ISSN : 2229-4333(Print) | ISSN : 0976-8491(Online)

Pattern Recognition to Detect Breast Cancer Thermogram Images Based on Fuzzy Inference System Method

[1]Kodrat Iman Satoto, [2]Oky Dwi Nurhayati, [3]R.Rizal Isnanto, Indonesia
[1,2,3]Computer Engineering Dept., Faculty of Engineering, Diponegoro University, Semarang

Abstract
The techniques of pattern recognition to classify the characteristics of images have been based on the similarities of certain features. To detect possible breast cancer, some bioinformatics inspired thermogram features are to be used to classify the associated stages. In this research we specifically applied Fuzzy Inference System method to a total of 60 thermograms of normal breasts as well as of those in early and advance cancer. In addition four standard image processing were also used to enhance the detection capabilitity. The results show that the Fuzzy Inference System method alone has the accuracies of the average RMSE for mean value, standard deviation, entropy, skewness, and kurtosis after 100 epochs of learning fuzzy are 0,1628; 0,24; 0,36; 2,35; 0,4481 respectively. With additional wiener filtering, histogram equalization, and region growing on the original images, the average RMSE have decreased 0,0298; 0,147; 0,3424; 0,0322; 0,3476 respectively.

Keywords
Thermogram images, Fuzzy Inference System, wiener filtering, histogram equalization

I. Introduction

Thermogram images of human bodies represent the temperature distributions, which in turn indicate the associated health conditions of the persons. A 2-D array of infrared censors is the heart of the thermograph. Certain segments of a body captured in the image which appear distinctly in brightness from the rest, indicate higher thermal energy emitted and so higher bioactivities. Therefore, thermogram images have become parts of those for medical diagnosis (Ovechkim, 2003) [8].

The shapes and intensity distributions of objects of interest in a thermogram image signify their associated particular characteristics and on qualities. However, infrared says are directed to the skins, most of their energies are absorbed by the outer layers of the skins (keyserling, 2002) [3].

Breast thermography reveals the infrared heat emittion distributions differently from normal and abnormal breasts. The abnormalities can be due to the existence of cysts, infections, and malignant tumors or other fibrocystic desease. A proper image processing scheme is to be developed to enhance the readability of the thermograms to ease the doctor's diagnosis.

II. The Underlying Theory

A. Image Preprocessing

The most important technique for removal of blur in images due to linear motion or unfocussed optics is the Wiener filter. The best method to solve inverse filtering fails in some circumstances because the sinc function goes to 0 at some values of x and y. Real pictures contain noise which becomes amplified to the point of destroying all attempts at reconstruction of an Fest is to use Wiener filtering. This tool solves an estimate for F according to the following equation:

Fest(u,v) = |H(u,v)|^2.G(u,v)/(|H(u,v)|^2.H(u,v) + K(u,v)) (1)

K is a constant chosen to optimize the estimate. This equation is derived from a least squares method. An example of Wiener filtering is given in Fig. 1. The next step image preprocessing used in our research is histogram equalization. This method usually increases the global contrast of many images, especially when the usable data of the image is represented by close contrast values. Through this adjustment, the intensities can be better distributed on the histogram. This allows for areas of lower local contrast to gain a higher contrast without affecting the global contrast. Histogram equalization accomplishes this by effectively spreading out the most frequent intensity values.

The method is useful in images with backgrounds and foregrounds that are both bright or both dark. In particular, the method can lead to better views of bone structure in x-ray images, and to better detail in photographs that are over or under-exposed. A key advantage of the method is that it is a fairly straightforward technique and an invertible operator. So in theory, if the histogram equalization function is known, then the original histogram can be recovered. The calculation is not computationally intensive.

When one wishes to compare two or more images on a specific basis, such as texture, it is common to first normalize their histograms to a "standard" histogram. This can be especially useful when the images have been acquired under different circumstances. The most common histogram normalization technique is histogram equalization where one attempts to change the histogram through the use of a function b = f(a) into a histogram that is constant for all brightness values. This would correspond to a brightness distribution where all values are equally probable. Unfortunately, for an arbitrary image, one can only approximate this result. For a "suitable" function f(*) the relation between the input probability density function, the output probability density function, and the function f(*) is given by:

$$p_b(b)db = p_a(a)da \Rightarrow df = \frac{p_a(a)da}{p_b(b)} \qquad (2)$$

From equation we see that "suitable" means that f(*) is differentiable and that df/da >= 0. For histogram equalization we desire that $p_b(b)$ = constant and this means that:

$$f(a) = (2^B - 1) \bullet P(a) \qquad (3)$$

where P(a) is the probability distribution function. In other words, the quantized probability distribution function normalized from 0 to 2^B-1 is the look-up table required for histogram equalization. The last step in image preprocessing is region growing.

This method takes a set of seeds as input along with the image. The seeds mark each of the objects to be segmented. The regions are iteratively grown by comparing all unallocated neighbouring pixels to the regions. The difference between a pixel's intensity value and the region's mean, δ, is used as a measure of similarity. The pixel with the smallest difference measured this way is allocated to the respective region. This process continues until all pixels are allocated to a region.

Seeded region growing requires seeds as additional input. The segmentation results are dependent on the choice of seeds. Noise in the image can cause the seeds to be poorly placed. Unseeded region growing is a modified algorithm that doesn't require explicit seeds. It starts off with a single region A_1 – the pixel chosen here does not significantly influence final segmentation. At each iteration it considers the neighbouring pixels in the same way as seeded region growing. It differs from seeded region growing in that if the minimum δ is less than a predefined threshold T then it is added to the respective region A_j. If not, then the pixel is considered significantly different from all current regions A_i and a new region A_{n+1} is created with this pixel. Fig. 1 illustrate the effect of wiener filter and histogram equalization on breast thermogram.

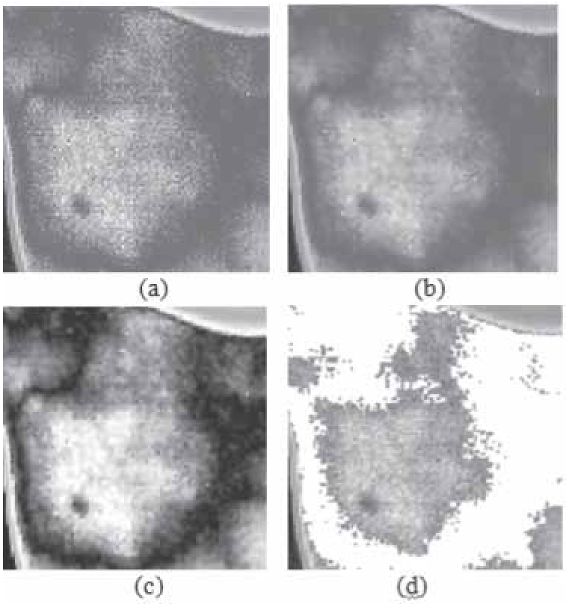

Fig. 1. (a) Original image, (b) filtering wiener, (c) filtering wiener and histogram equalization, (d) filtering wiener, histogram equalization, and region growing

B. Feature Extraction

Feature extraction is a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant then the input data will be transformed into a reduced representation set of features. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which overfits the training sample and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy. In this research, several features are extracted from thermograms. These features are mean, standard deviation, entropy, skewness, and kurtosis.

C. Classification

Fuzzy inference is the process of formulating the mapping from a given input to an output using fuzzy logic. Two types of fuzzy inference systems in the toolbox Matlab are Mamdani-type and Sugeno-type. These two types of inference systems vary somewhat in the way outputs are determined.

Fuzzy inference system is a rule based system consists of three conceptual components. These are: (1) a rule-base, contains fuzzy if-then rules, (2) a data-base, defines the membership function and (3) an inference system, combines the fuzzy rules and produces the system results. First phase of fuzzy logic modeling is the determination of membership functions of input –output variables, second is the construction of fuzzy rules and the last is the determination of output characteristics, output membership function and system results [Firat etal., 2006] [11]. Two methods, called as back propagation algorithm and hybrid-learning algorithm, provide learning of the ANFIS and construction of the rules, are used To determine the membership function of the input-output variables. The ANFIS is a multilayer feed-forward network uses ANN learning algorithms and fuzzy reasoning to characterize an input space to an output space. A general structure of fuzzy system is demonstrated in Fig. 2 The "subtractive fuzzy clustering" function offering the effective result by less rules, is applied to solve the problem in ANFIS modeling.

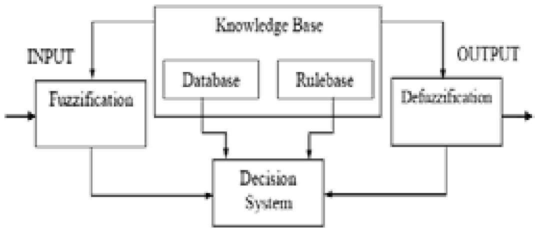

Fig. 2 : The General Structure of the fuzzy Inference System

III. Materials and Methods

The present research was performed at Dr.Sarjito Hospital Yogyakarta, 60 women were examined. Patient sits on a chair and after fifteen minutes long equilibration within air-conditioned room three pictures are taken. Digital thermal camera Fluke is used for thermogram acquisition. Three groups were assigned; Healthy Group consisted of 20 women, Early Breast Cancer Group of 20 women, and Advanced Breast Cancer Group consisted of 20 women. Several methods were applied to the image processing algorithm. They are image pre-processing with filtering wiener, histogram equalization, and segmentation with region growing method. The next method is feature extraction which extract several features of the image such as mean value, standard deviation, entropy, skewness, and kurtosis. The last method is classification of class thermograms. Features extracted is used as input classifier FIS-subtractive clustering to separate the types of thermograms. Steps of the three procedures involved in the proposed approach is shown in Fig. 3: (1) preprocessing by filtering wiener, histogram equalization, and region growing, (2) feature extraction of the thermograms, (3) Pattern recognition using fuzzy inference system subtractive clustering to group each pixel of the segments into certain clusters.

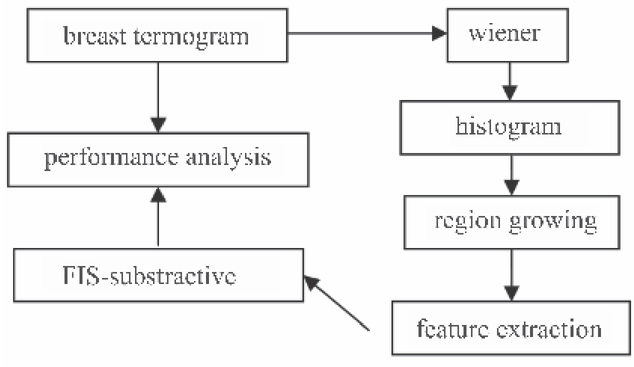

Fig. 3 : Steps of the research

IV. Results

Table 1 shows error rates occurred in the training of the FIS subtractive clustering method in each statistical value with diferrent influence range. The average RMSE (Root Mean Squared Error) for mean value, std, entropy, skewness, and kurtosis after 100 epochs of learning are 0,1628; 0,24; 0,36; 2,35; 0,4481 respectively. Fig. 4 displays the error measure (RMSE) as function of epoch number for training dataset.

Table 1 : FIS Substractive Clustering without image processing

ra	RMSE				
	Mean	Std	Entropy	Skewness	Kurtosis
0,05	0,163	0,240	0,360	2,350	0,448
0,1	0,331	0,352	0,395	0,484	0,455
0,2	0,371	0,450	0,423	0,522	0,576
0,3	0,450	0,449	0,483	0,600	0,580
0,4	0,471	0,463	0,466	0,603	0,600
0,5	0,478	0,476	0,465	0,615	0,602

Plot FIS subtractive clustering without image processing was shown in Fig. 4. The black points are data centers.

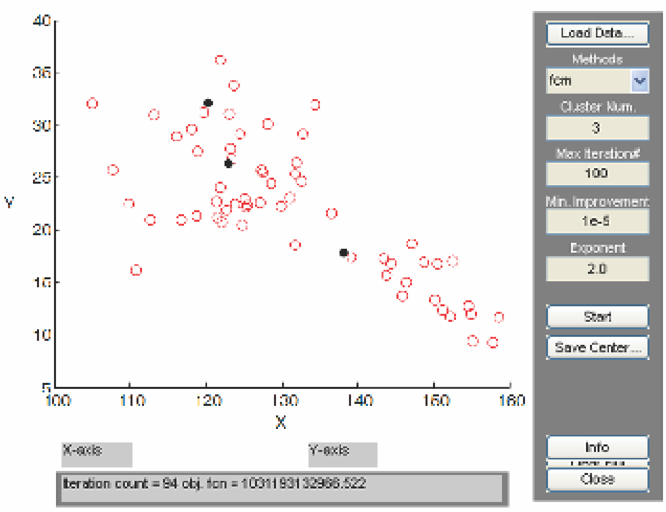

Fig. 4 : FIS clustering data with subtractive method without image processing. The black points are data centers

The comparisons of error required to extract features from a region growing method is shown in Table 2 below. Table 2 shows FIS subtractive clustering method with image processing. The average RMSE (Root Mean Squared Error) for mean value, std, entropy, skewness, and kurtosis after 100 epochs of learning are 0,0298; 0,147; 0,3424; 0,0322; 0,3476 receptively.

Table 2 : FIS Substractive Clustering with image processing

ra	RMSE				
	Mean	Std	Entropy	Skewness	Kurtosis
0,05	0.030	0.147	0.342	0.032	0.348
0,1	0.077	0.386	0.355	0.077	0.410
0,2	0.140	0.397	0.367	0.139	0.425
0,3	0.160	0.401	0.400	0.161	0.412
0,4	0.190	0.401	0.403	0.168	0.418
0,5	0.170	0.398	0.401	0.170	0.420

Plot FIS subtractive clustering with image processing shown in Fig. 5. The black points are data centers.

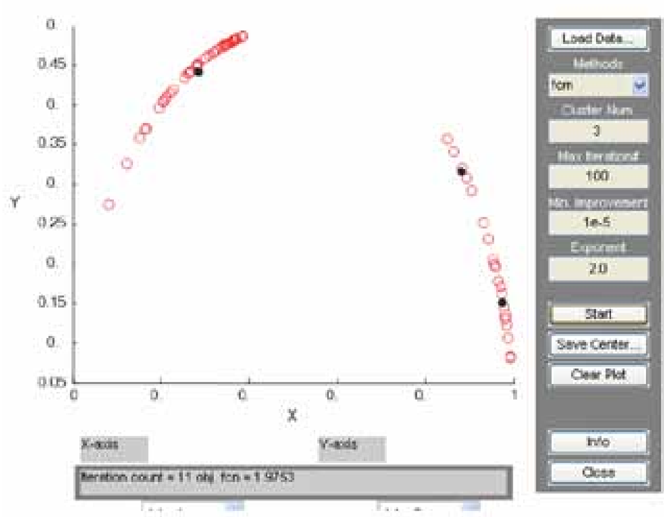

Fig. 5 : FIS substractive clustering with image processing. The black points are data centers

V. Conclusions

The goal of this work was to find the most useful method performed for increasing performance of breast thermograms. We performed comparative studies of the performances between the pattern recognition of the breast thermogram images based on FIS subtractive clustering classification with and without image processing methods. As can be seen, our classifier has good performance in FIS subtractive clustering with image processing. We compare our classifiers with two different treatments. FIS subtractive clustering with image processing could reduced error rate of classification.

References

[1] Jain, A.K., "Fundamentals of Digital Image Processing", Prentice-Hall, Inc., A Division of Simon & Schuster Engelwood Cliffs, New Jersey, 1989

[2] Gonzalez, R.C., Richard E. Woods, "Digital Image Processing", Prentice-Hall, Inc., Upper Saddle River, New Jersey, 2008.

[3] Keyserling, J.R, P.D.Ahlgren, E.Yu, N.B.Belliveau, "Overview of functional infrared imaging as part of a multi-imaging strategy for breast cancer detection and therapeutic monitoring", Proc. 2nd Joint IEEE EMBS/BMES Conf., Houston, TX, pp. 1126-8, 2002.

[4] Schalkoff, R.J., "Digital Image Processing and Computer Vision", John Wiley & Sons, Inc, Monticello, 1999.

[5] Sivanandam, S.N, S.Sumathi, S.N Deepa, "Introduction to Fuzzy Logic using MATLAB". [Online] Available: http://www.google/Introduction to Fuzzy Logic using Matlab.pdf.

[6] Stoitsis,J.,I.Valavanis, S.G.Mougiakakou, S.Gole-matti, A.Nikita, K.S.Nikita, "Computer aided diagnosis based on medical image processing and artificial intelligence methods". Elsevier Journal, 2006.

[7] Ovechkim, "Infrared Thermography as a Method to Diagnose of the Breast Cancer", International medical online journal, 2(1), 2003

[8] Ovechkim, Yoon G, "New Possibilities of infrared Thermography in Diagnosing Regional Metastases of Breast Cancer". International medical online journal, 2(1)2003:17-33, ISSN 1684-2989.

[9] Jang, J. S. R., C. T. Sun, "Neuro-fuzzy modeling and control," Proceedings of the IEEE, March1995. pp. 378-406.

[10] Fırat M, Güngör M. "River flow estimation using adaptive

neuro-fuzzy inference system, Mathematics and Computers in Simulation", 2006

[12] Tan, T.Z., C.Quek, G.S. Ng, E.Y.K.Ng, "Breast Cancer Diagnosis Using Thermography and Complementary Learning Fuzzy Neural Network". School of Computer Engineering, Nanyang Tech-nological University, 2004.

Born in Madiun, October 28th, 1963. Graduated from Department of Electrical Engineering, Faculty of Engineering, Gadjah Mada University, Yogyakarta. Master from Post Graduate Program, Department of Electrical Engineering, Gadjah Mada University, Yogyakarta, Indonesia in 2000. From 1993-present, He's a lecturer in electrical engineering, Faculty of Engineering. In the period 2008-present, He's also a lecturer in the Department of Computer Engineering, Faculty of Engineering, Diponegoro University, Indonesia. Current and interest research are information and communication technology, and software engineering.

Born in Semarang, 2nd October 1979. Graduated from the Department of Telecommunication Engineering, Faculty of Engineering, Telecommunication Engineering Institute, Indonesia in 2002. Master from Post Graduate Program, Department Electrical Engineering, Gadjah Mada University, Indonesia in 2008. In the period of 2008— present she is a Ph. D. Candidate in the Department of Electrical Engineering, Faculty of Engineering, Gadjah Mada University, Yogyakarta, Indonesia. From 2009—present she is a lecturer in the Department of Computer Engineering, Faculty of Engineering, Diponegoro University, Indonesia. Current and interest research is in biomedical image processing and software engineering. Oky Dwi Nurhayati, S.T, M.T., is also a member of IAENG (International Association of Engineer, and IMAVIS (International Journal on Image and Vision Computing).

Born in Wonosobo, July 27th 1970. Graduated from Department of Electrical Engineering, Faculty of Engineering, Gadjah Mada University, Yogyakarta, Indonesia in 1994. Master from Post Graduate Program, Department of Electrical Engineering, Gadjah Mada University, Indonesia in 2002. In the period of 2009— At present he is a Ph. D. Candidate in the Department of Electrical Engineering, Faculty of Engineering, Gadjah Mada University, Indonesia. From 2009— Now, he is a lecturer in the Department of Computer System, Faculty of Engineering, Diponegoro University, Indonesia. Current and interest research is in biomedical image processing. R. Rizal Isnanto.