

A Computer-Aided Diagnosis for Locating Abnormalities in Bone Scintigraphy by a Fuzzy System With a Three-Step Minimization Approach

Tang-Kai Yin*, *Member, IEEE*, and Nan-Tsing Chiu

Abstract—Bone scintigraphy is an effective method to diagnose bone diseases such as bone tumors. In the scintigraphic images, bone abnormalities are widely scattered on the whole body. Conventionally, radiologists visually check the whole-body images and find the distributed abnormalities based on their expertise. This manual process is time-consuming and it is not unusual to miss some abnormalities. In this paper, a computer-aided diagnosis (CAD) system is proposed to assist radiologists in the diagnosis of bone scintigraphy. The system will provide warning marks and abnormal scores on some locations of the images to direct radiologists' attention toward these locations. A fuzzy system called characteristic-point-based fuzzy inference system (CPFIS) is employed to implement the diagnosis system and three minimizations are used to systematically train the CPFIS. Asymmetry and brightness are chosen as the two inputs to the CPFIS according to radiologists' knowledge. The resulting CAD system is of a small-sized rule base such that the resulting fuzzy rules can be not only easily understood by radiologists, but also matched to and compared with their expert knowledge. The prototype CAD system was tested on 82 abnormal images and 27 normal images. We employed free-response receiver operating characteristics method with the mean number of false positives (FPs) and the sensitivity as performance indexes to evaluate the proposed system. The sensitivity is 91.5% (227 of 248) and the mean number of FPs is 37.3 per image. The high sensitivity and moderate numbers of FP marks per image shows that the proposed method can provide an effective second-reader information to radiologists in the diagnosis of bone scintigraphy.

Index Terms—Biomedical imaging, bones, fuzzy systems, medical diagnosis, minimization methods.

I. INTRODUCTION

BONE scintigraphy is a useful tool in diagnosing bone diseases such as bone tumors [1], metabolic bone disease [2], Reiter's syndrome [3], Paget's disease [4], and Gardner's syndrome [5]. It is highly sensitive. By manually checking the whole-body scans of patients, radiologists can find the abnormal

Manuscript received May 15, 2002; revised January 30, 2004. This work was supported in part by the National Science Council under Grant NSC-89-2213-E-041-007. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was C. Mayer. *Asterisk indicates corresponding author.*

*T. K. Yin is with the Department of Management Information Science, Chia-Nan University of Pharmacy and Science, Tainan, Taiwan, R.O.C. (e-mail: qtkyin@mail.chna.edu.tw).

N. T. Chiu is with the Department of Nuclear Medicine, College of Medicine, National Cheng Kung University, Tainan, Taiwan, R.O.C.

Digital Object Identifier 10.1109/TMI.2004.826355

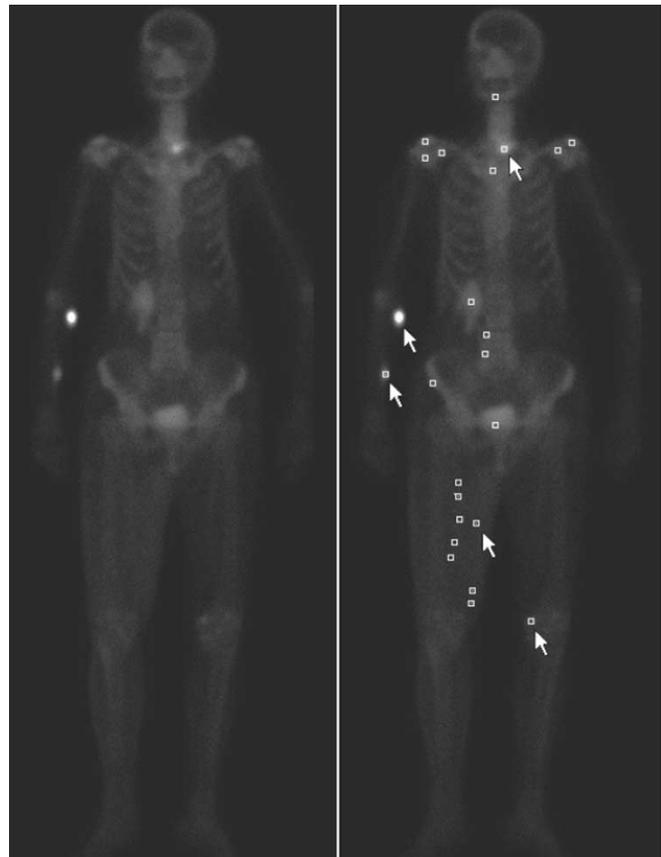


Fig. 1. One sample image of bone scintigraphy. The left is the original image, and the right is its CAD by adding possible abnormal marks to help radiologists. Five of these marks at the arrow-pointed locations are true abnormalities. The other 19 marks are FPs. There are no missing abnormalities for this sample image.

locations and their distributions in bone. The diagnosis is based on the expertise and also subjective assessment of radiologists by visually inspecting the bone scintigraphic images. In this paper, we propose a computer-aided diagnosis (CAD) system to aid radiologists by providing an automatic classification and objective scores of possible abnormalities on the whole-body scans.

Fig. 1 is a sample image of bone scintigraphy. The left-hand side of Fig. 1 is the original image, and the right-hand side is the result after the diagnosis of the proposed system. When

a new scintigraphic image is opened, the system provides abnormal marking as squares and scores of all these squares on the status bar when the mouse is clicked on the squares. The CAD system acts as a second reader to provide warning marks. For example, 24 squares are labeled in the right-hand side of Fig. 1 as warning marks. All of them deserve careful screening. Five arrow-pointed squares are true abnormalities. Especially, the arrow-pointed mark at the right thigh may be overlooked since its pixel value is lower than many other pixels in the image. The trained CAD system can mark such abnormalities timely and effectively.

One major application of CAD on medical images is the diagnosis of mammograms. Nowadays, only two commercially CAD systems for mammograms are available [6]. They are the "ImageChecker" system (R2, Los Altos, CA.) [7]–[9], and the "Second Look" CAD system (CADx Medical Systems, Quebec, Canada) [6]. The sensitivities and false positive (FP) marks per image of these two systems are suitable comparison baselines for new CAD systems for mammograms or other images. In [10], Verma and Zakos used fuzzy-neural and feature extraction techniques to diagnose microcalcifications' patterns. Cheng *et al.* used a fuzzy logic technique to detect microcalcifications even in very dense breast mammograms [11]. Yu and Guan employed neural networks and wavelet transforms to detect clusters of microcalcifications [12]. Li *et al.* utilized morphological enhancement, contextual segmentation and neural networks to mammographic mass detection [13], [14]. Some other CAD applications are the detections for lung nodules by neural networks [15], the frontal lobe atrophy by fuzzy k-means [16], glioblastoma-multiforme tumors by a knowledge-based technique [17], ocular fundus vessels by fuzzy c-means [18], and malignant melanoma by several segmentation algorithms and feature selections [19]. In this paper, we tried to apply CAD to a new area of bone scintigraphy as shown in Fig. 1. Bone scintigraphy has some similar properties as in the above researches. Like mammograms, abnormalities are few locations widely distributed in the whole scan. They can be detected by brightness and shape, but for bone scintigraphy, shape is not an effective parameter. We will employ a new parameter, asymmetry, in our CAD system. This is based on the symmetry structure of human skeleton. Asymmetry is one of the major parameters employed by radiologists in checking bone scintigraphic images.

In this paper, in order to take into account of experts' knowledge, we use a fuzzy-logic technique in our CAD system. In addition to some of the aforementioned papers, fuzzy logic was also applied in medical images such as segmentation of intrathoracic airway trees [20], model-free functional MRI analysis [21], segmentation of dynamic neuroreceptor single-photon emission tomography images [22], enhancement of computerized tomography images [23], classification analysis of exercise-induced lower leg pain [24], and feature extraction for brain MRI segmentation [25]. Many of the researches employed fuzzy c-means clustering, but to have linguistic interpretation of the CAD system, we will take the original form of fuzzy inference rules as in Zadeh's original introduction of fuzzy logic [26]–[28]. We will follow our previous research in [29], [30] in which a characteristic-point-based fuzzy inference system (CPFIS) was proposed. A new training

method by a three-step minimization approach is employed in this paper to train the CPFIS in a systematic way. The minimization approach consists of a gradient-projection method, a Gauss–Jordan-elimination-based column elimination, and a back-propagation tuning. The number of fuzzy rules are gradually decreased from the original number of training samples to a small number in the process of minimization. The resulting CPFIS has a small-sized rule base. The advantage is that the linguistic interpretation and graphical presentation of few fuzzy rules can enable the underlying CAD diagnosing mechanism to be easily understood by radiologists [31], [32]. In our experiment, three or four rules were obtained at the end of training. Interpreted by the input variables, asymmetry and brightness, and the output variable, score, the proposed CAD system can be matched to and compared with human knowledge about the diagnosis of bone scintigraphy.

This paper is organized as follows. The image preprocessing of bone scintigraphy is discussed in Section II. A local-maximum-based segmentation is first employed to divide the image into a number of not overlapped connected areas. Then the vertically central line of body is calculated. The two variables, asymmetry and brightness, are defined and calculated for each connected area. Section III presents the employed fuzzy system, CPFIS, and the accompanying three-step minimization training. The algorithms of the three minimizations are listed in the Appendix. Section IV delineates the whole process of training. The training samples are keyed in by radiologists and then some of them are removed by the abnormal-point cancellations. Six CPFISs are employed for the head, vertebrae, chest, pelvic, hand, and leg regions of body. Section V discusses our experiment. The prototype CAD system was implemented and trained by 20 sample images. After training, another 82 abnormal images and 27 normal images were employed to test the performance of the proposed system. The results in terms of sensitivity and FP marks per image were satisfactory. Finally, Section VI concludes this paper.

II. BONE-SCINTIGRAPHY PREPROCESSING

The original Tc-99m bone scintigraphy consists of two 256-pixel-wide and 1024-pixel-high images for whole-body front and back scans. Each pixel is with a 16-bit value in gray scale. The first step to deal with these images is to choose some variables that extract information from the images. The choices must be the most utilized parameters by radiologists in diagnosing bone scintigraphy, in order to include their expert knowledge in fuzzy systems. After discussions with radiologists, we propose to employ asymmetry and brightness as the input variables of fuzzy systems. Since the bone structure of human being is vertically symmetric, the asymmetric property of local areas should be checked. Besides, abnormal bone will have high scintigraphic values than normal one; thus, brightness is also important in diagnosis.

A. A Local-Maximum-Based Segmentation

The bone image is checked by areas, but not by pixels. A single pixel may have a high scintigraphy value, but it is not taken into account because noise exists in the original image.

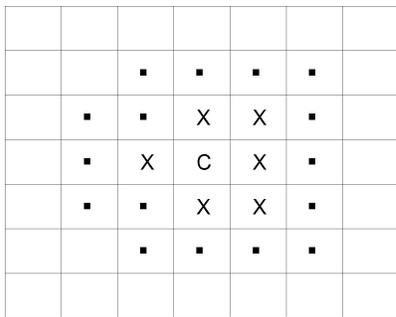


Fig. 2. Growing of a local-maximum area in the segmentation algorithm. “C”: the center and the maximum pixel, “X”: the pixels in this local-maximum area, and “.”: the neighborhood.

We use a local-maximum-around segmentation to generate not overlapped connected areas.

Denote the scintigraphic value at (i, j) is $s(i, j)$. Let T be a threshold value. Denote

$$A = \{(i, j) | s(i, j) \geq T, s(i, j) \geq s(a, b), \\ \forall a = i - 1, i, i + 1, b = j - 1, j, j + 1\}.$$

Thus, A is a set of local maxima. Then sort A and denote the result as a list, L , of elements in an order of descending values

$$L = \{t_i | t_i \cdot s \geq t_j \cdot s, \forall i < j\}$$

where $t_i \cdot s$ stands for the gray value of the pixel t_i . Therefore, we have a list of pixels, t_0, t_1, \dots, t_{m-1} , in an order of decreasing values. The last element t_{m-1} is greater than or equal to the threshold T . In order to obtain the pixels around each local maximum, we develop the following algorithm. The detailed discussion follows the algorithm. Let t_c be the center of a local area and r be a chosen ratio that any pixel belonging to the area around t_c should be of a value greater than or equal to $t_c \cdot s \times r$, $0 < r < 1$. In our experiment, r is 2/3 and T is 30. As shown in Fig. 2, the segmentation operation starts with each local maximum, checks all the neighboring pixels around this maximum pixel with values greater than or equal to $t_c \cdot s \times r$. Denote all these local areas as the set $S_{\text{local}}(t)$, $\forall t \in A$.

Step 1) Let $I(i) = 0, \forall i = 0, 1, \dots, N-1$. N is the number of whole-body pixels.

Step 2) Choose t_i , so that for all the pixels, t_0, t_1, \dots, t_{i-1} , their indicator-function values $I(j)$, are equal to one, where j is the index of these pixels in the image.

If no such t_i exists, then stop.

Let $t_c = t_i$ and $Q_{\text{check}} = S_{\text{local}}(t_c) = \{t_c\}$.

Step 3) If $Q_{\text{check}} = \Phi$, then go to Step 2).

Remove the first element p of Q_{check} and add p into $S_{\text{local}}(t_c)$.

Step 4) The location of p is (i, j) .

Denote all the eight elements at $(i - 1, j - 1), \dots, (i + 1, j + 1)$, around (i, j) , as p_k .

If any element p_k of these eight elements satisfies the conditions that $p_k \cdot s \geq t_c \cdot s \times r$ and its indicator-function value $I(j)$ being zero, j being the index of p_k in the image, then add p_k into the set Q_{check} as the last element of the set and set $I(j) = 1$.

Step 5) Go to Step 3).

In Step 1), we define an indicator function $I(i)$ that

$$I(i) = \begin{cases} 0, & \text{if the pixel } i \text{ is not yet checked in the algorithm} \\ 1, & \text{if the pixel } i \text{ has been checked in the algorithm.} \end{cases}$$

Initially, all the N pixels of the whole-body image are not checked. In Step 2), we take one pixel t_i out of L for checking. We sequentially look at t_0, t_1, \dots, t_{m-1} and choose the first t_i that it is not yet checked. Then, we will grow an area around the chosen t_i . The variable t_c is the center of this area. The set $S_{\text{local}}(t_c)$ is the set of the pixels in this area. The queue Q_{check} is a queue of pixels for checking. The following steps 3)–5) add the neighboring pixels around the center t_c into $S_{\text{local}}(t_c)$. In Step 3), if no more neighboring pixels are available for checking, then the $S_{\text{local}}(t_c)$ is done and go back to Step 2) for another center t_c ; otherwise, remove the first element of Q_{check} for checking and also add this element to $S_{\text{local}}(t_c)$. In Step 4), the growing of the area is performed by adding more neighboring pixels into Q_{check} . Check the eight pixels around a pixel $p \in S_{\text{local}}(t_c)$. If any pixel p_k of these eight elements satisfies the conditions that its indicator-function value $I(j)$ being zero, and $p_k \cdot s \geq t_c \cdot s \times r$, then add p_k into the queue Q_{check} as the last element of the queue and set $I(j)$ to be one. In Step 5), the process goes back to Step 3) for checking another pixel in the queue Q_{check} .

B. Asymmetry and Brightness Calculations

To calculate asymmetry values, we have to find the vertically central line of body. Let $c(j)$ be the function such that $(c(j), j)$ is on the central line. We use two steps to obtain $c(j)$. The first step is a centroid method, and the second step is a minimum-of-difference method to fine-tune the result $c_1(j)$ of the first step. To reduce the noise effect, we use the sum of neighboring pixels

$$d(i, j) = \sum_{i-2 \leq i_1 \leq i+2} \sum_{j-2 \leq j_1 \leq j+2} s(i_1, j_1)$$

instead of $s(i, j)$ in the follows.

Step 1) Centroid method

$$c_1(j) = \frac{\sum_{0 \leq i \leq 255} d(i, j) \times i}{\sum_{0 \leq i \leq 255} d(i, j)}, j = 0, \dots, 1023. \quad (1)$$

Step 2) Fine-tuning

$$c(j) = \arg \min_{k} \sum_{n \leq k \leq c_1(j) + n} \sum_{1 \leq a \leq m} \sum_{b=j-f, j, j+f} |d(k+a, b) - d(k-a, b)| \quad (2)$$

where n, m, f are suitably chosen constants. In our experiment, f is 15 and this distance can reduce most abnormal-pixel effect that can add bias to the calculations of the central line. For the head region, n is 30 and m is 70; for the other regions, n is 10 and m is 120.

Denote $asym(i, j)$ and $brg(i, j)$ as the asymmetry and brightness values of the local area $S_{\text{local}}(t_c) = S_{\text{local}}(t_c \text{ at } (i, j)) = S_{\text{local}}(i, j)$, respectively. If the pixel

at (i, j) is not the center, t_c , of a local area, both $asym(i, j)$ and $brg(i, j)$ are set to be 0. Otherwise, they are calculated as follows.

Let a and b be the arguments in the minimum of the following differences:

$$\arg \min_{k, l} \min_{-5 \leq k \leq 5, -5 \leq l \leq 5} \|d(2c(j) - i + k, j + l) - d(i, j)\|. \quad (3)$$

Then

$$asym(i, j) = 100 \cdot \left\| \frac{d(2c(j) - i + a, j + b) - d(i, j)}{d(2c(j) - i + a, j + b) + d(i, j)} \right\|. \quad (4)$$

From the above calculation, $asym(i, j)$ is the difference between the neighboring pixels and the symmetrically corresponding pixels across the central line. The minimization in (3) is to roughly eliminate the error in finding the corresponding left and right bone elements of human bone by (2). It is to match the pixels in a small square range to take into account of small body movement. For example, the left coxal bone and the right coxal bone at pelvic in Fig. 1 are not at the same horizontal line. The left coxal bone is slightly lower than the right one. To precisely find the coxal bones, it is necessary to do translations and rotations of individual pixels in the bone-scintigraphy images, because the positions of the patients are not exactly symmetric in the images. To simplify the image processing, the rough method of calculating $asym(i, j)$ is employed.

Let m_{body} be the average of the whole body. The brightness $brg(i, j)$ is calculated as

$$brg(i, j) = \frac{20}{m_{\text{body}}} \frac{\sum_{\substack{-4 \leq k \leq 4, -4 \leq l \leq 4, \\ \text{excluding } k=0 \text{ and } l=0}} (s(i, j) - s(i + k, j + l))}{80} \quad (5)$$

where for convenience we denote $(i, j) \in S_{\text{local}}$ in which (i, j) is the location of one element t . From the equation, $brg(i, j)$ is the average of the differences between the center pixel (i, j) and its 80 neighboring pixels with the scaling factor $20/m_{\text{body}}$. Most abnormalities are brighter spots over normal bone or tissue. Therefore, instead of the center pixel value itself, the differences between the center pixel and its neighboring pixels are used here.

To reduce unnecessary spot samples and, thus, decrease computation time in the later learning, the following three conditions are used to remove many normal local maxima. The values of the thresholds and the parameters in these conditions are decided by the later experiments.

Condition 1: If $brg(i, j) < T_{brg}$, then $asym(i, j) = brg(i, j) = 0$.

In bone scintigraphy, abnormalities have high pixel values. Therefore, a threshold of $brg(i, j)$ can be suitably chosen to remove low-pixel-value normal samples.

Condition 2: If $\|S_{\text{local}}\| > T_{\text{size}}$, then $asym(i, j) = brg(i, j) = 0$.

Both abnormalities and individual bone elements can have high pixel values, but the sizes of the former are usually smaller than those of the latter. Thus, if the size of some local maximum is greater than a threshold T_{size} , the local region is on a normal bone element.

Condition 3: If $brg(i, j) < L_0 + L_1(asym(i, j) - L_2)$, then $asym(i, j) = brg(i, j) = 0$.

It is observed that if the asymmetry values of true abnormalities are larger than a positive number L_2 , their brightness values will be greater than $asym(i, j)$ -based values, $L_0 + L_1(asym(i, j) - L_2)$. All these three conditions and the parameters and thresholds in them are only based on the observations on samples. When they are applied to other images, abnormalities may be mistakenly included to become false negatives (FNs). In our experiments, no abnormalities in either the training images or the testing images were taken to be normal by these three conditions. Therefore, we adopt these conditions to remove many normal local maxima in this paper.

III. A FUZZY SYSTEM WITH A THREE-STEP MINIMIZATION APPROACH

A. Characteristic-Point-Based Fuzzy Inference System

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

$$\begin{aligned} \text{If } x^{(1)} \text{ is } A_1^{(1)} \text{ and } \dots \text{ and } x^{(m)} \text{ is } A_1^{(m)}, \text{ then } y \text{ is } B_1, \\ \vdots \\ \text{If } x^{(1)} \text{ is } A_z^{(1)} \text{ and } \dots \text{ and } x^{(m)} \text{ is } A_z^{(m)}, \text{ then } y \text{ is } B_z, \end{aligned}$$

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

$$\mu_{A_j^{(k)}}(x^{(k)}) = \exp\left(-\frac{(x^{(k)} - m_{in,j}^{(k)})^2}{2(\sigma_{in,j}^{(k)})^2}\right). \quad (6)$$

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 σ_j :

$$\mu_{B_j}(y) = \exp\left(-\frac{(y - m_j)^2}{2\sigma_j^2}\right). \quad (7)$$

There are three steps in making an inference of a CPFIS:

Step 1) Calculate the firing strength w_j , $1 \leq j \leq z$, for each fuzzy rule

$$w_j(\mathbf{x}) = \mu_{A_j^{(1)}}(x^{(1)}) \times \dots \times \mu_{A_j^{(m)}}(x^{(m)}). \quad (8)$$

Step 2) Form the output fuzzy sets

$$w_j \mu_{B_j}(y), \quad 1 \leq j \leq z.$$

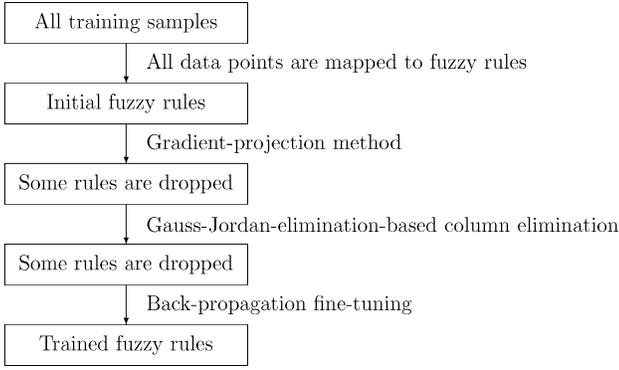


Fig. 3. A systematic approach by three minimization steps is proposed to train the CPFIS.

Step 3) Defuzzify the output fuzzy sets by using the simulated center-of-area method of Lin and Lee [33]

$$y = \frac{\sum_{j=1}^z w_j m_j \sigma_j}{\sum_{j=1}^z w_j \sigma_j}. \quad (9)$$

The training of CPFIS is to decide the number of fuzzy rules, and the parameters of these rules. A systematic approach by three minimization steps is proposed as shown in Fig. 3. The first minimization is based on σ_j , the spreads of the membership functions of the output fuzzy sets. The number of fuzzy rules will be reduced from n , the number of training data, to s , $s \leq n$. Then the second minimization is based on m_j , the means of the membership functions of the output fuzzy sets. After this step, the number of fuzzy rules is further reduced to be z , $z \leq s$. The number z is a small number by adjusting a threshold value of column sums. Finally, the third minimization is based on σ_j , m_j , and $\sigma_{in,j}$, the spreads of the membership functions of the input fuzzy sets. This step is a fine-tuning of parameters in fuzzy sets but not a determination of the number of fuzzy sets. The aim of this fine-tuning is to enhance the precision performance of CPFIS. The follows are the details of the proposed systematic training process of CPFIS.

B. 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) , $\mathbf{x}_i \in \mathbb{R}^m$, $y_i \in \mathbb{R}$, 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,

$$\begin{aligned} m_{in,i}^{(t)} &= x_i^{(t)}, \quad t = 1, \dots, m, \quad i = 1, \dots, n, \\ m_i &= y_i, \quad i = 1, \dots, n. \end{aligned}$$

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, σ_j , remains to be set. We set

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

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

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

and the fuzzy inference output is

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

The weights σ_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:

$$\begin{aligned} \min_{\sigma_1, \dots, \sigma_n} \sum_{i=1}^n \left[y_i - \frac{\sum_{j=1}^n y_j \sigma_j A(\mathbf{x}_i, \mathbf{x}_j)}{\sum_{p=1}^n \sigma_p A(\mathbf{x}_i, \mathbf{x}_p)} \right]^2 \\ \text{subject to } \sum_{j=1}^n \sigma_j = 1, \sigma_j \geq 0, j = 1, \dots, n. \quad (10) \end{aligned}$$

A gradient projection method is employed to solve this problem. The details are included in Appendix I. It is noted that (10) 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 σ_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

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

where $r(\cdot)$ is a function that indicates $r(j)$ 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 $r(\cdot)$.

C. Gauss–Jordan-Elimination-Based Column Elimination

The weights obtained from the gradient-projection method for all the s fuzzy rules are further utilized in the Gauss–Jordan-elimination-based column elimination. The minimization of this step is based on m_j

$$\min_{m_{r(1)}, \dots, m_{r(s)}} \sum_{i=1}^n \left[y_i - \frac{\sum_{j=1}^s m_{r(j)} \sigma_{r(j)} A(\mathbf{x}_i, \mathbf{x}_{r(j)})}{\sum_{p=1}^s \sigma_{r(p)} A(\mathbf{x}_i, \mathbf{x}_{r(p)})} \right]^2. \quad (11)$$

This is a quadratic function of the variables $m_{r(1)}, \dots, m_{r(s)}$. 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, (11) can be rewritten as

$$\min_{m_{r(1)}, \dots, m_{r(s)}} \sum_{i=1}^n \left[y_i - \sum_{j=1}^s m_{r(j)} B(i, j) \right]^2 = \min_{m_{r(1)}, \dots, m_{r(s)}} f(m_{r(1)}, \dots, m_{r(s)}). \quad (12)$$

The necessary condition for the solution is that the first derivatives of f are zero. Take derivatives of (12), we have

$$\begin{aligned} & \frac{\partial f}{\partial m_{r(k)}} \\ &= (-2) \sum_{i=1}^n \left[y_i - \sum_{j=1}^s m_{r(j)} B(i, j) \right] B(i, k) = 0, k = 1, \dots, s \\ &\Rightarrow \sum_{i=1}^n \sum_{j=1}^s m_{r(j)} B(i, j) B(i, k) = \sum_{i=1}^n y_i B(i, k) \\ &\Rightarrow \begin{bmatrix} \sum_{i=1}^n B(i, 1) B(i, 1) & \sum_{i=1}^n B(i, 2) B(i, 1) & \vdots \\ \sum_{i=1}^n B(i, 1) B(i, 2) & \sum_{i=1}^n B(i, 2) B(i, 2) & \vdots \\ \vdots & \vdots & \vdots \\ \sum_{i=1}^n B(i, 1) B(i, s) & \sum_{i=1}^n B(i, 2) B(i, s) & \vdots \end{bmatrix} \begin{bmatrix} m_{r(1)} \\ m_{r(2)} \\ \vdots \\ m_{r(s)} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{i=1}^n y_i B(i, 1) \\ \sum_{i=1}^n y_i B(i, 2) \\ \vdots \\ \sum_{i=1}^n y_i B(i, s) \end{bmatrix}. \end{aligned} \quad (13)$$

Then, a modified Gauss–Jordan-elimination method is employed to solve the linear equations, (13). The details of the algorithm are included in Appendix II. During the process, a threshold is set to trim redundant columns in (13). The corre-

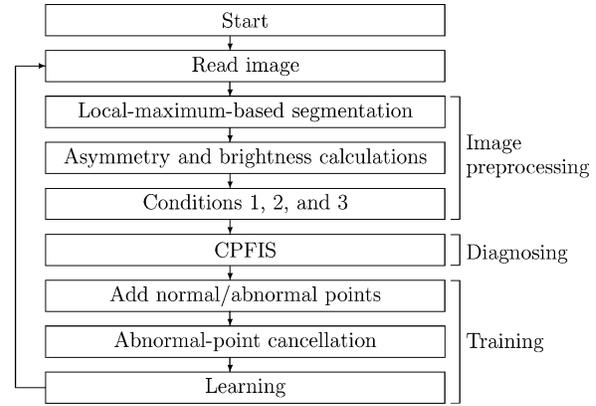


Fig. 4. Flow chart of training.

sponding $m_{r(j)}$ of these redundant columns are also removed. It is noted that the column-sum threshold here is aimed to choose the most representational columns, but not only to avoid the matrix inverse of a zero-determinant matrix in the original Gauss–Jordan elimination. Thus, the values of the threshold are increased gradually and the number of fuzzy rules are then decreased. After this step, the number of fuzzy rules have been decided. The numbering function $r(\cdot)$ is changed to indicate that the j th column, which is not removed in the algorithm, is the $r(j)$ -th fuzzy rule in the original n fuzzy rules.

D. Back-Propagation Tuning

After the second minimization, there remain $z, z \leq s$, fuzzy rules. The third minimization is based on $\sigma_{in,r(1)}, \dots, \sigma_{in,r(z)}, m_{r(1)}, \dots, m_{r(z)}, \sigma_{r(1)}, \dots, \sigma_{r(z)}$:

$$\min_{\sigma_{in,r(1)}, \dots, \sigma_{in,r(z)}, m_{r(1)}, \dots, m_{r(z)}, \sigma_{r(1)}, \dots, \sigma_{r(z)}} \sum_{i=1}^n \left[y_i - \frac{\sum_{j=1}^z m_{r(j)} \sigma_{r(j)} A(\mathbf{x}_i, \mathbf{x}_{r(j)})}{\sum_{p=1}^z \sigma_{r(p)} A(\mathbf{x}_i, \mathbf{x}_{r(p)})} \right]^2. \quad (14)$$

The details of the back-propagation process is described in Appendix III. It is noted that the means of the membership functions of the input fuzzy sets, $\mathbf{m}_{in,r(z)}$, are not variables in the minimization. We assign $m_{in,r(j)}^{(k)} = x_{r(j)}^{(k)}, k = 1, \dots, m, j = 1, \dots, z$. These $\mathbf{x}_{r(j)}$ positions are the obtained CPs from the Gauss–Jordan-elimination-based column elimination.

IV. TRAINING

The flow chart of training is shown in Fig. 4. The system reads a scintigraphic image, calculates the asymmetry and brightness values of the segmented areas of the image, makes inference to CPFIS, and then shows the result to radiologists. Radiologists inspect the image and add missing abnormal locations as training samples. Besides abnormal locations, radiologists can also add normal locations as training samples when some area of the image should be diagnosed as normal pixels.

Let sample (i, j) denote a sample has asymmetry value i , and brightness value j . These training samples are further processed by the following cancellation.

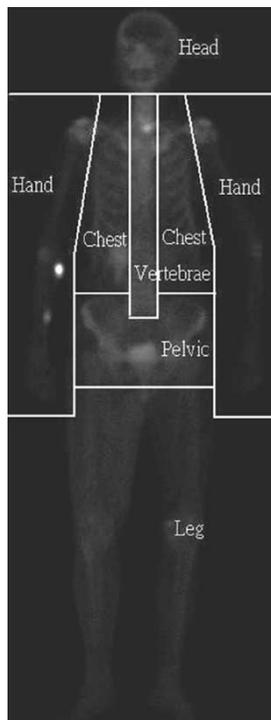


Fig. 5. Rough segmentation of six parts on the sample image in Fig. 1.

Abnormal-point cancellation: for an abnormal sample $(i1, j1)$ and a normal sample $(i2, j2)$, if $i1 \leq i2$ and $j1 \leq j2$, then delete $(i1, j1)$.

To the radiologists' knowledge, if both the asymmetry and brightness values of one point are smaller than a normal point, then this point should be also a normal point. Thus, if this point is an abnormal point, then we delete it. The removing of contradictory points can make the training of CPFIS more effective. After the cancellations of training samples, the three-step minimization learning is applied. The process is iterated until no new images exist or radiologists are satisfied with the current system performance.

To take into account varied sensitivity in different parts of body, six CPFISs are used for 6 regions of body: head, vertebrae, chest, spine, hand, and leg. Fig. 5 shows the segmentation of six parts on the sample image in Fig. 1. In this paper, the segmentation is a rough and automatic operation by assigning fixed ratios between parts. It is better that the segmentation can be adaptive since human body is of various sizes and shapes, but for simplicity, we take fixed segmentation operations. The results were acceptable as shown in the following experiment.

V. EXPERIMENT

To test the proposed system, we implemented a prototype system and performed experiments on a database in the department of nuclear medicine, National Cheng Kung University Hospital, Tainan, Taiwan, R.O.C. The scintigraphy-imaging device is a gamma camera (Siemens, E-cam, USA). The patients were injected with Technetium-99m-MDP (methylene diphosphonate). First, 20 whole-body scintigraphic images were chosen to train the proposed CPFIS. Some of them had minute abnormalities, or contained abnormalities in the head

part which was the most infrequent part having abnormalities. Altogether, they consisted of much information about scintigraphy abnormalities and, thus, became good training samples. After training, 82 abnormal images and 27 normal images were used to test the system. These testing cases were taken from the same database, excluding the 20 training cases and the cases in which abnormalities were located on back views. For simplicity, we considered only front-view diagnosis in the experiment. The results are detailed in the follows.

A. Analysis of the 20 Training Images

When a sample image was opened, the local-maximum-based segmentation was first performed. As shown in Table I, there were 13 952 normal spots and 104 abnormal spots in the 20 training sample images after the segmentation. Originally, radiologists labeled 110 abnormal spots on these 20 images. Thus, there are six (110–104) abnormal spots which are not local maxima. The minimum of $brg(i, j)$ of these 104 abnormal spots is 12. Therefore, T_{brg} was set to be 12 in Condition 1. After the process of Condition 1, normal spots had been greatly eliminated to be 8486, while none of the 104 abnormal spots was removed.

The maximum of $\|S_{local}\|$ of the 104 abnormal spots is 92. Thus, T_{size} was set to be 92 in Condition 2. After Condition 2, normal spots were further decreased to be 8214. To decide the parameters in Condition 3, the distribution of abnormal spots in $asym(i, j)$ and $brg(i, j)$ was plotted and the line was suitably chosen as shown in Fig. 6. The parameters L_0 , L_1 , and L_2 were set to be 0, 5/6, and 40, respectively. There is no definite rule in choosing the line. To be conservative, we chose the line at a distance to the right of all the abnormal points. After Condition 3, the number of normal spots was 8186. Although only 28 normal spots were removed by Condition 3, they were all high-asymmetry spots and very likely misdiagnosed.

After the image preprocessing, there were 8186 normal spots and 104 abnormal spots in the 20 training sample images. Only 5.5% (6 of 110) abnormal spots were eliminated, but 41.3% (5766 of 13 952) normal spots including a lot of normal pixels of large asymmetry or brightness values were also eliminated.

Table II shows the means and standard deviations of the asymmetry and brightness values in the 8186 normal spots and the 104 abnormal spots after the preprocessing of Condition 3. The mean of asymmetry values in abnormal spots is larger than that in normal spots, and the difference is 23.11 which is greater than the standard deviation of asymmetry values in normal spots and that in abnormal spots. Similarly, the mean of brightness values in abnormal spots is larger than that in normal spots, and the difference is 37.85 which is greater than the standard deviation of brightness values in normal spots and slightly smaller than that in abnormal spots. These significant differences show that the two variables $asym(i, j)$ in (4) and $brg(i, j)$ in (5) are effective variables in classifying abnormal and normal spots.

B. Training

The training process was as described in Section IV. There were six CPFISs for the six parts: head, vertebrae, chest, pelvic, hand, and leg. Initially, one default abnormal point of large asymmetry and brightness values, 80 and 120 in this experiment, and one default normal point of both zero asymmetry and

TABLE I
NUMBER OF NORMAL AND ABNORMAL SPOTS IN THE 20 TRAINING IMAGES AFTER THE SEGMENTATION AND CONDITIONS

Image Preprocessing	Normal Spots	Abnormal Spots
Original	-	110
After the local-maximum-based segmentation	13952	104
After Condition 1	8486	104
After Conditions 1 and 2	8214	104
After Conditions 1, 2, and 3	8186	104

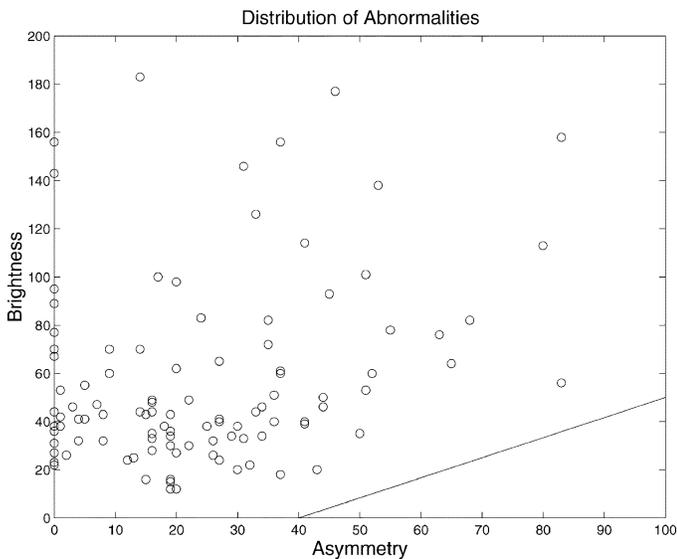


Fig. 6. Distribution of abnormal spots in the 20 training images after the process of Condition 2. To the right of the line, there are no abnormal spots. This line is used in Condition 3.

TABLE II
MEANS AND STANDARD DEVIATIONS OF THE ASYMMETRY AND BRIGHTNESS VALUES IN NORMAL AND ABNORMAL SPOTS

Parameter	Normal Spots	Abnormal Spots
Mean (SD) of asymmetry values	1.45 (5.52)	24.56 (20.69)
Mean (SD) of brightness values	20.34 (12.00)	58.19 (40.33)

brightness values were added to each part. Different scaling factors, 20 and 30, were used for asymmetry and brightness values, respectively. Therefore, initially two input-output pairs $(4, 4) \rightarrow 1$, and $(0, 0) \rightarrow 0$ were added to all six parts. Table III lists the numbers of true abnormalities, the added abnormal and normal points of the 20 training images by the radiologist's diagnosis. The samples S1, S2, S3, S6, S14, and S18 had added abnormal points in all six regions. Besides the added abnormal points, the samples S3, S3, etc. provided corrections for too many abnormalities marked by the CAD system; some normal points were added in these cases. It is noted that the diagnosis and the learning were alternately operated. One time of learning of the six CPFISs followed one diagnosis of an image. Thus, the diagnosis and learning were in an incremental mode, but not in a batch mode that only one time of learning was performed at the end of diagnosing the last image S20. It also merits attention that the added abnormal points were not necessarily the real abnormal positions in the training images.

Although these added points are actually not real abnormalities, radiologists may think some points of the training images had better be marked as abnormal points in order to increase the diagnosis performance of the proposed system.

The training was in the sequence: diagnosing and adding abnormal points to S1 \rightarrow learning \rightarrow diagnosing and adding normal points to S1 \rightarrow learning \rightarrow diagnosing and adding abnormal points to S2 $\rightarrow \dots \rightarrow$ diagnosing and adding normal points to S20 \rightarrow learning. Some images had no added abnormal/normal points and then their following learning operations were not performed. After the learning of S20, one sequence of training was finished. It is better to check if some abnormalities marked by adding abnormal points in the earlier samples may have been changed to be normal, or too many more abnormal marks have been made due to the learning of the later samples. Thus, the sequence of training can be conducted once more. In our experiment, few marks were added by the radiologist in this second sequence of training. Therefore, the training was finished after two sequences of training.

The left and right sides of Fig. 7 show the marks of S1 and S20 after the two sequences of training. There were two abnormal points (circles) on the pelvic, one abnormal point on the leg, and three normal point (triangle) on the pelvic region of S1 added by the radiologist. The CAD system labeled 23 abnormalities (squares) on S1. The four arrow-pointed locations are true abnormalities. For S20, no abnormal or normal points were added. The CAD system labeled 10 abnormalities on the whole body. One arrow-pointed location at the right hand is a true abnormality.

Figs. 8–10 show the fuzzy rules of the trained system. All the input and output membership functions are of Gaussian shape. These six figures correspond to the six CPFISs for the head, vertebrae, chest, pelvic, hand, and leg regions, respectively. In each figure, one row represents one fuzzy rule. We use the fuzzy terms *small*, *slightly small*, *medium*, *slightly large*, and *large* to label the input membership functions which centers are approximately at 0, 1, 2, 3, and 4, and the output membership functions which centers are approximately at 0.0, 0.4, 0.5, 0.6, and 1.0, respectively.

In all the six fuzzy systems, the first two fuzzy rules can be conveniently described as follows.

If asymmetry is *large*, and brightness is *large*, then y is *large*.

If asymmetry is *small*, and brightness is *small*, then y is *small*.

The first rule says that if both the asymmetry and brightness values are large, then the location is abnormal. Contrarily, the second rule says that if both the asymmetry and brightness values are small, then the location is normal. All the other five

TABLE III
RADIOLOGIST'S DIAGNOSIS OF THE 20 TRAINING IMAGES DURING TRAINING FROM IMAGE S1 TO IMAGE S20 IN A SEQUENCE

Fig.	Abnormalities	Normal Points Added	Abnormal Points Added
S1	4	3 (3 in pelvic)	3 (2 in pelvic, 1 in leg)
S2	2	0	1 (1 in chest)
S3	6	3 (2 in vertebrae, 1 in hand)	4 (1 in head, 2 in vertebrae, 1 in hand)
S4	0	2 (1 in vertebrae, 1 in hand)	0
S5	1	0	0
S6	19	0	3 (1 in chest, 1 in pelvic, 1 in leg)
S7	1	2 (2 in pelvic)	0
S8	12	2 (1 in hand, 1 in leg)	0
S9	2	0	0
S10	0	0	0
S11	1	0	0
S12	10	4 (4 in hand)	0
S13	0	0	0
S14	5	1 (1 in hand)	2 (1 in hand, 1 in leg)
S15	3	0	0
S16	1	1 (1 in head)	0
S17	2	0	0
S18	7	1 (1 in chest)	1 (1 in chest)
S19	3	1 (1 in chest)	0
S20	1	0	0

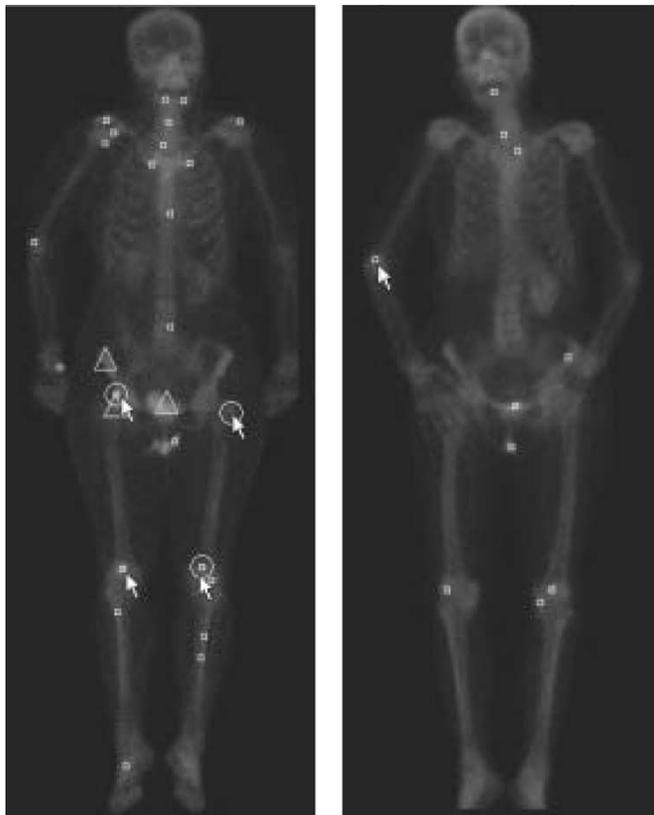


Fig. 7. The training abnormal points (circles) and normal points (triangles) were gradually added to the 20 training images. The left is the first image S1, and the right is the final image S20.

rule sets have the same two fuzzy rules. These two rules match the same knowledge as radiologists have when they inspect and diagnose scintigraphic images.

In the left-hand side of Fig. 8, the third rule can be conveniently described as follows.

If asymmetry is *small*, and brightness is *slightly large*, then y is *slightly large*.

In the right-hand side of Fig. 8 and the right-hand side of Fig. 10, the third rules can be conveniently described as follows.

If asymmetry is *small*, and brightness is *medium*, then y is *medium*.

In the left-hand side of Fig. 9, the third rule can be conveniently described as follows.

If asymmetry is *slightly small*, and brightness is *slightly small*, then y is *slightly large*.

In the right-hand side of Fig. 9, the third and fourth rules can be conveniently described as follows.

If asymmetry is *small*, and brightness is *slightly large*, then y is *medium*.

If asymmetry is *slightly small*, and brightness is *slightly small*, then y is *large*.

Finally, in the left-hand side of Fig. 10, the third and fourth rules can be conveniently described as follows.

If asymmetry is *slightly small*, and brightness is *slightly large*, then y is *medium*.

If asymmetry is *medium*, and brightness is *small*, then y is *slightly small*.

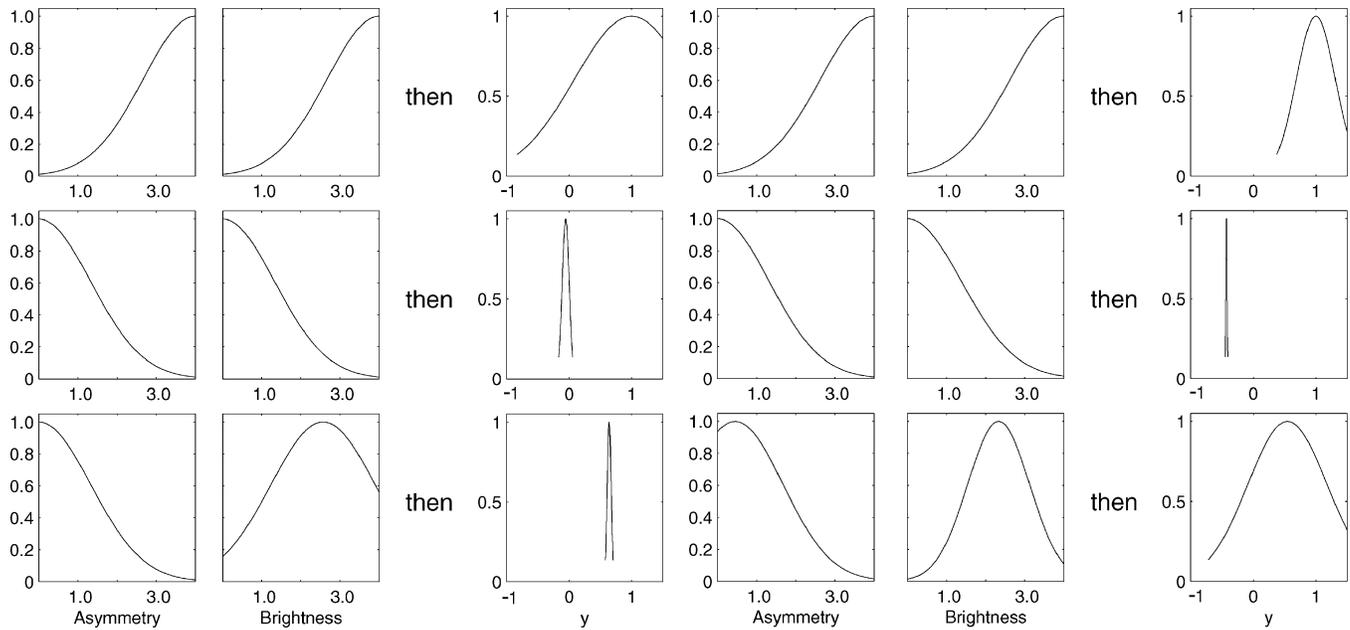


Fig. 8. Left: the three fuzzy rules for the head part; right: the three fuzzy rules for the vertebrae part after training.

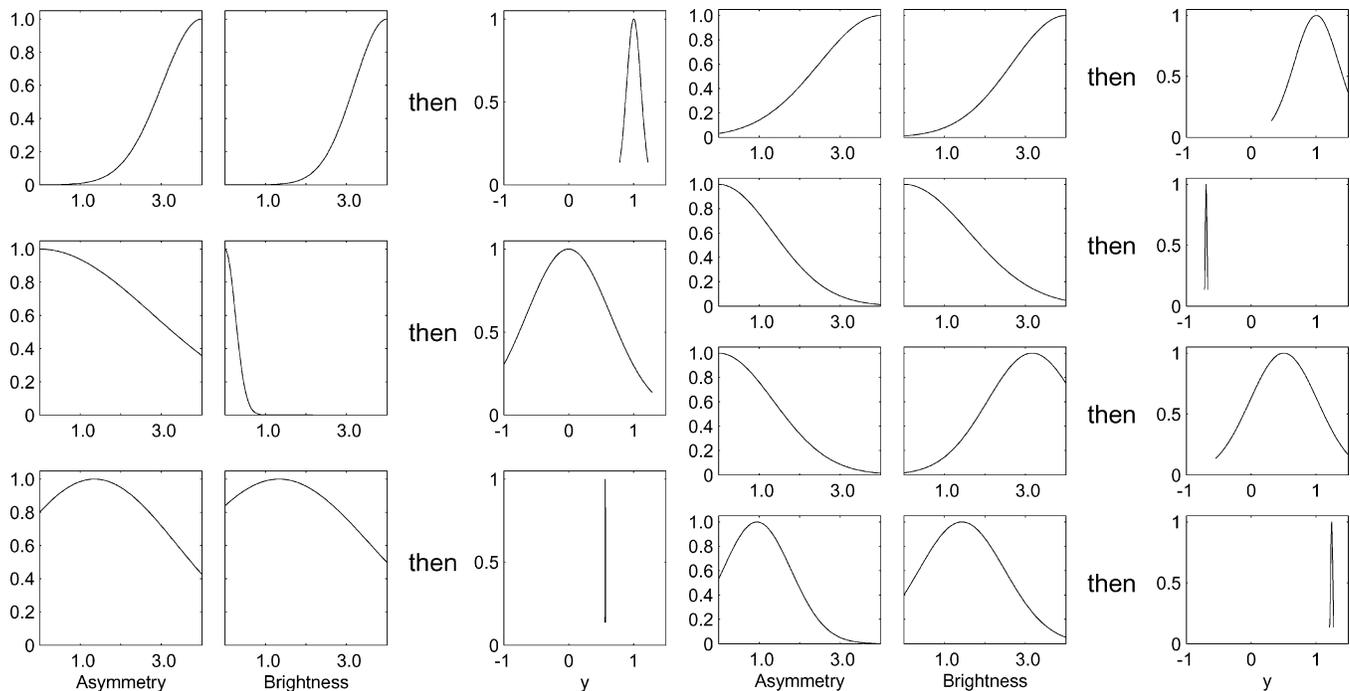


Fig. 9. Left: the three fuzzy rules for the chest part; right: the four fuzzy rules for the pelvic part after training.

All these third and fourth fuzzy rules describe diagnosis rules around the threshold $y = 0.5$. If the asymmetry or brightness values of one location are near the threshold, the local property may not be effectively described by the first two rules. Thus, by learning, the CPFIS generates the third and the fourth fuzzy rules to increase the diagnosis sensitivity around the threshold. In these fuzzy-rule figures, a spike-like output membership function makes its fuzzy rule less important than the other fuzzy rules due to the near-zero σ_j in (9). For example, the second rule in the right of Fig. 8 is less important than the other two fuzzy rules.

These diagnosis rules can be further illustrated by plotting the threshold curves on the asymmetry-brightness plane as in Fig. 11. The diagnosis of the CPFISs for the six regions are plotted in Fig. 11(a)–(f), respectively. All these plots describe a fundamental knowledge in radiologists that if both the asymmetry and brightness values of one location are large, then it is an abnormality. The y value near the left-below corner is less than 0.4. Contrarily, the y value near the right-up corner is more than 0.6. The third and fourth fuzzy rules make these threshold curves not simply left-up to right-below curves. For example, the 0.6-threshold curve in Fig. 11(d) has a turning at (1.5, 1.1).

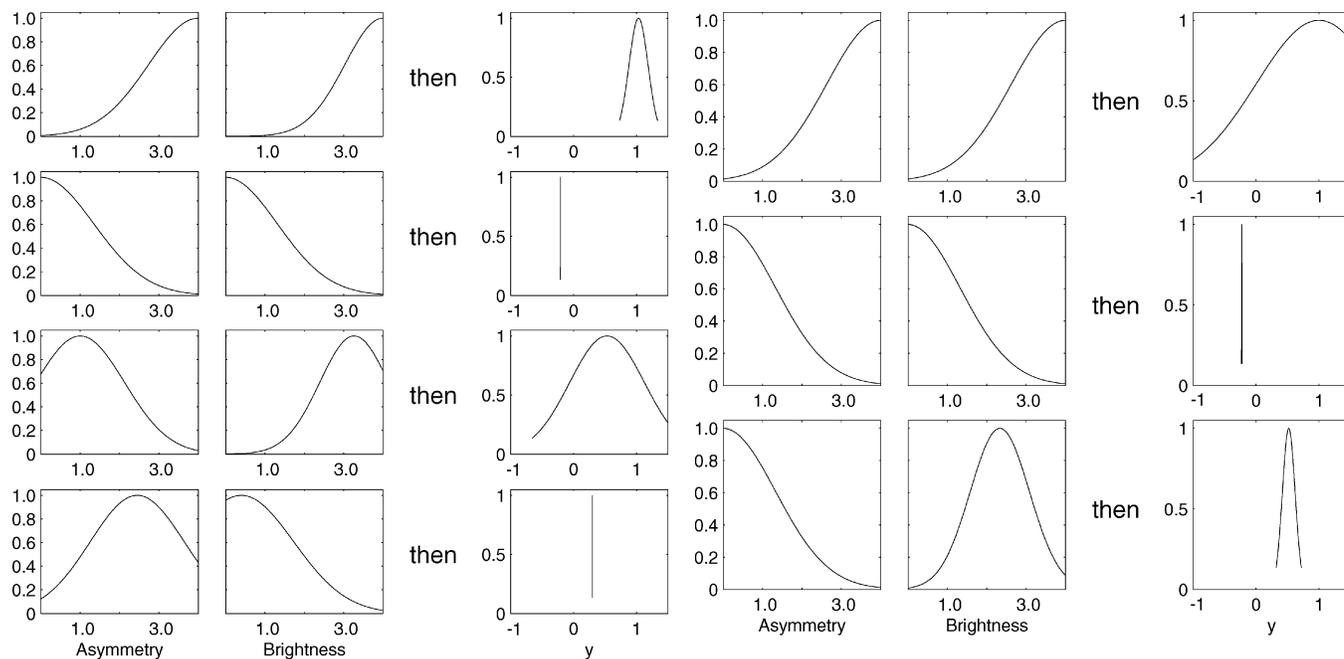


Fig. 10. Left: the four fuzzy rules for the hand part; right: the three fuzzy rules for the leg part after training.

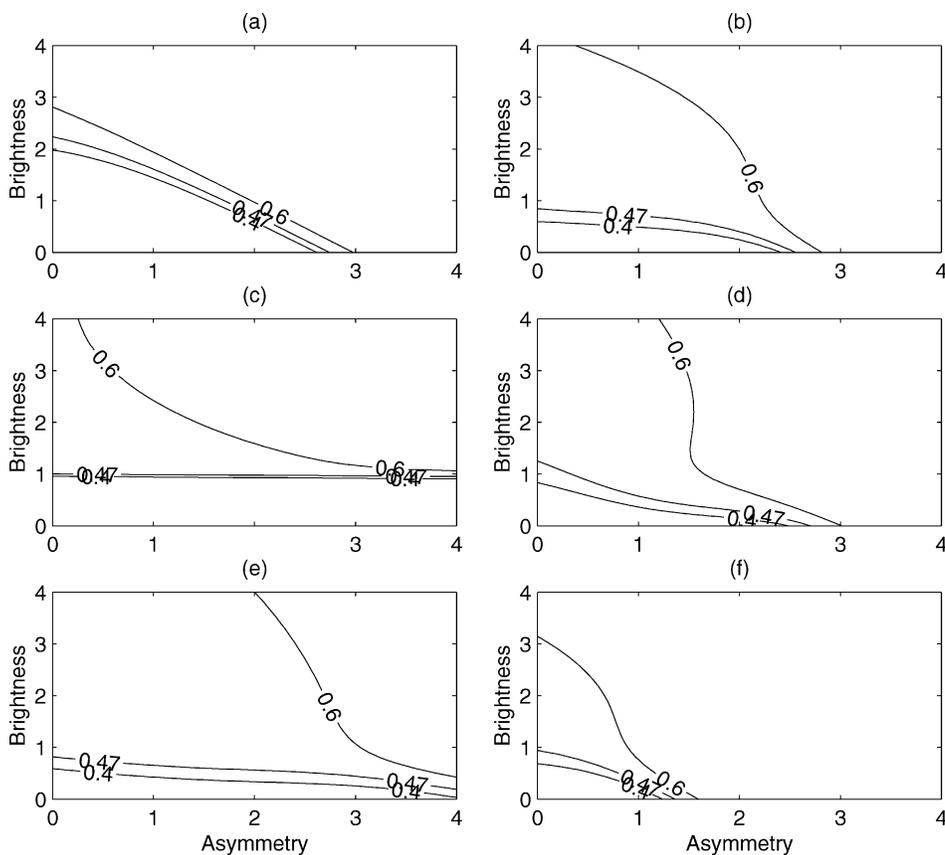


Fig. 11. The plotting of the marking made by the proposed system after training on the whole asymmetry-brightness plane. (a) Head, (b) vertebrae, (c) chest, (d) pelvic, (e) hand, and (f) leg. Three contours for $y = 0.4, 0.47,$ and 0.6 are shown.

The graphical representation of the fuzzy rules can help radiologists understand the underlying mechanism of computer diagnosis. In the following, we discuss the setting of the three minimization algorithms as in Section III.

1) *Gradient-Projection Algorithm:* The training data keyed in by radiologists as in Table III were further processed by the

abnormal-point cancellation as described in Section IV. This operation was performed before each time of learning. The size of training data was small in this experiment. Nonetheless, it was enough to build an effective CPFIS. For example, the numbers of training data were 4, 7, 7, 10, 12, and 6 for the six CPFISs in the six regions of the body, respectively.

TABLE IV
ABNORMAL LOCATIONS OF THE 82 TESTING (ABNORMAL) IMAGES

Abnormal Location	Image No.
Clavicle	4,15,24,26,37,43,47,54,65,81
Costochondral	6,49,52,63
Di us	28,74,78
Feet	8,11,16,24,70
Femur	3,10,26,36,39,41,47,60,71
Hand	3,22,27,45,46,47,52,67,81
Kidney	13,79
Knee	1,3,8,9,18,21,23,25,27,31,32,45,47,55,57,68,72,76
Mandible	2,4,24,37
Pelvic	1,7,8,14,17,20,26,29,30,34,41,47,50,51,56,59,71,80
Pubis	65
Rib	1,7,16,18,20,24,46,50,53,64,66,69,77,82
Sacrum	30
Scapula	30,37,54
Shoulder	33,50,54,55,63,82
Skull	17,23,42,48
Spine	1,3,10,12,15,19,25,31,34,35,41,44,49,50,51,58,62,66,68,73,75,76,79,82
Sternum	5,10,19,40,45,59,73
Stomach	48,72
Thyroid	19,27,38,41,61,76
Tibia	1,79

Suppose the number of data in some region is n . Initially, the spread of the output fuzzy membership function σ_j was set to be $1/n$. The threshold of σ_j was set to be 10^{-8} . If any σ_j was less than this threshold, then $\sigma_j \approx 0$. Thus, from (10), $\sigma_j \geq 0$ became an active constraint. The corresponding training data for this active constraint, $\sigma_j \geq 0$, was removed from the candidates of being characteristic points. Initially, all n training data were all candidates, but after 200 epochs of the gradient-projection algorithm, there remained n' candidates, $n' \leq n$. After processing, the algorithm not only reduced the number of candidates, but also provided relative weights between training data by the final values of σ_j .

2) *Gauss-Jordan-Elimination-Based Column Elimination*: From (13), the column elimination is based on $B(i, j)$, which is decided by σ_j values after the gradient-projection algorithm. When the threshold of column sums increases, more candidate points are removed. The threshold of column sums was initially set to be 0.001, and then doubled at each iteration until the threshold was $0.001 \times 2^6 = 0.064$ after six epochs. After the algorithm, the number of the CPs were determined. Let it be n'' , $n'' \leq n'$. In our experiment, n'' was either 3 or 4. n'' is the number of fuzzy rules.

3) *Back-Propagation Tuning*: After the n'' number of CPs were found, the other parameters were tuned by back-propagation learning. There are $n'' \times 2 = 2n''$ number of $\sigma_{in,r(j)}^{(k)}$, the spreads of the membership functions of the input fuzzy sets, $n'' \times 1 = n''$ number of $m_{r(j)}$, the means of the membership functions of the output fuzzy sets, and $n'' \times 1 = n''$ number of

$\sigma_{r(j)}$, the spreads of the membership functions of output fuzzy sets. The back-propagation learning was performed with 20 000 epochs. It took less than 20 s of computation on a PC of Pentium IV 2.4-GHz CPU, and 512 MB memory. The operating system is Windows 2000 Server, and the programming software is JBuilder 4.

C. Testing

There were 82 abnormal images and 27 normal images for testing. The real abnormal locations in the 82 abnormal images are listed in Table IV. The abnormal locations are broadly distributed over the whole body from skull to knee. Among these locations, spine is the area containing the largest number of real abnormalities, followed by pelvic and knee.

We employed a free-response receiver operating characteristic (FROC) [34], [11], [15] in terms of true positive (TP) fraction and the mean number of FPs per image to evaluate the performance of the system. Fig. 12 shows the diagram of the FROC curve. When the mean numbers of FPs per image increase from 2.0 to 63.6, the sensitivities increase from 0.359 to 0.960. Table V shows the performance of the CAD system at thresholds 0.47 and 0.5. At the threshold 0.47, the sensitivity and the mean number of FP per image are 0.92 and 37, respectively. If the threshold is set to a higher value, 0.5, the sensitivity and the mean number of FP per image are 0.86 and 24, respectively.

Tables VI and VII show the CAD performance in different regions when the threshold is set to be 0.47. There are 248

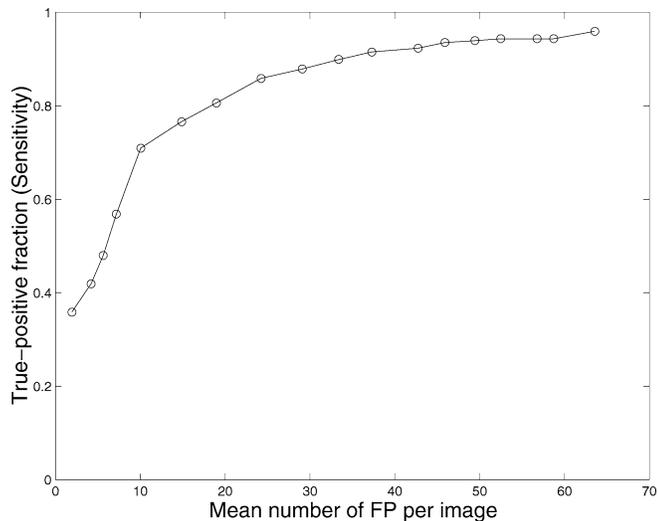


Fig. 12. The diagram of the FROC curve for the proposed CAD system.

TABLE V
PERFORMANCE OF THE CAD SYSTEM AT THRESHOLDS 0.47 AND 0.5

Threshold	Sensitivity	FP per image
0.47	0.92 (227/248)	37 (4065/109)
0.5	0.86 (213/248)	24 (2642/109)

TABLE VI
DETECTION RATES AND FP MARKS OF THE 82 ABNORMAL TESTING IMAGES IN SIX PARTS AT THRESHOLD 0.47

Region	Abnormalities	Detected	FP marks
Head	7	4 (57%)	46
Vertebrae	47	40 (85%)	918
Chest	40	38 (95%)	390
Pelvic	50	43 (86%)	402
Hand	53	52 (98%)	531
Leg	51	50 (98%)	526
Total	248	227 (92%)	2813
Av. per image	3.0	2.8 (92%)	34

true abnormalities on the 82 abnormal images. Except head, the number of abnormalities in the other five regions are about the same. Both the hand region and the leg region have the highest detection rate, 98%, while the head region has the lowest detection rate, 57%. The other three regions have more than 85% detection rates. Totally, 227 abnormalities are detected. The detection rate is 92% (227 of 248). Besides detection rates, the numbers of FP marks in all regions are also listed. In the 82 abnormal images, the vertebrae region has the largest number of FP marks, 918, while the head region has the smallest number of FP marks, 46. In the 27 normal images, the vertebrae region has the largest number of FP marks, 410, while the head region has the smallest number of FP marks, 22. From these figures, the proposed CAD system performs better in the hand and leg regions, but worse in the head and vertebrae regions.

In many images, the heads are slightly rotated which makes the asymmetry variable less effective in comparing the left and

TABLE VII
FP MARKS OF THE 27 NORMAL TESTING IMAGES IN SIX PARTS AT THRESHOLD 0.47

Region	FP marks
Head	22
Vertebrae	410
Chest	203
Pelvic	215
Hand	210
Leg	192
Total	1252
Av. per image	46

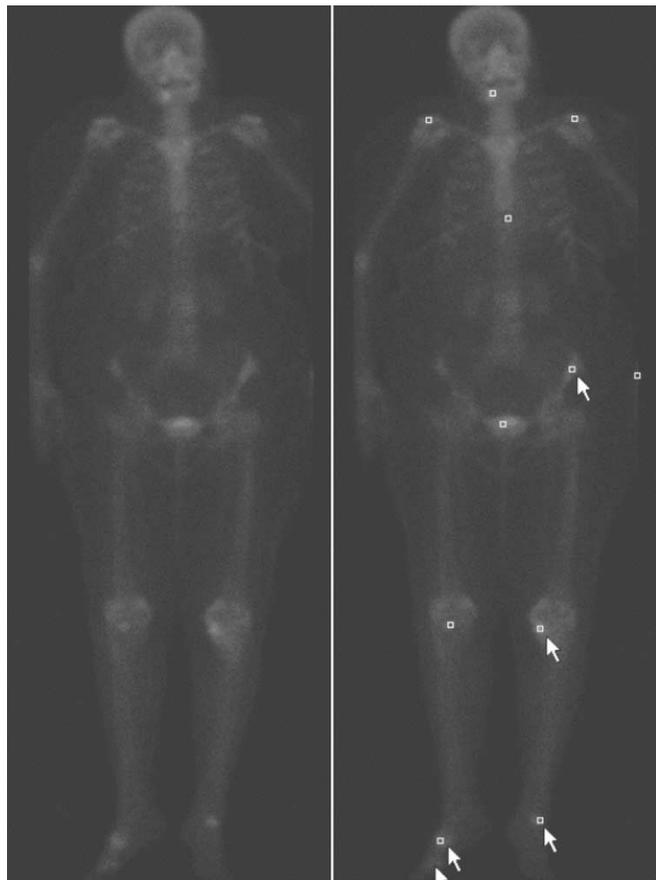


Fig. 13. Diagnosis result of abnormal image no. 8. Left: original, and right: diagnosis. The five arrow-pointed locations are true abnormal locations of the patient. There are 11 marks. Four of them are true abnormalities, and the other seven are FPs. One FN is at the right leg. The TP at the left pelvic is easy to be missed because normal pelvic at this location is also bright.

the right parts of head. For vertebrae, it is situated at the central line of body. Many abnormalities in vertebrae are at or near the central line or in equal distances to the central line in pairs. The asymmetry values in vertebrae tend to be zero in these cases. These properties partially account for the worse performance in the head and vertebrae regions. We would like to demonstrate the assistance given by the proposed system in finding and scoring minute abnormalities in the following two figures.

The first figure is Fig. 13, the no. 8 abnormal image of the 82 abnormal testing images. The five arrow-pointed locations

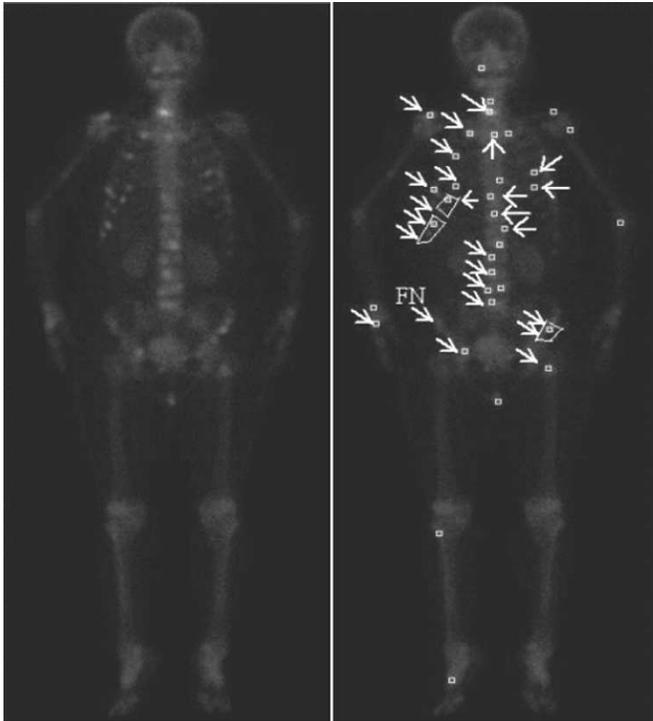


Fig. 14. Diagnosis result of abnormal image no. 28. Left: original, and right: diagnosis. The 26 arrow-pointed locations are true abnormal locations of the patient. There are 35 marks. Twenty-two of them are true abnormalities, and the other 13 are FPs. One FN is at the right pelvic. There are three TPs which are included in the same local-maximum areas of the other three neighboring marks as indicated by the three polygons. Although not all true abnormalities are labeled, the 35 warning marks can aid radiologists not to miss some abnormal locations when the patients have multiple abnormal locations.

are true abnormal locations of the patient. Eleven squares were marked in the right-hand side of Fig. 13. Four of them are true abnormalities, and the other seven are FPs. One FN is at the right leg. The TP at the left pelvic is easy to be missed because normal pelvic at this location is also bright. It was correctly marked by the CAD system. Radiologists were cautioned by the proposed system to inspect this location for not missing this highly possible abnormality.

The second figure is Fig. 14, the no. 28 abnormal image of the 82 abnormal testing images. The patient has a diffuse pattern that 26 abnormalities exist in rib, spine, pelvic, and hand. There are 35 marks. Twenty-two of them are true abnormalities, and the other 13 are FPs. There are three TPs which are included in the same local-maximum areas of the other three neighboring marks as indicated by the three polygons. Only one FN is at the right pelvic. Although not all true abnormalities are labeled, the 35 warning marks can aid radiologists not to miss some abnormal locations when the patients have multiple abnormal locations. When the user clicks the mouse on a mark, the asymmetry, brightness, and score values are shown in the status bar of the image by the system. The automatic marking and scoring ability of the system aid radiologists by providing extra information from a second reader.

VI. CONCLUSION

In this paper, a CAD system for bone scintigraphy was proposed. After consulting radiologists, asymmetry and brightness

were the two chosen variables to be the inputs of the CAD system. Since the diagnosis is by areas but not by pixels, their values were calculated on a number of not overlapped connected areas. These areas were formed by a local-maximum-based segmentation on the whole body. Also, in order to calculate the asymmetry, a method to generate the vertically central line of body was presented. A CPFIS was employed to implement the CAD system. Three minimization algorithms, gradient-projection method, Gauss–Jordan-elimination-based column elimination, and back-propagation tuning, were used to train the proposed CPFIS. By these algorithms, the number of fuzzy rules were gradually decreased from the number of training samples to a small number. To make the learning of CPFIS more effective, abnormal-point cancellations were performed to filter off improper samples keyed in by radiologists. The whole body was roughly divided into the head, vertebrae, chest, pelvic, hand, and leg regions, and six CPFISs were used for each region. Experiments were conducted to test the proposed method. Twenty samples were used for training, and another 82 abnormal images and 27 normal images were used for testing. We employed a FROC method with the mean number of false positives (FPs) and the sensitivity as performance indexes to evaluate the proposed system. The sensitivity was 91.5% (227 of 248) and the mean number of FPs was 37.3 per image. By the different detection rates and numbers of FP marks in six regions, the proposed CAD system performed better in the hand and leg regions, but worse in the head and vertebrae regions. Two of the testing images were further elucidated to demonstrate the assistance provided by the CAD system in finding minute abnormalities. The high sensitivity and moderate numbers of FP marks per image shows that the proposed method can provide an effective second-reader information to radiologists in the diagnosis of bone scintigraphy.

APPENDIX I GRADIENT-PROJECTION METHOD

The gradient-projection method is a method for solving constrained minimization problems. The following gradient-projection algorithm is modified from [35] to tackle our problem. The modification is primarily for the settings of precision thresholds of numerical calculations. Let $\sigma_{\text{threshold}}$ and $d_{\text{threshold}}$ be some chosen thresholds for the spreads of the membership functions of output fuzzy sets, σ_j , and the elements of the projection vector \mathbf{d} of the gradient, respectively. $\sigma_{\text{threshold}}$ and $d_{\text{threshold}}$ are set to be 10^{-8} in this paper. If $\sigma_j < \sigma_{\text{threshold}}$, then the constraint $\sigma_j \geq 0$ is taken as an active constraint. Let n be the number of training data, and the vector $\boldsymbol{\sigma}$ be $[\sigma_1, \dots, \sigma_n]^T$. The function f in the following is the (10).

Step 1) Let $I = \{i | 1 \leq i \leq n, \sigma_i < \sigma_{\text{threshold}}\}$. A matrix \mathbf{A}_q is set as

$$\begin{bmatrix} 1 & 1 & \dots & 1 & 1 \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & -1 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}. \quad (15)$$

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 \mathbf{A}_q : $[0 \cdots 0 - 1 \ 0 \cdots 0]$, -1 being at the j th column. Thus, \mathbf{A}_q is a matrix of $|I| + 1$ rows and n columns.

Step 2) Calculate $\mathbf{P} = \mathbf{I} - \mathbf{A}_q^T (\mathbf{A}_q \mathbf{A}_q^T)^{-1} \mathbf{A}_q$, and $\mathbf{d} = -\mathbf{P} \nabla f(\boldsymbol{\sigma})^T$.

Step 3) If $\max_{1 \leq i \leq n} |d_i| \geq d_{\text{threshold}}$, find α_1 and α_2 achieving, respectively

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

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

Set $\boldsymbol{\sigma}$ to $\boldsymbol{\sigma} + \alpha_2 \mathbf{d}$ and return to Step 1).

Step 4) If $\max_{1 \leq i \leq n} |d_i| \leq d_{\text{threshold}}$, find $\boldsymbol{\lambda} = -(\mathbf{A}_q \mathbf{A}_q^T)^{-1} \mathbf{A}_q \nabla f(\boldsymbol{\sigma})^T$.

Case (a) If $\lambda_j \geq 0$, for all j corresponding to active constraints, stop;

$\boldsymbol{\sigma}$ satisfies the Kuhn-Tucker conditions,

Case (b) Otherwise, delete the row from \mathbf{A}_q corresponding to the constraint with the most negative component of $\boldsymbol{\lambda}$ and return to Step 2).

After the algorithm, σ_j are decided, which can be taken as weights between training data.

APPENDIX II

GAUSS–JORDAN–ELIMINATION–BASED COLUMN ELIMINATION

The Gauss–Jordan elimination is a method to solve linear equations [36]. We modified the method to perform column elimination. From (13), we have $\mathbf{C}\mathbf{m} = \mathbf{D}$, 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 algorithm in Appendix I. Thus, s is the number of training data minus the number of active constraints. Choose a threshold $C_{\text{threshold}} > 0$.

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 $\sum_{1 \leq j \leq s} |C_{ji}| < C_{\text{threshold}}$, then the i th column of \mathbf{C} and the i th element of \mathbf{m} are both deleted, and go to Step 5); otherwise, find $p = \arg \max_{1 \leq j \leq s} |C_{ji}|$.

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 $-F_{ji}/F_{ii} \times F_{ik}$ to F_{jk} , $k = i, \dots$, (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).

After the algorithm, the row dimension of \mathbf{m} , s , is decided. The larger value $C_{\text{threshold}}$ is, the smaller number s is.

APPENDIX III

BACK-PROPAGATION TUNING

For convenience, the derivatives in the back-propagation tuning are listed in the following:

$$\frac{\partial A(\mathbf{x}_i, \mathbf{x}_{\mathbf{r}(j)})}{\partial \sigma_{in,r(j)}^{(k)}} = A(\mathbf{x}_i, \mathbf{x}_{\mathbf{r}(j)}) \frac{(x_i^{(k)} - x_{r(j)}^{(k)})^2}{(\sigma_{in,r(j)}^{(k)})^3}$$

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

$$\cdot (-1) \cdot \left[\frac{m_{r(j)} \sigma_{r(j)} \frac{\partial A(\mathbf{x}_i, \mathbf{x}_{\mathbf{r}(j)})}{\partial \sigma_{in,r(j)}^{(k)}}}{\sum_{p=1}^s \sigma_{r(p)} A(\mathbf{x}_i, \mathbf{x}_{\mathbf{r}(p)})} \right]$$

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

$$\cdot \left[\sigma_{r(j)} \frac{\partial A(\mathbf{x}_i, \mathbf{x}_{\mathbf{r}(j)})}{\partial \sigma_{in,r(j)}^{(k)}} \right]$$

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

$$\cdot (-1) \cdot \frac{\sigma_{r(k)} A(\mathbf{x}_i, \mathbf{x}_{\mathbf{r}(k)})}{\sum_{p=1}^s \sigma_{r(p)} A(\mathbf{x}_i, \mathbf{x}_{\mathbf{r}(p)})}$$

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

$$\cdot (-1) \cdot \left[\frac{m_{r(j)} A(\mathbf{x}_i, \mathbf{x}_{\mathbf{r}(k)})}{\sum_{p=1}^s \sigma_{r(p)} A(\mathbf{x}_i, \mathbf{x}_{\mathbf{r}(p)})} + (-1) \frac{\sum_{j=1}^s m_{r(j)} \sigma_{r(j)} A(\mathbf{x}_i, \mathbf{x}_{\mathbf{r}(j)})}{\left[\sum_{p=1}^s \sigma_{r(p)} A(\mathbf{x}_i, \mathbf{x}_{\mathbf{r}(p)}) \right]^2} \cdot A(\mathbf{x}_i, \mathbf{x}_{\mathbf{r}(k)}) \right].$$

The algorithm is an iteration of the updating of parameters, $\sigma_{in,r(j)}^{(k)}$, $m_{r(j)}$, and $\sigma_{r(j)}$

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

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

$$\begin{aligned}\Delta m_{r(k)}(t) &= -\eta \left. \frac{\partial E}{\partial m_{r(k)}} \right|_t + \alpha \Delta m_{r(k)}(t-1), \\ m_{r(k)}(t) &= m_{r(k)}(t-1) + \Delta m_{r(k)}(t) \\ \Delta \sigma_{r(k)}(t) &= -\eta \left. \frac{\partial E}{\partial \sigma_{r(k)}} \right|_t + \alpha \Delta \sigma_{r(k)}(t-1), \\ \sigma_{r(k)}(t) &= \sigma_{r(k)}(t-1) + \Delta \sigma_{r(k)}(t)\end{aligned}$$

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

REFERENCES

- [1] K. Van Laere, K. Casier, D. Uyttendaele, W. Mondelaers, C. De Sadeleer, M. Simons, and R. Dierckx, "Technetium-99m-MDP scintigraphy and long-term follow-up of treated primary malignant bone tumors," *J. Nucl. Med.*, vol. 39, no. 9, pp. 1563–1569, Sept. 1998.
- [2] T. Leitha, "Rapid changes in the scintigraphic pattern in Tc-99 m DPD whole-body scanning in metabolic bone disease," *Clin. Nucl. Med.*, vol. 23, no. 11, pp. 784–785, Nov. 1998.
- [3] S. H. Kim, S. K. Chung, Y. W. Bahk, Y. H. Park, S. Y. Lee, and H. S. Sohn, "Whole-body and pinhole bone scintigraphic manifestations of Reiter's syndrome: Distribution patterns and early and characteristic signs," *Eur. J. Nucl. Med.*, vol. 26, no. 2, pp. 163–170, Feb. 1999.
- [4] F. Pons, L. Alvarez, P. Peris, N. Guanabens, S. Vidal-Sicart, A. Monegal, J. Pavia, A. M. Ballesta, J. Munos-Gomez, and R. Herranz, "Quantitative evaluation of bone scintigraphy in the assessment of Paget's disease activity," *Nucl. Med. Commun.*, vol. 20, no. 6, pp. 525–528, June 1999.
- [5] B. D. Nguyen, B. B. Chin, and D. P. Beall, "Gardner's syndrome with bone scintigraphic and CT demonstration," *Clin. Nucl. Med.*, vol. 23, no. 4, pp. 234–235, Apr. 1998.
- [6] A. Malich, C. Marx, M. Facius, T. Boehm, M. Fleck, and W. A. Kaiser, "Tumour detection rate of a new commercially available computer-aided detection system," *Eur. Radiol.*, vol. 11, pp. 2454–2459, 2001.
- [7] H. Sittek, K. Herrmann, C. Perlet, I. Kunzer, M. Kessler, and M. Reiser, "Computer-assisted evaluation of mammography images. Initial clinical experiences" (in German), *Radiologe*, vol. 37, no. 8, pp. 610–616, Aug. 1997.
- [8] L. Garvican and S. Field, "A pilot evaluation of the R2 image checker system and users' response in the detection of interval breast cancers on previous screening films," *Clin. Radiol.*, vol. 56, no. 10, pp. 833–837, Oct. 2001.
- [9] I. Leichter, S. Fields, R. Nirel, P. Bamberger, B. Novak, R. Lederman, and S. Buchbinder, "Improved mammographic interpretation of masses using computer-aided diagnosis," *Eur. Radiol.*, vol. 10, no. 2, pp. 377–383, 2000.
- [10] B. Verma and J. Zakos, "A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques," *IEEE Trans. Inform. Tech. Biomed.*, vol. 5, pp. 46–54, Mar. 2001.
- [11] H. D. Cheng, Y. M. Lui, and R. I. Freimanis, "A novel approach to microcalcification detection using fuzzy logic technique," *IEEE Trans. Med. Imag.*, vol. 17, pp. 442–450, June 1998.
- [12] S. Yu and L. Guan, "A CAD system for the automatic detection of clustered microcalcification in digitized mammogram films," *IEEE Trans. Med. Imag.*, vol. 19, pp. 115–126, Feb. 2000.
- [13] H. Li, Y. Wang, K. J. R. Liu, S. C. B. Lo, and M. T. Freedman, "Computerized radiographic mass detection-Part I: Lesion site selection by morphological enhancement and contextual segmentation," *IEEE Trans. Med. Imag.*, vol. 20, pp. 289–301, Apr. 2001.
- [14] —, "Computerized radiographic mass detection-Part II: Decision support by featured database visualization and modular neural networks," *IEEE Trans. Med. Imag.*, vol. 20, pp. 302–313, Apr. 2001.
- [15] M. G. Penedo, M. J. Carreira, A. Mosquera, and D. Cabello, "Computer-aided diagnosis: A neural-network-based approach to lung nodule detection," *IEEE Trans. Med. Imag.*, vol. 17, pp. 872–880, Dec. 1998.
- [16] K. Sato, K. Sugawara, Y. Narita, and I. Namura, "Consideration of the method of image diagnosis with respect to frontal lobe atrophy," *IEEE Trans. Nucl. Sci.*, vol. 43, pp. 3230–3239, Dec. 1996.
- [17] M. C. Clark, L. O. Hall, D. B. Goldfog, R. Velthuisen, F. R. Murtagh, and M. S. Silbiger, "Automatic tumor segmentation using knowledge-based techniques," *IEEE Trans. Med. Imag.*, vol. 17, pp. 187–201, Apr. 1998.
- [18] Y. A. Toliás and S. M. Panas, "A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering," *IEEE Trans. Med. Imag.*, vol. 17, pp. 263–273, Apr. 1998.
- [19] H. Ganster, A. Pinz, R. Röhner, E. Wildling, M. Binder, and H. Kittler, "Automated melanoma recognition," *IEEE Trans. Med. Imag.*, vol. 20, pp. 233–239, Mar. 2001.
- [20] W. Park, E. A. Hoffman, and M. Sonka, "Segmentation of intrathoracic airway trees: A fuzzy logic approach," *IEEE Trans. Med. Imag.*, vol. 17, pp. 489–497, Aug. 1998.
- [21] K. H. Chuang, M. J. Chiu, C. C. Lin, and J. H. Chen, "Model-free functional MRI analysis using kohonen clustering neural network and fuzzy c-means," *IEEE Trans. Med. Imag.*, vol. 18, pp. 1117–1128, Dec. 1999.
- [22] P. D. Acton, L. S. Pilowsky, H. F. Kung, and P. J. Ell, "Automatic segmentation of dynamic neuroreceptor single-photon emission tomography images using fuzzy clustering," *Eur. J. Nucl. Med.*, vol. 26, no. 6, pp. 581–590, June 1999.
- [23] M. Hanmandlu, S. N. Tandon, and A. H. Mir, "A new fuzzy logic based image enhancement," *Biomed. Sci. Instrum.*, vol. 33, pp. 590–595, 1997.
- [24] P. R. Innocent, M. Barnes, and R. John, "Application of the fuzzy ART/MAP and MINMAX/MAP neural network models to radiographic image classification," *Artif. Intell. Med.*, vol. 11, no. 3, pp. 241–263, Nov. 1997.
- [25] R. P. Velthuisen, L. O. Hall, and L. P. Clarke, "Feature extraction for MRI segmentation," *Artif. Intell. Med.*, vol. 9, no. 2, pp. 85–90, Apr. 1999.
- [26] L. A. Zadeh, "Fuzzy sets," *Inform. Contr.*, vol. 8, pp. 338–353, 1965.
- [27] —, "Fuzzy algorithm," *Inform. Contr.*, vol. 12, pp. 94–102, 1968.
- [28] —, "Outline of a new approach to the analysis of complex systems and decision processes," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, pp. 28–44, 1973.
- [29] T. K. Yin, "Fuzzy modeling and control: A characteristic-point approach," Ph.D. dissertation, Purdue Univ., West Lafayette, IN, 1996.
- [30] T. K. Yin and C. S. G. Lee, "A characteristic-point-based fuzzy inference system," in *Proc. 1996 Asian Fuzzy Syst. Symp.*, Taiwan, 1996, pp. 533–538.
- [31] S. Guillaume, "Designing fuzzy inference systems from data: An interpretability-oriented review," *IEEE Trans. Fuzzy Syst.*, vol. 9, pp. 426–443, June 2001.
- [32] L. A. Zadeh, "Soft computing and fuzzy logic," *IEEE Software Mag.*, pp. 48–56, Nov. 1994.
- [33] C. T. Lin and C. S. G. Lee, "Neural-network-based fuzzy logic control and decision system," *IEEE Trans. Comput.*, vol. 40, pp. 1320–1336, 1991.
- [34] J. A. Swets and R. M. Pickett, *Evaluation of Diagnostic Systems—Methods from Signal Detection Theory*. New York: Academic, 1982.
- [35] D. G. Luenberger, *Linear and Nonlinear Programming*, 2nd ed. Reading, MA: Addison-Wesley, 1989.
- [36] S. Nakamura, *Applied Numerical Methods in C*. Englewood Cliffs, NJ: Prentice-Hall, 1993.