

Corresponding Author: Tang-Kai Yin is with the Department of Computer Science and Information Engineering, National University of Kaohsiung, 700, Kaohsiung University Rd., Kaohsiung, Taiwan, 811. E-mail: tkyin@nuk.edu.tw.

Nan-Tsing Chiu is with the Department of Nuclear Medicine, College of Medicine, National Cheng Kung University, 138, Sheng Li Rd., Tainan, Taiwan.

Manuscript received 28 Aug. 2003; revised 1 Dec. 2003; accepted 21 June. 2004.

elimination [10], and back-propagation [11]. The CPs are determined by iterative computations of the first two minimization algorithms, after which the resulting fuzzy sets are further fine-tuned by the third algorithm.

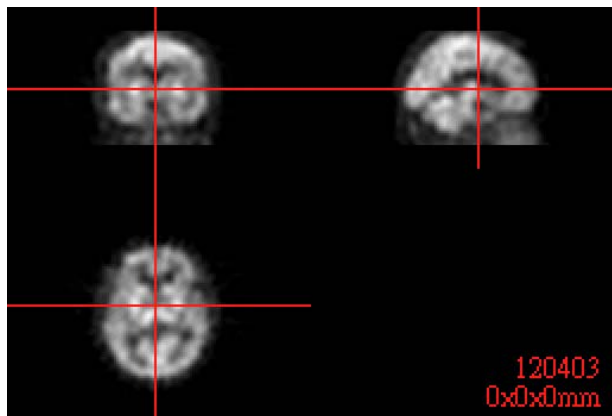


Figure 1. SPECT volume of a patient of Alzheimer's disease.

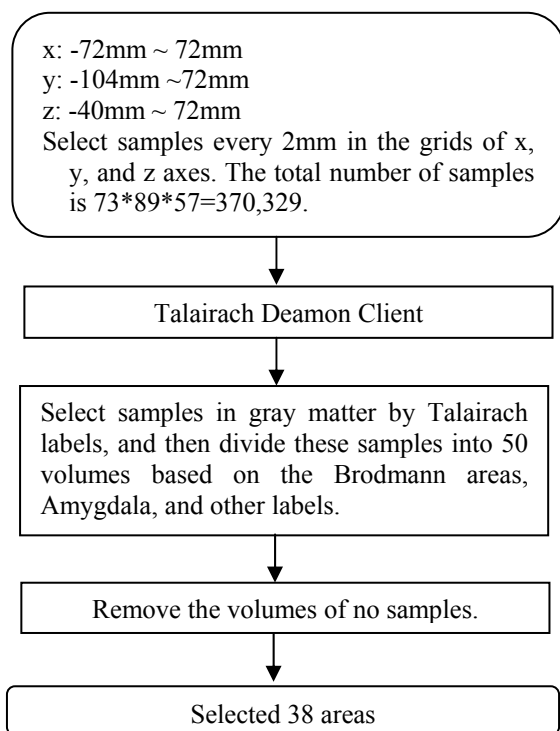


Figure 2. Process of selecting the 38 areas from the SPECT volumes.

This paper is organized as follows. The selection of system input variables from SPECT volumes is presented in Section 2. The structure and training of CPFIS is described in Section 3. Experiment results and discussions are given in Section 4. Finally, we state our conclusion in Section 5.

Table 1. Talairach labels and numbers of samples in the left and the right brains of the thirty-eight selected areas. BA: Brodmann area, GP: Globus Pallidus

Label	Left	Right	Label	Left	Right
BA 3	5	2	BA 31	17	22
BA 4	1	1	BA 32	13	9
BA 5	6	11	BA 34	4	1
BA 6	75	74	BA 37	11	11
BA 7	84	54	BA 38	17	17
BA 8	60	28	BA 39	20	14
BA 9	60	64	BA 40	44	51
BA 10	54	75	BA 41	1	2
BA 11	19	14	BA 44	8	3
BA 13	10	18	BA 46	25	18
BA 17	5	4	BA 47	4	7
BA 18	29	36	Amygdala	33	33
BA 19	36	41	Hippocampus	5	3
BA 20	6	3	Lateral GP	33	39
BA 21	8	14	Medial GP	8	9
BA 22	9	9	Putamen	283	289
BA 24	19	10	Thalamus	472	471
BA 25	3	1	Caudate	181	176
BA 30	1	1	Cerebellum	699	685

## 2. System Input Variables

There were 58 brain technetium-99m hexamethyl-propylene amine oxime (99m-Tc-HMPAO) SPECT volumes of the patients of Alzheimer's disease in this study. The ages of the 58 patients (22 male, and 36 female) were between 46 and 88 years old, with the average and standard deviation being 73.8 and 7.9, respectively. These volumes were from the National Cheng Kung University Hospital, Tainan, Taiwan, R.O.C. The imaging device was a triple-headed rotating gamma camera (Multispect3; Siemens, Hoffman Estates, USA).

### A. Selecting Thirty-Eight Areas Based on Talairach Labels

Fig. 1 shows a SPECT volume of a patient of Alzheimer's disease. The top left, top right, and bottom left are coronal, sagittal, and transverse views of the brain, respectively. The first step in estimating CASI scores from SPECT volumes is to select many areas of interest, and calculate the averages of these areas as the system inputs. We consider gray matter of brain in this study. Fig. 2 shows the process of selecting the 38 areas from the SPECT volumes. In Talairach space, the 2mm grids between  $x = -72\text{mm}$  and  $x = 72\text{mm}$ ,  $y = -104\text{mm}$  and  $y = 72\text{mm}$ ,  $z = -40\text{mm}$  and  $z = 72\text{mm}$  are taken as samples. The total number of these samples is  $73 \times 89 \times 57 = 370,329$ .

We use the Talairach Daemon Client Version 1.1 (Research Imaging Center, University of Texas Health Science Center at San Antonio) to obtain the Talairach labels of the selected samples. Consider Brodmann areas 1~11, 13, 17~25, 27~47, Amygdala, hippocampus,

lateral globus pallidus, medial globus pallidus, putamen, thalamus, caudate, and cerebellum. If any top, down, left, right, front, or rear sample next to one sample is not inside an area, it is taken as a boundary sample. To eliminate the effect of area boundary samples in calculating area averages, the boundary samples are not considered. After this elimination, some areas contain no samples. These areas are not considered. After these selecting steps, there are 38 selected areas. Table 1 shows the Talairach labels and numbers of samples in the left and the right brains of the 38 selected areas.

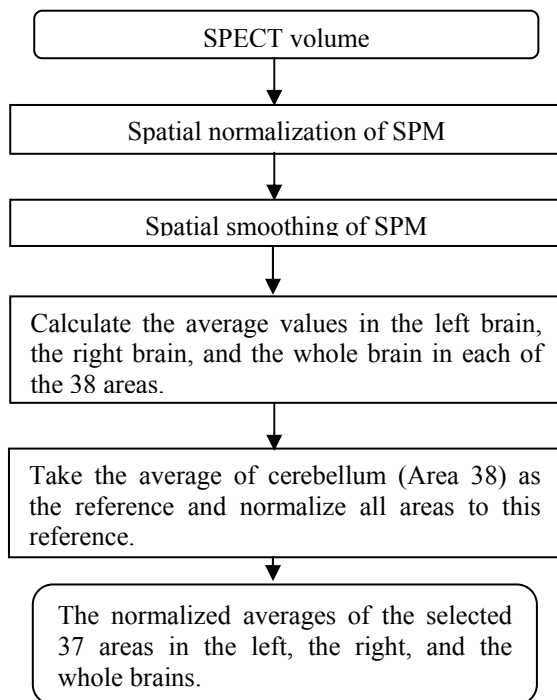


Figure 3. Normalized averages of the selected 37 areas.

### B. Calculating Thirty-Seven Area Ratios of the Chosen Brodmann Areas

Each SPECT volumes was stereotactically normalized (12 parameters affine transformation) to a standard SPECT template of 79\*95\*68 voxels in the Montreal Neurological Institute (MNI) stereotactic space, and then smoothed with the FWHM (Full Width Half Maximum) being set to be 6mm in all three dimensions by the SPM99 (Statistical Parametric Mapping) software from the Wellcome Department of Cognitive Neurology, London, UK (<http://www.fil.ion.ucl.ac.uk/spm>) [12].

Since SPM99 uses MNI brain, the coordinates of Brodmann areas are slightly different from those of Talairach space. We used the MATLAB function **tal2min.m** (<http://www.mrc-cbu.cam.ac.uk/Imaging>) to map the coordinates of Talairach space to those of MNI space. Because cerebellum is spared with Alzheimer's disease [13], the obtained samples were further normalized to the average voxel value of cerebellum.

Fig. 3 summarizes the process of deriving the normalized averages of the selected 37 areas.

Table 2. P values of the null hypotheses that there are no correlations between the nine items of CASI and Talairach areas. The areas of  $P < 0.05$  are listed. One or two more areas are listed for RCM, ABS, LAN, and DRA in order to increase input variables of CPFIS. Bold and underlined: CPFIS input variable

CASI	Label	All	Left	Right
RMM	BA 21	.0497	<b><u>.0082</u></b>	-
	BA 37	.0333	<b><u>.0019</u></b>	-
	Lateral GP	-	<b><u>.032</u></b>	-
	Medial GP	-	.0372	<b><u>.0127</u></b>
RCM	BA 17	-	-	<b><u>.0004</u></b>
	BA 22	-	-	<b><u>.0973</u></b>
ATT	BA 20	<b><u>.0068</u></b>	.0081	-
	BA 21	<b><u>.0333</u></b>	-	-
	BA 22	-	<b><u>.0256</u></b>	-
	BA 38	<b><u>.0278</u></b>	.0316	.0434
	BA 44	-	<b><u>.0065</u></b>	-*
MEN	Lateral GP	-	-	<b><u>.0381</u></b>
	BA 17	-	<b><u>.0183</u></b>	-
ORI	BA 18	-	<b><u>.0287</u></b>	-
	BA 21	.0268	-	<b><u>.0189</u></b>
	Amygdala	.044	-	<b><u>.0378</u></b>
ABS	Hippocampus	-	-	<b><u>.0332</u></b>
	BA 17	-	<b><u>.1839</u></b>	-
LAN	Medial GP	-	<b><u>.1275</u></b>	-
	BA 20	.019	<b><u>.0068</u></b>	-
	BA 21	-	<b><u>.0775</u></b>	-
DRA	BA 30	.0343	<b><u>.0088</u></b>	-
	BA 31	-	<b><u>.0798</u></b>	-
ANM	BA 25	.0463	.0494	<b><u>.0347</u></b>
	BA 32	.0317	<b><u>.0114</u></b>	-
	BA 34	.0056	<b><u>.0054</u></b>	-
	BA 47	.0437	-	<b><u>.0119</u></b>

### C. Areas Linearly Correlated with CASI

After calculating the normalized averages of the selected 37 areas in the left, the right, and the whole brains, there are  $37 \times 3 = 111$  values. On the other hand, there are 9 CASI items, which have been corrected with respect to age and education. Readers are referred to [5] for details. We employed the MATLAB correlation coefficient function **CORRCOE** to calculate the correlation coefficients between the 9 CASI items and the 111 area averages for the 58 patients of Alzheimer's disease.

Let the linear correlation coefficient be  $r$ . We used the (7.48) equation of [14]

$$t_{n-2} = r / [(1 - r^2) / (n - 2)]^{1/2}$$

to test the null hypothesis  $H_0: \rho = 0$  and two-sided alternative  $H_1: \rho \neq 0$ . The value  $t$  is the Student  $t$  distribution. For example, we want to calculate the relationship between the Brodmann Area 21 in the left brain and the RMM scores. There are  $n$  SPECT volumes and the corresponding  $n$  CASI scores. For

each volume, we have one average value of the Brodmann Area 21 in the left brain, and one RMM score for one patient. Use **CORRCOEF** to calculate the correlation between the  $n$  values from the Brodmann Area 21 in the left brain, and the  $n$  RMM scores. Denote the correlation as  $r$ . From  $n$  and  $r$ , we have the Student  $t$  value, and the corresponding P value, 0.0082, as shown in Table 2. Table 2 shows the P values of the null hypotheses  $H_0$  that there are no correlations between the nine items of CASI and Talairach areas. The areas of  $P < 0.05$  are listed. One or two more areas are listed for RCM, ABS, LAN, and DRA in order to increase input variables of CPFIS. There are 4, 2, 6, 2, 3, 2, 2, 2, 4 variables for RMM, RCM, ATT, MEN, ORI, ABS, LAN, DRA, and ANM of the 9 CASI items, respectively. We built a CPFIS for each of these 9 CASI items as discussed in the next section.

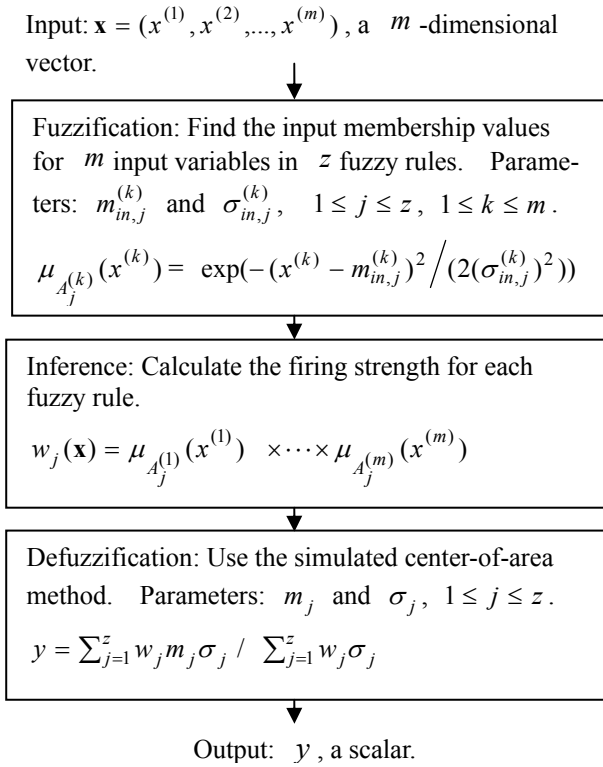


Figure 4. Flowchart of making a fuzzy inference of CPFIS.

### 3. Characteristic-Point-Based Fuzzy Inference System

There are 9 CPFISs for the 9 CASI items representing the input-output relationship for each of them. Each CPFIS is independently trained and inferred after training. The discussions in this section are applicable to each of the 9 CPFISs. For simplicity, the notations are not further indexed to distinguish these CPFISs here,

but it is kept in mind that their values are different in the 9 CPFISs.

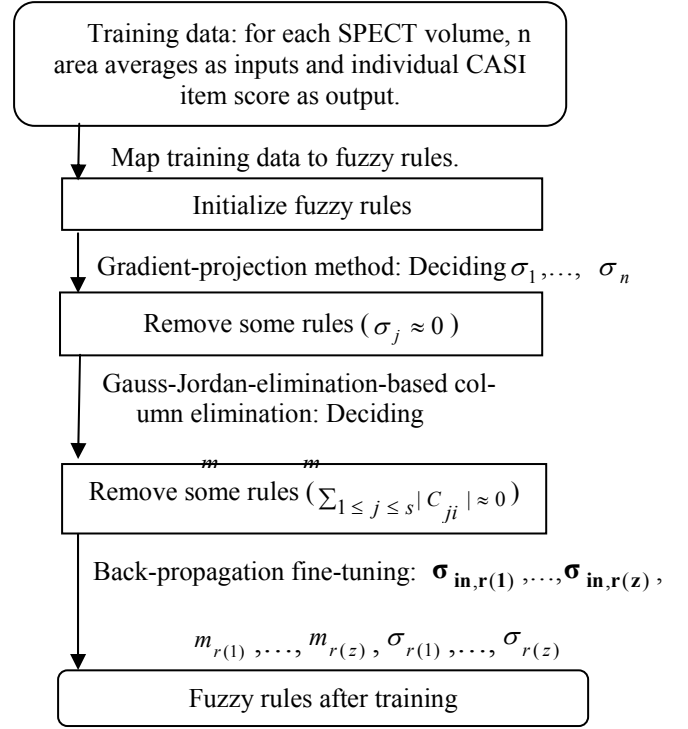


Figure 5. Training process of CPFIS.

#### A. Rule Base

Fig. 4 shows the flowchart of making a fuzzy inference of CPFIS, consisting of fuzzification, calculation of firing strengths, and defuzzification. Let  $z$  be the number of fuzzy rules and  $m$  be the dimension of the input  $\mathbf{x}$ . The rule base of a CPFIS is

if  $x^{(1)}$  is  $A_j^{(1)}$  and ... and  $x^{(m)}$  is  $A_j^{(m)}$ ,  
then  $y$  is  $B_j$ ,

where  $A_j^{(k)}$  and  $B_j$ ,  $1 \leq j \leq z$ ,  $1 \leq k \leq m$ , are fuzzy sets in the antecedent and consequent parts of fuzzy rules. The membership function of  $A_j^{(k)}$  is a bell-shaped function with center  $m_{in,j}^{(k)}$  and spread  $\sigma_{in,j}^{(k)}$ :

$$\mu_{A_j^{(k)}}(x^{(k)}) = \exp\left(-\frac{(x^{(k)} - m_{in,j}^{(k)})^2}{2(\sigma_{in,j}^{(k)})^2}\right).$$

We call  $(m_{in,j}^{(1)}, \dots, m_{in,j}^{(m)})$ ,  $1 \leq j \leq z$ , as characteristic points (CPs). The membership function of  $B_j$  is also chosen as a bell-shaped function with center  $m_j$  and spread  $\sigma_j$ :

$$\mu_{B_j}(y) = \exp\left(-\frac{(y - m_j)^2}{2\sigma_j^2}\right).$$

To make an inference of a CPFIS, calculate the firing strength  $w_j(\mathbf{x}) = \mu_{A_j^{(1)}}(x^{(1)}) \times \dots \times \mu_{A_j^{(m)}}(x^{(m)})$ ,  $1 \leq j \leq z$ , for each fuzzy rule. Then defuzzify the output fuzzy sets by using the simulated center-of-area method:

$$y = \sum_{j=1}^z w_j m_j \sigma_j / \sum_{j=1}^z w_j \sigma_j.$$

The training of CPFIS is to decide the number of fuzzy rules, and the parameters of these rules. The following three minimization steps are proposed to perform the training of CPFIS.

## B. Training

The training of CPFIS is to decide the number of fuzzy rules, and the parameters of these rules. Fig. 5 shows the proposed three minimization steps to perform the training of CPFIS. The details of the algorithms are listed in the Appendix.

### B.1. Gradient-Projection Method

Initially, all training data are mapped to fuzzy rules. The mapping is performed on all data points  $(\mathbf{x}_i, y_i)$ ,  $\mathbf{x}_i \in \mathfrak{R}^m$ ,  $y_i \in \mathfrak{R}$ , that  $x_i^{(1)}, \dots, x_i^{(m)}$  are assigned to be the means of the membership functions of the input fuzzy sets, and  $y_i$  are assigned to be the means of the membership functions of the output fuzzy sets. That is,

$$m_{in,i}^{(t)} = x_i^{(t)}, \quad t = 1, \dots, m, \quad i = 1, \dots, n,$$

$$m_i = y_i, \quad i = 1, \dots, n.$$

Each data point is mapped to a fuzzy rule. Thus, if there are  $n$  data points, then initially there are  $n$  fuzzy rules. After the mapping, the spreads of the membership functions of the input fuzzy sets,  $\sigma_{in,j}$ , and the spreads of the membership functions of the output fuzzy sets,  $\sigma_j$ , remain to be set. We set

$$\sigma_{in,j}^{(t)} = \frac{1}{a} (\max_{1 \leq i \leq n} x_i^{(t)} - \min_{1 \leq i \leq n} x_i^{(t)}), \quad t = 1, \dots, m,$$

where  $a$  is a chosen constant. In this paper,  $a$  is set to be 2 or 3.  $\sigma_{in,j}$  are the same for all fuzzy rules in this step. They will be different and fine-tuned in the later back-propagation process. The firing strength of a rule  $j$  is

$$A(\mathbf{x}, \mathbf{x}_j) = \exp\left(-\frac{(x^{(1)} - m_j^{(1)})^2}{2(\sigma_{in,j}^{(1)})^2} - \dots - \frac{(x^{(m)} - m_j^{(m)})^2}{2(\sigma_{in,j}^{(m)})^2}\right),$$

and the fuzzy inference output is

$$y = \sum_{j=1}^n y_j \sigma_j A(\mathbf{x}, \mathbf{x}_j) / \sum_{p=1}^n \sigma_p A(\mathbf{x}, \mathbf{x}_p).$$

The weights  $\sigma_j$ ,  $j = 1, \dots, n$ , are initially set to be

$1/n$  before the minimization.

After these settings, we can take the training of CPFIS as a constrained minimization problem:

$$\min_{\sigma_1, \dots, \sigma_n} \sum_{i=1}^n \left[ y_i - \frac{\sum_{j=1}^n y_j \sigma_j A(\mathbf{x}_i, \mathbf{x}_j)}{\sum_{p=1}^n \sigma_p A(\mathbf{x}_i, \mathbf{x}_p)} \right]^2 \quad (1)$$

$$\text{subject to } \sum_{j=1}^n \sigma_j = 1, \quad \sigma_j \geq 0, \quad j = 1, \dots, n.$$

A gradient projection method is employed to solve this problem. It is noted that (1) is a general nonlinear equation in the variables  $\sigma_1, \dots, \sigma_n$ . Thus, the solution obtained by the gradient-projection method is usually a local minimum.

After the algorithm, many constraints  $\sigma_j \geq 0$  become active constraints, i.e.,  $\sigma_j \approx 0$ . The approximation “ $\approx$ ” used here is to take into account the precision of numerical calculations. In the experiments, we label a constraint as an active constraint if  $\sigma_j$  is less than a small positive number. The training data of these active constraints can be removed from being candidates of CPs, since the weights,  $\sigma_j \approx 0$ , of these fuzzy rules are much smaller than those of inactive constraints. It is from the approximation

$$\sum_{i=1}^n \left[ y_i - \frac{\sum_{j=1}^n y_j \sigma_j A(\mathbf{x}_i, \mathbf{x}_j)}{\sum_{p=1}^n \sigma_p A(\mathbf{x}_i, \mathbf{x}_p)} \right]^2 \approx$$

$$\sum_{i=1}^n \left[ y_i - \frac{\sum_{j=1}^s y_{r(j)} \sigma_{r(j)} A(\mathbf{x}_i, \mathbf{x}_j)}{\sum_{p=1}^s \sigma_{r(p)} A(\mathbf{x}_i, \mathbf{x}_{r(p)})} \right]^2$$

where  $r(\cdot)$  is a function that indicates  $r(\cdot)$  is the numbering of the  $j$ th fuzzy rules in the original  $n$  fuzzy rules. Thus, the indexes of active constraints are not in the output domain of  $r(\cdot)$ .

### B.2. Gauss-Jordan-Elimination-Based Column Elimination

Suppose the number of fuzzy rules is after some rules are removed in the last step. The following is a minimization of a quadratic function:

$$\min_{m_{r(1)}, \dots, m_{r(s)}} \sum_{i=1}^n \left[ y_i - \frac{\sum_{j=1}^s m_{r(j)} \sigma_{r(j)} A(\mathbf{x}_i, \mathbf{x}_{r(j)})}{\sum_{p=1}^s \sigma_{r(p)} A(\mathbf{x}_i, \mathbf{x}_{r(p)})} \right]^2. \quad (2)$$

Let

$$B(i, j) = \frac{\sigma_{r(j)} A(\mathbf{x}_i, \mathbf{x}_{r(j)})}{\sum_{p=1}^s \sigma_{r(p)} A(\mathbf{x}_i, \mathbf{x}_{r(p)})}.$$

Then, (2) can be written as

$$\min_{m_{r(1)}, \dots, m_{r(s)}} \sum_{i=1}^n \left[ y_i - \sum_{j=1}^s B(i, j) m_{r(j)} \right]^2 =$$

$$\min_{m_{r(1)}, \dots, m_{r(s)}} f(m_{r(1)}, \dots, m_{r(s)}).$$

The necessary condition for the solution is that the first derivatives of  $f$  are zero. Taking derivatives of and reformulate the equation, we have

$$\begin{bmatrix} \sum_{i=1}^n B(i,1)B(i,1) & \dots & \sum_{i=1}^n B(i,s)B(i,1) \\ \sum_{i=1}^n B(i,1)B(i,2) & \dots & \sum_{i=1}^n B(i,s)B(i,2) \\ \vdots & \vdots & \vdots \\ \sum_{i=1}^n B(i,1)B(i,s) & \dots & \sum_{i=1}^n B(i,s)B(i,s) \end{bmatrix} \begin{bmatrix} m_{r(1)} \\ m_{r(2)} \\ \vdots \\ m_{r(s)} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i B(i,1) \\ \sum_{i=1}^n y_i B(i,2) \\ \vdots \\ \sum_{i=1}^n y_i B(i,s) \end{bmatrix}. \quad (3)$$

By using the Gauss-Jordan elimination in solving linear equations, the minimization can be solved [9]. Suppose in the Gauss-Jordan elimination of  $[C_1 \dots C_{i-1} C_i C_{i+1} \dots] \mathbf{m} = \mathbf{D}$ , the process has been done for the first  $i-1$  columns. Suppose the sum of the absolute values of the elements of column  $i$  is less than a small specified threshold  $C_{threshold}$ . Thus,  $\sum_{1 \leq j \leq s} |C_{ji}| < C_{threshold}$ . Suppose  $C_{threshold}$  is very small, then  $\sum_{1 \leq j \leq s} |C_{ji}| \approx 0$ . Thus, in the solution of (3),  $m_{r(i)}$  will be very big. Since  $m_{r(i)}$  is the center position of the output membership function of rule  $r_i$ , we would like  $m_{r(i)}$  to be in the neighborhood of the output range of the training data. Hence, we remove column  $i$ . The fuzzy rule according to this column can be removed to avoid the redundancy in the above C matrix. Then, we have  $z$  number of fuzzy rules.

### B.3. Back-Propagation Fine-Tuning

The third minimization is based on  $\sigma_{\mathbf{in}, \mathbf{r}(1)}, \dots,$

$$\sigma_{\mathbf{in}, \mathbf{r}(z)}, m_{r(1)}, \dots, m_{r(z)}, \sigma_{r(1)}, \dots, \sigma_{r(z)} : \min_{\sigma_{m, r(1)}, \dots, \sigma_{m, r(z)}, m_{r(1)}, \dots, m_{r(z)}, \sigma_{r(1)}, \dots, \sigma_{r(z)}} \sum_{i=1}^n \left[ y_i - \frac{\sum_{j=1}^z y_{r(j)} \sigma_{r(j)} A(\mathbf{x}_i, \mathbf{x}_{r(j)})}{\sum_{p=1}^z \sigma_{r(p)} A(\mathbf{x}_i, \mathbf{x}_{r(p)})} \right]^2 \quad (4)$$

Back-propagation method can be found in [11]. Readers are referred to [15] for more detailed discussions of CPFIS.

### B.4. Discussions

The proposed three minimizations are based on an idea of learning the number of fuzzy rules and then tuning the parameters in these fuzzy rules. It is a general method which can be applied to other input-output relations. We have demonstrated in [15]

that the proposed CPFIS have the smallest number of fuzzy rules compared with many known methods on benchmark problems by experiments. However, no rigorous proofs are provided regarding the operations of these three minimizations until now. We show the proposed methods work, but need more research effort to know their exact properties.

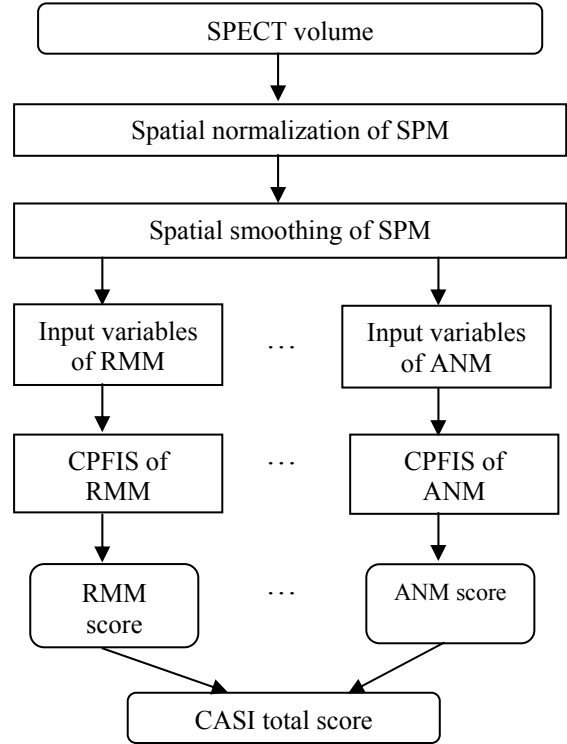


Figure 6. Process of estimating the CASI scores from the SPECT volumes.

In the gradient-projection method, initially all the training data are mapped to fuzzy rules by direct assignments. Then the first minimization is used to remove many fuzzy rules. The running time of this constrained minimization depends on the number of training data. If the number of training data is more than 500, it takes a long running time. Therefore, the size of the training data is limited to about 500 in the CPFIS.

The firing strength  $w_j(\mathbf{x}) = \mu_{A_j^{(1)}}(x^{(1)}) \times \dots \times \mu_{A_j^{(m)}}(x^{(m)})$  can be very small for all  $w_j(\mathbf{x})$  if the data  $\mathbf{x}$  is not near any fuzzy rules. Then in the defuzzification,  $y = \sum_{j=1}^z w_j m_j \sigma_j / \sum_{j=1}^z w_j \sigma_j$ , all the weights  $w_j$  in the numerator and denominator are very small, causing numerical problems in calculations. Therefore, if a testing sample is not in the domain of the trained fuzzy rules, the defuzzification may cause

problems. At these testing samples, the CPFIS cannot make inferences. Thus, the domain of training samples must be large enough to cover all the testing samples.

Both the gradient-projection method and back-propagation tuning are based on steepest descent. Hence, the solutions of the minimizations are usually local minima. It results in suboptimal fuzzy rule bases and suboptimal tuned parameters. For back-propagation, we can try different learning rates and momentum constants as in the Appendix to get better performance [11]. However, local-minimum problems cannot be solved. Genetic algorithm may be one solution, but it is not pursued in this research.

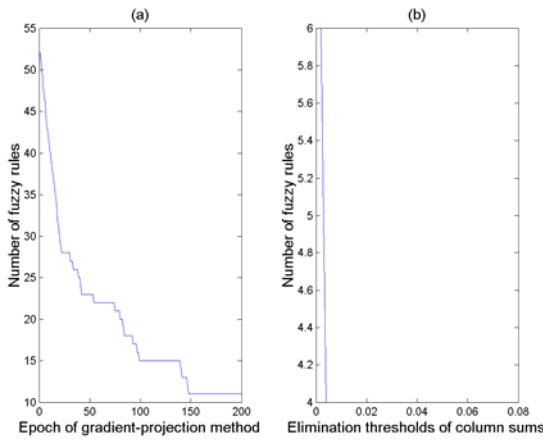


Figure 7. The number of candidate fuzzy rules decreased during (a) the gradient-projection iteration, and further during (b) the Gauss-Jordan-elimination-based column elimination for the CPFIS of MEN in the first validation of the ten-fold cross validation.

### 4. Results

To reduce noise in the training data, a smoothing is performed. Denote

$$d^{(t)} = \left( \max_{1 \leq i \leq n} x_i^{(t)} - \min_{1 \leq i \leq n} x_i^{(t)} \right), \quad t = 1, \dots, m.$$

Then, the new sample output is

$$y(\mathbf{x}) = \text{average of } y(\mathbf{x}_i), \text{ for all } \mathbf{x}_i \text{ that}$$

$$\left| x_i^{(1)} - x^{(1)} \right| < d^{(1)} / 3, \dots, \left| x_i^{(t)} - x^{(t)} \right| < d^{(t)} / 3.$$

To make the evaluation of the proposed diagnosis method more independent of individual selections of training and testing volumes, a ten-fold cross validation was performed. The 58 Alzheimer's volumes were randomly divided into ten subsets. First, the first subset was taken as testing samples and the other samples were training samples. Next, the second subset was taken as testing samples and the other samples were training ones. The procedure was repeated until the tenth subset was done.

Fig. 6 shows the process of estimating the CASI scores from the SPECT volumes. Each testing volume has 9 item scores from the 9 trained CPFISs for RMM, ..., and ANM, respectively. Using the mean  $mean_{RMM,i,j}$  and standard deviation  $SD_{RMM,i,j}$ , the RMM score is

$$RMM * SD_{RMM,i,j} + mean_{RMM,i,j}$$

where i and j are the age class and education class [5]. The other items are similarly calculated. The total CASI score is then obtained from these 9 CPFISs.

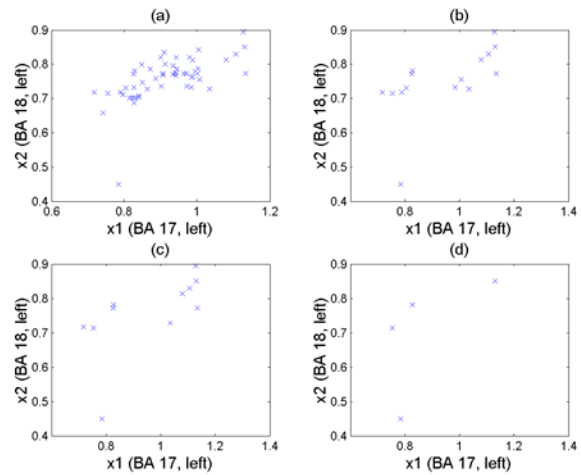


Figure 8. The candidate data points to be CPs are gradually reduced for the CPFIS of MEN in the first validation of the ten-fold cross validation. (a) Original training data. (b) After 100 epochs of the gradient-projection iteration. (c) After 200 epochs of the gradient projection iteration. (d) After the Gauss-Jordan-elimination-based column elimination.

#### A. Training of CPFIS

There are 90 CPFISs in the experiments. All of them are trained in the same way. To show the process of finding the CPs and the fuzzy rules, the results of the CPFIS of MEN in the first validation of the ten-fold cross validation are presented here. The results of all the other 89 CPFISs are similar.

Fig. 7(a) shows the number of inactive constraints during the iterations of the gradient projection algorithm performed on the 52 training data. The initial values of the spreads of the output fuzzy membership function  $\sigma_j$  were 1/52. The threshold of  $\sigma_j$  was set at 0.0001, which is less than  $0.01 * (1/52)$ . That is, when the weighting of a fuzzy rule is less than 0.01 of the average weighting, its contribution to the defuzzification is much smaller (0.01) than the average rules, this fuzzy rule is removed. Initially, all 52 training data were candidates for characteristic points. After 200 epochs of the gradient projection algorithm, 11 training data were still candidates. Fig. 8 shows their distribution. Fig. 8(a) shows the original 52 training data; Fig. 8(b), the

remaining candidates after 100 epochs of the gradient-projection algorithm; and Fig. 8(c), the remaining candidates after 200 epochs of the algorithm.

Starting with the threshold of  $C_{threshold} = 0.001$ , the threshold of column sums was doubled at each iteration until reaching  $64 \times 0.001 = 0.064$ . Fig. 7(b) shows the gradual reduction of candidate fuzzy rules in the column-elimination process. After 7 iterations, four training data remained. These four points were taken as CPs. Fig. 8(d) shows their distribution. The four points are separately located among the original training data in Fig. 8(a).

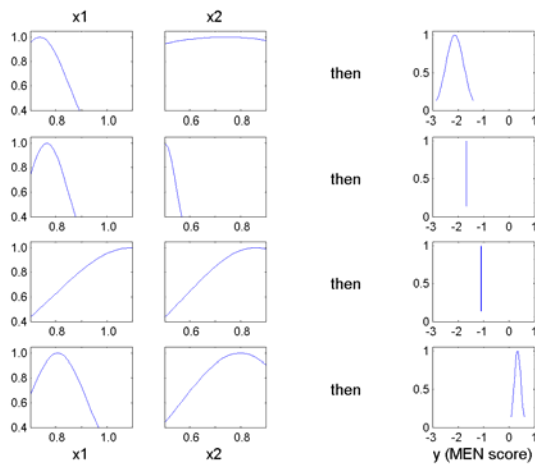


Figure 9. After training, there were four fuzzy rules for the CPFIS of MEN in the first validation of the ten-fold cross validation. Each row is a fuzzy rule.

The back-propagation learning was performed to tune the other parameters. They are

- $4 \times 2 = 8$  parameters of  $\sigma_{in,r(j)}$ , the spreads of the membership functions of input fuzzy sets,
- $4 \times 1 = 2$  parameters of  $m_{r(j)}$ , the means of the membership functions of output fuzzy sets, and
- $2 \times 1 = 2$  parameters of  $\sigma_{r(j)}$ , the spreads of the membership functions of output fuzzy sets, respectively.

The learning was performed with 20,000 epochs, taking 13 seconds of computation. Fig. 9 shows the four tuned fuzzy rules. The rule two and the rule three are fuzzy singletons. Due to their much smaller weights than those of the rules one and four, they are less important in making inferences. Rule one is the representative of low outputs, while rule four is the representative of high outputs from the centers of the output membership functions in Fig. 9. The output

centers of rule one and rule four are -1.69 and -1.37, respectively. Let  $p$  be  $(-1.69 + -1.37)/2$ . Then, the samples can be divided into two groups by the output values as  $y_j \leq p$  and  $y_j > p$ . Fig. 10 shows the plot of the two groups. It is clear that the two CPs are the representatives of local regions around them.

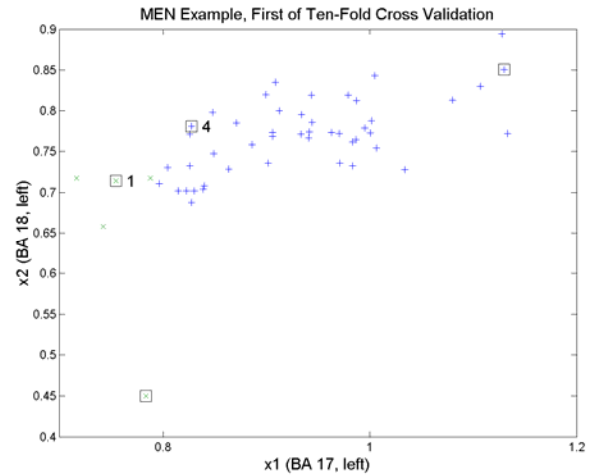


Figure 10. Four CPs are the four representatives of training samples in the first validation of the ten-fold cross validation. Squares: CPs; (+) and (x): the output is greater or smaller than the average of the outputs of rule one and rule four, respectively. Rule one and rule four are also labeled.

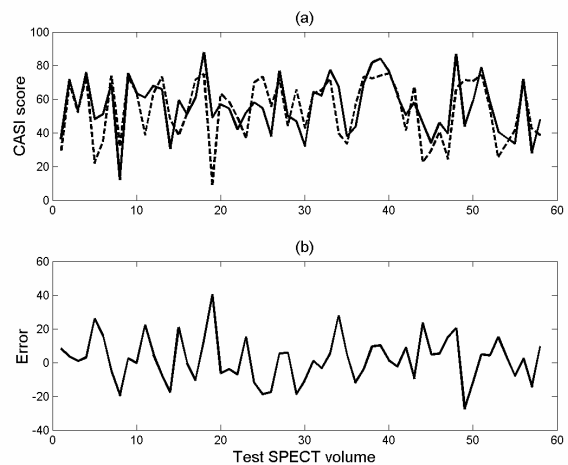


Figure 11. Testing result of the CASI total scores using the ten-fold cross-validation. (a) Solid: true CASI scores, dashed: estimated CASI scores by the CPFIS. (b) Absolute values of the differences between both.

The use of Gaussian membership functions, instead of triangular or trapezoidal membership functions, is based on the ease and availability of the derivative operations in the gradient-projection iteration and the back-propagation. The interpretation of fuzzy rules through Gaussian membership functions can also roughly made. As shown in Fig. 9, the output fuzzy

sets of the second and the third fuzzy rules are fuzzy singletons. They come from local patterns for one or few training samples. For interpretation, rule 1 and rule 4 can be described as

If  $x_1$  (Brodmann Area 17, left brain) is *low*, then MEN score is *low*.

If  $x_1$  (Brodmann Area 17, left brain) is *slightly low*, and  $x_2$  (Brodmann Area 18, left brain) is *high*, then MEN score is *low*.

Table 3. Averages of absolute errors of CPFIS. Bold and underlined: P<0.05 CASI item. CASI item with \*: CPFIS better than SVR (comparing with Table 4)

CASI item	Averages of absolute errors of CPFIS	Averages of absolute differences from the mean values	P values (averages of absolute errors of CPFIS < averages of absolute differences from the mean values)
RMM*	1.168	1.373	0.1327
RCM	2.255	2.335	0.4000
ATT*	1.167	1.428	0.0751
MEN*	2.068	2.941	<b>0.0016</b>
ORI*	4.741	4.111	-
ABS*	1.498	1.361	-
LAN*	1.278	1.684	<b>0.0375</b>
DRA*	1.999	2.779	<b>0.0066</b>
ANM*	1.596	1.719	0.297
Total*	10.61	13.45	<b>0.0447</b>

Table 4. Averages of absolute errors of SVR. Bold and underlined: P<0.05 CASI item

CASI item	Averages of absolute errors of SVR	Averages of absolute differences from the mean values	P values (averages of absolute errors of SVR < averages of absolute differences from the mean values)
RMM	1.228	1.373	0.2115
RCM	2.183	2.335	0.3178
ATT	1.324	1.428	0.3361
MEN	2.125	2.941	<b>0.0029</b>
ORI	4.822	4.111	-
ABS	1.502	1.361	-
LAN	1.310	1.684	0.0511
DRA	2.005	2.779	<b>0.0061</b>
ANM	1.627	1.719	0.3487
Total	11.19	13.45	0.0924

Here the linguistic term *low*, *slightly low*, *medium*, *slightly high*, and *high* are employed for roughly linguistic description along with the graphical display of membership functions. If precise definitions of terms are desired, the fuzzy rules must be modified which may

deteriorate the performance. For example, in rule 1, the values of the membership function of  $x_2$  (Brodmann Area 18, left brain) are near one over the universal set of such that it can be ignored in the above linguistic description for simplicity. Rules 1 and 4 clearly describe the input-output relationship between SPECT volumes and the MEN score, an interesting result to radiologists. To be more significant, the number of training samples must be more than those used in this research, say, more than 100 or more. Nonetheless, the proposed CPFIS provides a fuzzy mechanism in relating SPECT volumes to CASI items.

Table 5. Averages of absolute errors of neural networks. Bold and underlined: P<0.05 CASI item

CASI item	Averages of absolute errors of SVR	Averages of absolute differences from the mean values	P values (averages of absolute errors of SVR < averages of absolute differences from the mean values)
RMM	1.430	1.373	-
RCM	2.328	2.335	0.4914
ATT	1.301	1.428	0.2725
MEN	2.157	2.941	<b>0.0039</b>
ORI	5.121	4.111	-
ABS	1.556	1.361	-
LAN	1.438	1.684	0.1604
DRA	2.038	2.779	<b>0.0075</b>
ANM	1.969	1.719	-
Total	11.26	13.45	0.0943

## B. Performance

Fig. 11 shows the testing result of the CASI total scores in the ten-fold cross-validation. The solid line in Fig. 11(a) is the true CASI scores, while the dashed line is the estimated CASI scores by the CPFIS. The absolute errors are shown in Fig. 11(b). Table 3 lists the average absolute errors of CPFIS. Calculate the P value for the hypothesis that the average of absolute errors of CPFIS is less than the average of absolute differences from the mean values. The P values for MEN, LAN, DRA, and total CASI score are less than 0.05. The differences are statistically significant. For comparison, a weighted support vector regression (SVR) machine [16] was tried. The result is listed in Table 4. Except RCM, CPFIS performs slightly better than SVR in this study.

From Tables 3 and 4, we can compare the performances of CPFIS and SVR in the averages of absolute errors in the RMM, RCM, ATT, MEN, ORI, ABS, LAN, DRA, ANM, and total scores. Use the Student's t hypothesis testing that CPFIS is less than SVR. The degree of freedom is 9, the t value is -1.64,

and the  $P=0.05$  threshold is -1.83. Since the t value is not less than the threshold, that the CPFIS is better than SVR in the absolute errors is not statistically significant ( $P=0.068$ ).

The proposed CP is similar to support vectors in SVR. Both are among the training data, chosen in the process of training. The CP is the center of the local region around it, but a support vector is on the boundary of the specified  $\varepsilon$  error. Due to the Gaussian membership functions in both the input and output fuzzy sets and the small number of fuzzy rules, the rule base of the CPFIS can be easily understood by users as shown in Fig. 9. This interpretability is not aimed in a SVR. However, the calculations of SVR are simpler than those of CPFIS, and thus, faster in training.

Artificial neural networks were also tried for comparison. We used the software MATLAB Version 6.1 here. We employed the command `newff` to build a feed-forward backpropagation network. There was one hidden layer with 10 nodes. The input and hidden layers used the hyperbolic tangent sigmoid transfer function (parameter: TANSIG), while the output layer used the linear transfer function (parameter: PURELIN). The model was the Levenberg-Marquardt backpropagation (parameter: TRAINLM), and the learning is the gradient descent with momentum weight/bias (parameter: LEARN\_GDM). The result with the 100-epoch training setting is listed in Table 5. The 1000-epoch setting was also tried, but the result is worse than the 100-epoch case due to over fitting of training data.

As before, we can compare the performances of neural networks, CPFIS, and SVR in the averages of absolute errors in the RMM, RCM, ATT, MEN, ORI, ABS, LAN, DRA, ANM, and total scores from Tables 3, 4, and 5. Using the Student's t hypothesis testing, we have  $P=0.0029$  that CPFIS is less than neural networks, and  $P=0.0042$  that SVR is less than neural networks. Since both the P values are statistically significant, both CPFIS and SVR performed better than neural networks in the experiments. Much more effort can be made using different settings of training epochs, network models, number of hidden nodes, and so on to improve the performance of neural networks and outperform CPFIS and SVR. However, the interpretation of the connected network is not as convenient as the linguistic fuzzy rules in CPFIS. It is usually viewed as a black box. This is the most distinct feature that distinguishes between CPFIS and neural networks.

## 5. Conclusion

In this paper, a characteristic-point-based fuzzy inference system (CPFIS) was proposed to predict the

CASI scores from SPECT volumes. The normalized averages of 37 areas of Talairach labels were selected. Considering the left, the right, and the whole brains of these areas, there are  $37 \times 3 = 111$  average values. After calculating the linear correlation coefficients between these average values and the 9 CASI item scores of training volumes, we had 4, 2, 6, 2, 3, 2, 2, 2, and 4 variables for RMM, RCM, ATT, MEN, ORI, ABS, LAN, DRA, and ANM of the 9 CASI items, respectively. We built a CPFIS for each of these 9 CASI items. Experiment results showed that CPFIS could provide a rough estimation of CASI scores that the P value for the hypothesis that the average of absolute errors of CPFIS is less than the average of absolute differences from the mean values was 0.0447. For comparison, a weighted SVR machine was also tried. The performance of CPFIS was slightly better than SVR in this study ( $P=0.068$ ). Due to the Gaussian membership functions in both the input and output fuzzy sets and the small number of fuzzy rules, the rule base of the CPFIS can be easily understood by radiologists.

## 6. Appendix

The details of the learning algorithms in Section 3 have been reported in [15]. For completeness and convenience, we list these algorithms here.

### A. Gradient-Projection Method

Let  $\sigma_{threshold}$  and  $d_{threshold}$  be some chosen thresholds for the spreads of the membership functions of output fuzzy sets,  $\sigma_j$ , and the elements of the projection vector  $d$  of the gradient, respectively. Let  $n$  be the number of training data, and the vector  $\sigma$  be  $[\sigma_1, \dots, \sigma_n]^T$ . The function  $f$  in the following is the (1).

Step 1) Let  $I = \{i \mid 1 \leq i \leq n, \sigma_i < \sigma_{threshold}\}$ . A matrix  $A_q$  is set at

$$\begin{bmatrix} 1 & 1 & \dots & 1 & 1 \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & -1 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix},$$

The first row is all ones. Then, the remaining rows are formed by the active constraints. Each element  $j$  in  $I$  corresponds to a row in  $A_q$ :  $[0 \dots 0 -1 0 \dots 0]$ , -1 being at the  $j$ th column. Thus,  $A_q$  is a matrix of  $|I| + 1$  rows and  $n$  columns.

Step 2) Calculate  $P = I - A_q^T (A_q A_q^T)^{-1} A_q$ , and

$$d = -P\nabla f(\sigma)^T.$$

Step 3) If  $\max_{1 \leq i \leq n} |d_i| \geq d_{threshold}$ , find  $\alpha_1$  and  $\alpha_2$  achieving, respectively,

$$\alpha_1 = \max\{\alpha : \sigma + \alpha d \text{ is feasible}\}.$$

$$\alpha_2 = \arg \min\{f(\sigma + \alpha d) : 0 \leq \sigma \leq \alpha_1\}.$$

Set  $\sigma$  to  $\sigma + \alpha_2 d$  and return to Step 1.

Step 4) If  $\max_{1 \leq i \leq n} |d_i| < d_{threshold}$ , find

$$\lambda = -(A_q A_q^T)^{-1} A_q \nabla f(\sigma)^T.$$

Case (a) If  $\lambda_j \geq 0$ , for all  $j$  corresponding to active constraints, stop;  $\sigma$  satisfies the Kuhn-Tucker conditions,

Case (b) Otherwise, delete the row from  $A_q$  corresponding to the constraint with the most negative component of  $\lambda$  and return to Step 2.

### B. Gauss-Jordan-Elimination-Based Column Elimination

From (3), we have  $\mathbf{Cm}=\mathbf{D}$ , where  $\mathbf{C}$  is  $s \times s$ ,  $\mathbf{m}$  is  $s \times 1$ ,  $\mathbf{D}$  is  $s \times 1$ , and  $s$  is  $n-|I|$ . The set  $I$  here is the final result from the above gradient-projection method. Choose a threshold  $C_{threshold} > 0$ .

Step 1) Form the matrix  $[\mathbf{C}|\mathbf{D}] = \mathbf{F}$ . Let the index  $i$  be 1. Some columns of  $\mathbf{C}$  will be deleted, and thus, the column dimension of  $\mathbf{C}$  is gradually decreased in the iteration of the algorithm.

Step 2) If  $\sum_{1 \leq j \leq s} |C_{ji}| < C_{threshold}$ , then the  $i$ th column of

$\mathbf{C}$  and the  $i$ th column of  $\mathbf{m}$  are both deleted, and go to Step 5; otherwise, find  $p = \arg_j \max_{1 \leq j \leq s} |C_{ji}|$ .

Step 3) Exchange the  $p$ th and  $i$ th rows of  $\mathbf{F}$ . Divide the elements of the  $i$ th row by  $F_{ii}$ .

Step 4) All the other rows,  $j=1, \dots, s, j \neq i$ , are updated by adding  $-(F_{ji}/F_{ii}) \times F_{ik}$ ,  $k = i, \dots, (\text{the column number of } \mathbf{C}) + 1$ .

Step 5) If  $i$  is equal to the column number of  $\mathbf{C}$ , then stop; otherwise,  $i$  is set to  $i+1$ , and go to Step 2.

### C. Back-Propagation Tuning

From (4), let  $E$  be

$$\sum_{i=1}^n [y_i - \frac{\sum_{j=1}^z y_{r(j)} \sigma_{r(j)} A(\mathbf{x}_i, \mathbf{x}_{r(j)})}{\sum_{p=1}^z \sigma_{r(p)} A(\mathbf{x}_i, \mathbf{x}_{r(p)})}]^2 \text{ where}$$

$$A(\mathbf{x}_i, \mathbf{x}_{r(j)}) = \exp\left(-\frac{(x_i^{(1)} - m_{r(j)}^{(1)})^2}{2(\sigma_{in,r(j)}^{(1)})^2} - \dots - \frac{(x_i^{(m)} - m_{r(j)}^{(m)})^2}{2(\sigma_{in,r(j)}^{(m)})^2}\right)$$

and  $A(\mathbf{x}_i, \mathbf{x}_{r(p)})$  is similar. Then we can calculate the

derivatives,  $\frac{\partial E}{\partial \sigma_{in,r(j)}^{(k)}}$ ,  $\frac{\partial E}{\partial m_{r(k)}}$ , and  $\frac{\partial E}{\partial \sigma_{r(k)}}$ . The

algorithm is an iteration of the updating of parameters,

$\sigma_{in,r(j)}^{(k)}$ ,  $m_{r(j)}$ , and  $\sigma_{r(j)}$ :

$$\Delta \sigma_{in,r(j)}^{(k)}(t) = -\eta \frac{\partial E}{\partial \sigma_{in,r(j)}^{(k)}} \Bigg|_t + \alpha \Delta \sigma_{in,r(j)}^{(k)}(t-1)$$

$$\sigma_{in,r(j)}^{(k)}(t) = \sigma_{in,r(j)}^{(k)}(t-1) + \Delta \sigma_{in,r(j)}^{(k)}(t)$$

$$\Delta m_{r(k)}(t) = -\eta \frac{\partial E}{\partial m_{r(k)}} \Bigg|_t + \alpha \Delta m_{r(k)}(t-1)$$

$$m_{r(k)}(t) = m_{r(k)}(t-1) + \Delta m_{r(k)}(t)$$

$$\Delta \sigma_{r(k)}(t) = -\eta \frac{\partial E}{\partial \sigma_{r(k)}} \Bigg|_t + \alpha \Delta \sigma_{r(k)}(t-1)$$

$$\sigma_{r(k)}(t) = \sigma_{r(k)}(t-1) + \Delta \sigma_{r(k)}(t)$$

where  $0 < \eta < (1/n)$  and  $0 < \alpha < 1$  are the learning rate and the momentum constant, respectively.

## 7. Acknowledgment

This paper is supported by the National Science Council, R.O.C. (NSC Project: 92-2218-E-041-002 2003/8/1 ~ 2004/7/31).

## 8. References

- [1] S. Kobashi, T. Takae, Y. Hata, Y. T. Kitamura, T. Yanagida, O. Ishikawa, and M. Ishikawa, "Automated segmentation of the cerebrospinal fluid and the lateral ventricles from human brain MR images," In *Joint 9th IFSA World Congress and 20th NAFIPS Int. Conf.*, vol. 4, pp. 1961–1966, Vancouver, BC, Canada, July 2001.
- [2] K. Sato, K. Sugawara, Y. Narita, and I. Namura, "Consideration of the method of image diagnosis with respect to frontal lobe atrophy," *IEEE Trans. Nuclear Science*, vol. 43, no. 6, pp. 3230–3239, Dec. 1996.
- [3] D. Wang, S. Rose, G. Cowin, D. Spooner, D. Barnes, G. Galloway, D. M. Doddrell, J. B. Chalk, and J. Semple, "Regional rates of brain

atrophy--can they be used as a reliable tool for early diagnosis of Alzheimer's disease," In *Joint 9th IFSA World Congress and 20th NAFIPS Int. Conf.*, vol. 4, pp. 1985–1990, Vancouver, BC, Canada, July 2001.

- [4] R. D. Terry, R. Katzman, and K. L. Bick, Eds., *Alzheimer Disease*, Raven Press, New York, 1994.
- [5] C. K. Liu, R. T. Lin, C. L. Lai, and C. T. Tai, "A normative study of Chinese version of the cognitive ability screening instrument [abstract]," *Acta Neurol Taiwan*, vol. 7, p. 142, 1998.
- [6] Y. Ushijima, C. Okuyama, S. Mori, T. Nakamura, T. Kubota, and T. Nishimura, "Relationship between cognitive function and regional cerebral blood flow in Alzheimer's disease," *Nucl. Med. Comm.*, vol. 23, pp. 779–784, 2002.
- [7] J. L. Eberling, B. R. Reed, M. G. Baker, and W. J. Jagust, "Cognitive correlates of regional cerebral blood flow in Alzheimer's disease," *Arch. Neurol.*, vol. 50, pp. 761–766, 1993.
- [8] N. Wolfe, B. R. Reed, J. L. Eberling, and W. J. Jagust, "Temporal lobe perfusion on single photon emission computed tomography predicts the rate of cognitive decline in Alzheimer's disease," *Arch. Neurol.*, vol. 52, pp. 257–262, 1995.
- [9] D. G. Luenberger, *Linear and Nonlinear Programming*, Addison-Wesley, second edition, 1989.
- [10] S. Nakamura, *Applied Numerical Methods in C*, P T R Prentice Hall, New Jersey, 1993.
- [11] S. Haykin, *Neural Networks: A Comprehensive Foundation*, IEEE Press, 1994.
- [12] K. J. Friston, J. Ashburner, J. B. Poline, C. D. Frith, and R. R. J. Frackowiak, "Spatial registration and normalization of images," *Hum. Brain Mapp.*, vol. 2, pp. 189–210, 1995.
- [13] D. Soonawala, T. Amin, K. P. Ebmeier, J. D. Steele, N. J. Dougall, J. Best, O. Migneco, F. Nobili, and K. Scheidhauer, "Statistical parametric mapping of 99m-Tc-HMPAO-SPECT images for the diagnosis of Alzheimer's disease: normalizing to cerebellar tracer uptake," *NeuroImage*, vol. 17, pp. 1193–1202, 2002.
- [14] P. W. Strike, *Statistical Methods in Laboratory Medicine*, Butterworth Heinemann, 1991.
- [15] T. K. Yin, "A characteristic-point-based fuzzy inference system aimed to minimize the number of fuzzy rules," *IEEE Trans. Fuzzy Syst.*, vol. 12, no. 2, pp. 250–273, April 2004.
- [16] J. A. K. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle, "Weighted least squares support vector machines: robustness and sparse approximation," *Neurocomputing*, vol. 48, pp. 85–105, 2002.



Tang-Kai Yin received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1990, and the M.S. degree in electrical engineering and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, in 1994 and 1996, respectively.

He worked in industry as a Senior Software Engineer on device drivers for network interface cards and scanners from 1996 to 1999. From 1999 to 2004, he was an Assistant Professor in the Department of Management Information Science, Chia Nan University of Pharmacy and Science, Tainan, Taiwan, R.O.C. He is currently an Assistant Professor in the Department of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung, Taiwan, R.O.C. His research interests include fuzzy systems, neural networks, and their applications in computer-aided diagnosis for medical imaging and web-based learning.