M – 20

Construction of Fuzzy Rules of Zero Order Takagi-Sugeno-Kang Fuzzy System Using Generalized Matrix Inverse Method and Its Application for Diagnosing Breast Cancer

Weni Safitri[1], Agus Maman Abadi[2]
Department of Mathematic, Yogyakarta State University
wenisafitri.weni@gmail.com[1]

***Abstract*—**Breast cancer is one of the most cancerous in Indonesia that the mortality rate is quite high. To date, the primary causes of breast cancer are still unknown. Therefore, early detection and diagnosis are needed to determine the possibility of breast cancer. One way to detect this cancer is with mammogram image. The aims of this study is to explain steps in applicating fuzzy system for breast cancer diagnosis using mammogram image and implement the result with GUI (Graphical User Interface). Another aims of this study is to compare the result with previous research. This study using 120 mammogram images that divided into 96 training data and 24 testing data. The fuzzy system used is zero order Takagi-Sugeno-Kang with 10 input variables and 3 output variables. To optimize the result, generalized matrix inverse method is used in building fuzzy rules and weight average method for defuzzification process. The accuracy, sensitivity, and specification of the zero order Takagi-Sugeno-Kang fuzzy system are 100% for training data. For testing data, the accuracy, sensitivity, and specification of the system are 50%, 77.77%, 66.67%, respectively. Whereas the accuracy of the Mamdani fuzzy system reached 96.875% and 100% for sensitivity, and specification in training data. In the testing data, the Mamdani fuzzy system reached 91.67%, 93.75%, and 87.5% in accuracy, sensitivity, and specification respectively. Based on these result, it can be concluded that the zero order Takagi-Sugeno-Kang fuzzy system is better than the Mamdani fuzzy system in training data, but need much improvement for testing data.

Keywords:*breast cancer diagnosis,mammogram image, fuzzy system, GUI*

I. INTRODUCTION

Breast cancer is a disease which the cells (tissue) of breast are grown and spread uncontrollably. The disease occurs almost entirely in women, but men can get it, too [1]. The breast cancer is one of malignant tumors which have been common to be found, WHO (World Health Organization) numerated breast cancer is the number one killer that threaten women's health [2]. According to 2012 data of GLOBOCAN (IARC) found that breast cancer is cancer disease with percentage of recent cases (controlled by age) the highest is 43.3% and mortality percentage (controlled by age) caused by breast cancer is 12.9% [3].

To date, the primary causes of breast cancer are still unknown. Even the symptom and the growth itself are hard to detect. Symptoms usually only discovered after the stage of the cancer develops further, because usually it does not cause any complaints in the early stages. Patients feel well, they do not feel any pains, and does not interfere to the activity. This condition causes many sufferers seek treatment in a state of advanced stage of cancer. This situation can complicate the healing process even increase the risk of death for patients. It will be much easier to do treatment when breast cancer can be found earlier [1].

Perceiving the importance of early detection and diagnosis of breast cancer makes many researchers conduct research in the diagnosis of breast cancer. Gerald Schaefer *et al* conduct research that aims to diagnose breast cancer using fuzzy classification method based on data calorimetry [4]. Hossein Ghayoumi Zadeh *et al* conduct research to detect breast cancer by combining breast thermography with Complementary Learning Fuzzy Neural Network (CLFNN) methods and built with image processing [5].

Shleeg and Ellabib diagnosed breast cancer using the method of Mamdani Fuzzy Inference System. They used the nine rules compared to Sugeno Fuzzy Inference System method [6]. In the same year Ali and Ayturk conducted research that aims to diagnose breast cancer using neuro-fuzzy classification method called NEFCLASS [7]. Research conducted by [8] to diagnose breast cancer with point operation using image mammogram data and implemented with a Graphical User Interface (GUI). In this study, the fuzzy inference process carried out by the mamdani method.

Based on the description above, it can be seen that the current research on the diagnosis of breast cancer has been conducted by researchers with a variety of methods. One widely used method associated with the diagnosis of breast cancer is fuzzy logic. Fuzzy logic is an appropriate way to map an input space into an output space. Fuzzy Logic will deliver value to an uncertainty as well, sufficient, and close [9].

Basically, there are a fuzzy system that can be used as a method in diagnosing breast cancer. For example, a diagnosis of breast cancer using mamdani fuzzy system that previously had been performed by [8]. In addition to the fuzzy system mamdani, other methods that may be used is zero-order Sugeno fuzzy system. Therefore, to compare between the two fuzzy system is then carried out further research into the diagnosis of breast cancer with the zero-order Sugeno fuzzy system.

In the establishment of zero-order Sugeno fuzzy systems there is the possibility that the rules were formed less than the consequent parameter estimated. Based on the rules and the consequent parameters can form a non-invertible matrix. Matrix non-invertible is hard to solve using the usual inverse so other methods are needed to solve them. The method that can be used is the inverse matrix of generalized or often referred to as pseudo-inverse. This method is general and can be used to find the inverse of any type of matrix including a non-invertible matrix. Based on the description above, underlie the author to conduct research on zero-order Sugeno fuzzy system in the diagnosis of breast cancer with generalized matrix inverse method for determining parameters of the consequent of fuzzy rules.

II. RESEARCH METHOD

The data used in this study were 120 mammograms a secondary data extraction results obtained from the image database Mammographic Image Analysis Society (MIAS) through the website http://peipa.essex.ac.uk/pix/mias. The data have previously been used by [8]. Based on the data of 120 mammograms were used in this study were divided into two, namely 96 training data and 24 testing data. Step-by-step analysis of the data presented in Figure 1.

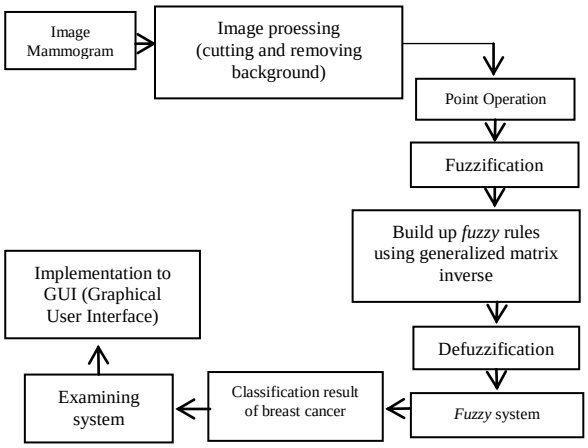

FIGURE 1. STEP STUDY CHART

After the built fuzzy system has good accuracy, the next step is to create the look of the system with a GUI (Graphical User Interface) so that the display system will be more attractive. The initial design GUI (Graphical User Interface) for breast cancer shown by Figure 2.

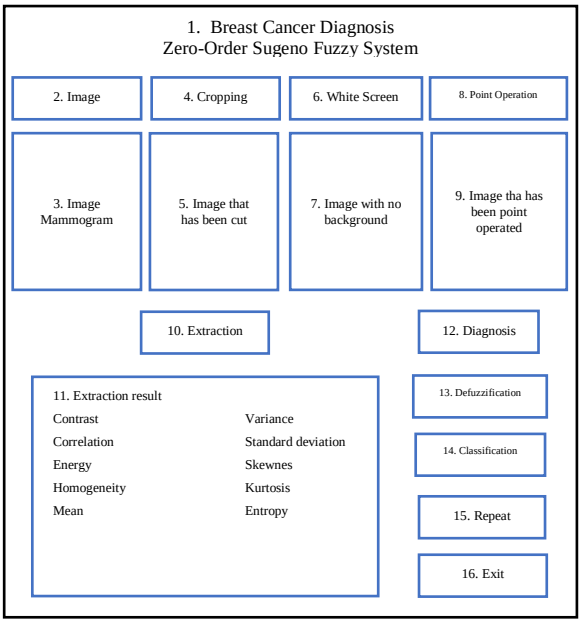

FIGURE 2. PRELIMINARY DRAFT GUI DISPLAY

III. DISCUSSION

The data used in this study is the mammogram image data that has previously been done image processing includes cutting, elimination of background and image enhancement with point operation. Image data that has been operated on point then extracted image so as to obtain 10 feature extractions, ie contrast, correlation, energy, homogeneity, mean, variance, standard deviation, skewnes, kurtosis, and entropy. Ten feature extractions are used as input variables. While the output variables used in this study was 3, which normally, benign and malignant. Image processing to obtain the extracted image has been done by [8].

A. Fuzzification

The next stage identifies the set of rules of input and output. To input a set of rules drawn from the minimum value and a maximum value of data from each extraction. The set of rules for each input that contrast = [0.134 0.235], correlation = [0.955 0.989], energy = [0.123 0.639], homogeneity = [0.939 0.979], mean [127.6 234], variance = [1973 7827], standard deviation = [44.42 88.47], skewnes = [-3.121 0.71], kurtosis = [1.36 13.13], entropy = [2.995 7.394]. While the output set of rules is in the interval [1 3]. One for normal, 2 for the benign, and 3 for malignant. Then define a fuzzy set for each input variable with 9 fuzzy set and is represented by gauss curve. Gauss curve representation for variable contrast is shown in Figure 3.

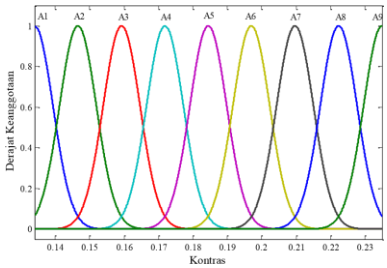

FIGURE 3. REPRESENTATION OF A SET OF FUZZY VARIABLE CONTRAST

B. *Constructing Fuzzy Rules*

Each value of the extracted image sought of the degree membership to the largest degree of membership then used to build the rules. The value of greatest degree of membership has been obtained from the respective image extraction feature are grouped in each variable fuzzy set membership degree input. The following is the result of the first training data and groupings in the fuzzy set.

TABEL 1. RESULT DATA EXTRACTION AND CLASSIFICATION 1 OF FUZZY INPUT IN SET

Image feature	Extraction result	Degree of membership	*Fuzzy* set
Contrast (x_1)	0,17765	0,562596	A_4
Correlation (x_2)	0,97352	0,701475	B_5
Energy (x_3)	0,26932	0,818787	C_3
Homogeneity (x_4)	0,96044	0,794505	D_5
Mean (x_5)	202,629	0,636253	E_7
Variance (x_6)	4336,078	0,86388	F_4
Standard deviation(x_7)	65,8489	0,968546	G_5
Skewnes (x_8)	-1,2388	0,977081	H_5
Kurtosis (x_9)	3,7231	0,650041	I_3
Entropy (x_{10})	5,5838	0,790291	J_6
Diagnosis			Normal

Thus the rules established for training data number 1 are as follows:

[Rule 1] If contrast is A_4 and correlation is B_5 and energy is C_3 and homogeneity is D_5 and mean is E_7 and variance is $F_4$and standard deviation is G_5 and skeunes is H_5 and kurtosis is I_3 and entropy is J_6 then the diagnosis is normal.

Using the same way to calculate and classify for another 95 data extraction result so that based on 96 training data which are used to construct system can be formed 96 fuzzy rules. This rule has been formed by [8]. Once obtained fuzzy ruels the next step is evaluating those rules in fuzzy rules of zero order Sugeno.

The form of fuzzy rules zero order of Sugeno is:

$$IF\ (x_1\ is\ A_{i1}\ ) \circ \ldots \circ (x_n\ is\ A_{in}\ )\ THEN\ y = k \qquad (1)$$

Where k is a constant. The fuzzy rules for fuzzy rules zero order system of sugeno become

[Rule 1] If the contrast is A_4 and correlation is B_5 and energy is C_3 and homogeneity is D_5 and mean is E_7 and variance is F_4 and standard deviation is G_5 and skewnes is H_5 and kurtosis is I_3 and entropy is J_6 then $y_1 = k_1$.

[Rule 2] If the contrast is A_4 and correlation is B_5 and energy is C_3 and homogeneity is D_5 and mean is E_7 and variance is $F_4$and standard deviation is G_4 and skewnes is H_5 and kurtosis is I_3 and entropy is J_6 then $y_2 = k_2$.

⋮

[Rule 96] If the contrast is A_5 and correlation is B_6 and energy is C_2 homogeneity is D_5 d and mean is E_2 and variance is F_5 and variance is G_6 and skewnes is H_8 and kurtosis is I_2 and entropy is J_7 then $y_{96} = k_{96}$. When $k_1, k_2, \dots, k_{96}$ are real constants which are going to be estimated. The process of determining k value using the method of generalized matrix inverse.

C. Defuzzification

The process of defuzzification in this resarch using the method of weight average with formula [10]:

$$y = \frac{\sum_{i=1}^{L} y_i(\mu_{i1}(x_1)\mu_{i2}(x_2)\dots\mu_{in}(x_n))}{\sum_{i=1}^{L} \mu_{i1}(x_1)\mu_{i2}(x_2)\dots\mu_{in}(x_n)} = \frac{\sum_{i=1}^{L} k_i(\mu_{i1}(x_1)\mu_{i2}(x_2)\dots\mu_{in}(x_n))}{\sum_{i=1}^{L} \mu_{i1}(x_1)\mu_{i2}(x_2)\dots\mu_{in}(x_n)} \quad (2)$$

The (3) equation can be stated as

$$y = \sum_{i=1}^{L} w_i k_i \quad (3)$$

where

$$w_i = \frac{\mu_{i1}(x_1)\mu_{i2}(x_2)\dots\mu_{in}(x_n)}{\sum_{i=1}^{L} \mu_{i1}(x_1)\mu_{i2}(x_2)\dots\mu_{in}(x_n)} \quad (5)$$

Then, variable of k can be sought by minimizing the function of

$$J = \sum_{k=1}^{N} \big(d(t) - y(t)\big)^2 = (d - Xk)^T (d - Xk) \quad [11] \quad (6)$$

Where d is the real *output* represented by matrix

$N \times 1$ (N the number of data) $d = \begin{bmatrix} d(1) \\ d(2) \\ \vdots \\ d(96) \end{bmatrix}$.

X is the matrix with size $N \times L$, where L is the amount of formed fuzzy rules.

$$X = \begin{bmatrix} w_1(1) & w_2(1) & \cdots & w_{96}(1) \\ w_1(2) & w_2(2) & \cdots & w_{96}(2) \\ \vdots & \vdots & \vdots & \vdots \\ w_1(96) & w_2(96) & \cdots & w_{96}(96) \end{bmatrix}$$

While k is the real constant that is going to be sought represented by $k = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_{96} \end{bmatrix}$, a matrix in size $L \times 1$. The J function will reach the minimum value if

$$d - Xk = 0 \text{ or } Xk = d \quad (7)$$

Then, can be determined parameter k using generalized matrix inverse [12]:

$$k = X^+ d \quad (8)$$

Where $X^+ = V\Sigma^+ U^T$ is generalized matrix inverse from matrix X. The variable U is orthogonal matrix $m \times m$. V is orthonormal matrix $n \times n$, and Σ is matrix $m \times n$ which all entries on the diagonal is a singular entri value where zero beyond the diagonal [12]. Based on the steps above will be obtained constants k as follows

$$k = \begin{bmatrix} 0 \\ \vdots \\ 1{,}4283 \\ \vdots \\ -0{,}0119 \\ \vdots \\ 0 \end{bmatrix}$$

Then k constant substitutable in 96 rules and gained rules for the training data number -1:

[Rule 1] If the contrast is A_4 and correlation is B_5 and energy is C_3 and homogeneity is D_5 and mean is E_7 and variance is F_4 and standard deviation is G_5 and skewnes is H_5 and kurtosis is I_3 and entropy is J_6 then $y_1 = 0$.

Further for the process of defuzzification can be calculated using equation (2). Once obtain the value of defuzzification, can be done a testing system to get accuracy level, ie: accuracy, sensitivity, and specification. The formula for the accuracy level [13]:

$$accuracy = \frac{the\ number\ of\ the\ correct\ data}{the\ amount\ of\ entirely\ data} \times 100\% \quad (9)$$

$$sensitivity = \frac{TP}{TP+FN} \times 100\% \quad (10)$$

$$specification\ \frac{TN}{TN+FP} \times 100\% \quad (11)$$

D. Result

Based on the defuzzification process done toward the train data achieved the classification result of breast cancer TP=64, TN=32, FP=0, and FN=0. Thus can be calculated the accuracy level from the training data which are the accuracy, sensitivity, and specification 100%. Whereas for the test data achieved the result of accuracy, sensitivity, and specification of the test data 50%, 77,7% dan 66,67%.

The accuracy obtained from mamdani fuzzy system had been done by [8] up to 96,875% and 100% for sensitivity and specification on the training data. For the test data in mamdani fuzzy system are up to 91,67%; 93,75%, and 87,5% for accuracy, sensitivity, and specification . The comparison of the classification result from both fuzzy systems shown at Table 2.

TABEL 2. COMPARISON OF THE RESULT ZER0-ORDER SUGENO FUZZY SYSTEM TO THE MAMDANI FUZZY SYSTEM

Examining system		*Fuzzy* system	
		Zero-order Sugeno	Mamdani
Training Data	Accuracy	100%	96,875%
	Sensitivity	100%	100%
	Specification	100%	100%
Test Data	Accuracy	50%	91,67%
	Sensitivity	77,7%	93,75%
	Specification	66,67%	87,5%

E. Implementation with GUI

The last stage in this research is the application GUI (Graphical User Interface) on a fuzzy system that has been built in order that the look of fuzzy system more attractive and interactive for users. Results GUI for all training data number -1 with normal early early diagnosis is shown in Figure 4.

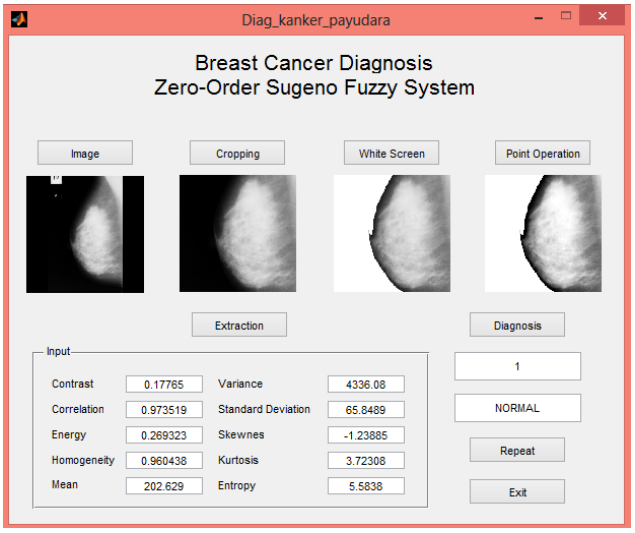

FIGURE 4. GUI DISPLAY RESULT OF TRAIN DATA 1

Figure 4 shows the classification results for the training data 1. In the column showing the image of the original image, the column cuts and a white background is an image that has been cut and removed her black background and then do the operation point on the operation point of the column. After the surgery performed extraction point in the image. Based on the extracted image is then calculated defuzzification value and classified the results. Based on Figure 4 shows that the results of data classification 1 is normal practice that has been in accordance with the initial diagnosis.

IV. CONCLUSION

The steps in constructing a zero-order Sugeno fuzzy system that starts with image processing to obtain extraction mammogram image, then fuzzification of data result of exctraction image. Based on training data used, will form 96 fuzzy rules in which the consequent determination of real constants do with generalized matrix inverse. Furthermore defuzzification process performed by the method of weight average. In the final stage, the system will be implemented fuzzy awoke with a GUI (Graphical User Interface) in order to see more attractive programs. Based on the level of accuracy obtained in this study it can be concluded that for the training data-order Sugeno fuzzy system is better in diagnosing breast cancer than using fuzzy system of mamdani. As for the test data is still needed improvements in order to improve accuracy.

Improvement and development that are likely to do in order to obtain better results of which add input variables besides extraction features such as age, weight, and quality of mammography tool that will affect the quality of the resulting image. It also adds the classification at the output, as is normal, stage 1, stage 2, stage 3 and stage 4. Furthermore, it can also make the selection of 10 extraction image is used as input to improve the accuracy and use other methods in determining the consequent parameters so that it can further optimize the results obtained.

REFERENCES

[1] S. B. Harahap, "Deteksi dini dan diagnosis dini kanker payudara," retrieved from www.mitrakeluarga.com on thursday, December 24th 2015, 19.45 p.m, 2015.

[2] Lembaga Internasional Pengobatan Kanker, retrieved from www.asiancancer.com on thursday, December 24th 2015, 20.00 p.m, 2015.

[3] Kementerian Kesehatan Republik Indonesia, "Infodatin kanker," Retrieved from http://www.depkes.go.id on tuesday, Februari 2nd 2016, 19.46 p.m, 2015.

[4] G. Schaefer., M. Zavisek, and T. Nakashima, "Thermography based breast cancer analysis using statistical features and fuzzy classifications," Pattern Recognition, 42 (6), pp. 1133 – 1137, 2009.

[5] H. G. Zadeh, O. Pakdelazar, J. Haddadnia, G. Rezai-Rad, and M. Mohammad-Zadeh, "Diagnosing breast cancer with the aid of fuzzy logic based on data mining of a genetic algorithm in infrared images," Middle East Journal of Cancer, 3(4), pp 119-129, 2012.

[6] A. A. Shleeg and I. M. Ellabib, "Comparison of mamdani and sugeno fuzzy interference system for the breast cancer risk," International Journal of Computer, Control, Quantum and Information Engineering, vol 7, no 10, pp. 695-699, 2013.

[7] A. Keles and A. Keles, "Extracting fuzzy rules for the diagnosis of breast cancer," Turkish Journal of Electrical Engineering and Computer Sciences, 21, pp. 1495-1503, 2013.

[8] K. A'yun, Optimisasi sistem fuzzy pada diagnosis kanker payudara menggunakan citra mammogram yang diimplementasikan dengan Graphical User Interface (GUI). Yogyakarta, Thesis: Yogyakarta State University, 2015.

[9] S. Kusumadewi, Analisis dan Desain Sistem Fuzzy Menggunakan Toolbox Matlab. Yogyakarta: Graha Ilmu, 2002.

[10] A. M. Abadi and D. U. Wutsqa, "Optimalisasi model neuro fuzzy untuk data time series dengan metode dekomposisi nilai singular," Jurnal Penelitian Saintek, vol. 18, nomor 1,pp. 44-54, 2013.

[11] J. Yen, L. Wang, and W. Gillespie, "Improving the interpretability of TSK fuzzy models by combining global learning and local learning," IEEE Transactions on Fuzzy System, 6(4), pp. 530-537, 1998.

[12] S. J. Leon, Linear Algebra with Applications. United States of America: Prentice-Hall International, 1998.

[13] M. Sharma and S. Mukherjee, "Artificial nueral network fuzzy inference system (ANFIS) for brain tumor detection," Advances in Intelligent Systemand Computing, vol. 177, pp. 329-339, 2013.