

Original Research Paper

A Genetic-Fuzzy System Algorithm Method for the Breast Cancer Diagnosis Problem

Normalisa^{1,2}

¹ Department of Computing, Faculty of Arts, Computing Industry Creative. Universiti Pendidikan Sultan Idris. Perak, Malaysia.

² Faculty of Computer Science Engineering, International University Liaison Indonesia. Indonesia.

Article History

Received:
30.10.2021

Revised:
17.12.2021

Accepted:
22.12.2021

*Corresponding Author:

Normalisa

Email

normalisa.halim@gmail.com

This is an open access article,
licensed under: [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/)



Abstract: Breast cancer is an important medical problem, especially for women, computer-aided medical diagnosis is very important in terms of prevention and early detection. This paper presents early detection of breast cancer using two methods, namely genetic algorithm and fuzzy inference system which will be used for early detection of breast cancer which will be used by doctors with computer assistance to obtain medical diagnosis of breast cancer in Indonesia. Our research shows that the diagnosis of breast cancer using these two methods has a high level of accuracy.

Keywords: Breast Cancer Diagnosis Problem, Computer-Aided Diagnosis, Fuzzy Inference System, Genetic Algorithm, High Level of accuracy.



1. Introduction

Breast cancer can attack in men and women. Although breast cancer cases in men are rarer, recent research indicates that breast cancer in men may be more deadly than women. "Men with breast cancer may be lacking standard care, which is different from what women get," said study author Dr. Jon Greif, a breast surgeon in San Francisco in the United States [1]. Greif said the survival rate for men who have breast cancer as a whole is lower than for women, at least when diagnosed at an early stage. It is planned that Greif will present his findings at an annual meeting of the American Society of Breast Surgeons in Phoenix. Greif and colleagues emphasize, however, that some of the differences they found might not be proven in clinical practice. Greif admitted, the research he did had limitations because they used a database of breast cancer patients who had died but it was not known what caused the death. So, it is difficult to know whether they died from cancer or something else.

Many men do not realize that they can get breast cancer. Data owned by the American Cancer Society estimates, there are about 2,200 new cases of breast cancer in men this year. It is estimated, 410 people will die of breast cancer in 2012 in the United States.

Most cancer cases in Indonesia are breast cancer which is 58,256 cases or 16.7% of the total 348,809 cancer cases in 2018. In 2019 the highest incidence of breast cancer was 42.1 per 100,000 population with an average mortality rate of 17 per 100,000 from the populations. And in 2020 Breast cancer has now overtaken lung cancer as the world's most commonly diagnosed cancer, according to statistics released by the International Agency for Research on Cancer (IARC) [2].

Early detection is very important for all diseases, especially breast cancer because early detection and diagnosis that has a high level of accuracy can make treatment success increase.

2. Literature Review

2.1. Genetic Algorithm

Genetic Algorithm (GA) method is a combination or collection of natural evolution concepts, which are usually used to initiate individuals randomly in a population [3] [4] [5] [6] [7] [8]. The Genetic Algorithm technique restarts identification techniques for optimization in search problems. This method uses a random selection process such as the evolution process. The methods often used to solve optimization problems are complex and difficult with conventional methods.

Table 1. Strength and Weakness of GA [9]

Strengths	Highly effective in forecasting.
Weakness	Being time-consuming and computationally intensive.

2.2. Fuzzy Inference System

Uncertainty is universa. For example, a computer system whose job is to recognize trees in visual images. Sources of uncertainty in this task include (but are not limited to) noise in sensing images, distortion due to poses and lenses conditions, class variability of interest (what is a "tree?"), fidelity of features used to describe trees, missing features, spatial context (trees in the forest versus trees in New York City), temporal context (trees in summer versus trees in winter), the choice of introduction algorithm, and so on [10] [11] [12] [13].

Table 2. Strength and Weakness of FIS [14]

Strengths	<ul style="list-style-type: none"> The method is strong and reliable (best) and contributes to algorithm design. Good algorithm for detects areas of image features that are less clear (faint).
Weakness	<ul style="list-style-type: none"> Change the image size during the segmentation process which allows the loss of important pixel information. Fuzzification results are affected by the fuzzy set, fuzzy set domain, and membership functions used.

If multispectral (or hyperspectral) imagery is available or multiple algorithms are applied for decision making aspects, then the problem of how to integrate compensation or even information conflicts becomes important. The historical framework for dealing with uncertainty is probability theory. It is a powerful tool that has served science well in modeling situations where s umber of major uncertainty is randomness. In some cases, we argue that uncertainty takes another form. Oftentimes, instead of asking if something is true, we ask how true, that is, how many specific properties are exhibited in specific examples. For example, we might want to know how many specific objects match the ideal prototype [13].

2.3. Genetic-Fuzzy System

The genetic algorithm rules here will utilize the Rule Base of the Fuzzy Inference System, where during the process it will not change the dataset of the database itself. In principle, using the Pittsburgh approach the rules will be that each chromosome of the genetic algorithm will represent a set of rules from fuzzy logic. The result is that each fuzzy set will form as a string in the Pittsburgh method, then it will end up in a decision table reference [15].

3. Methodology

3.1. Genetic-Fuzzy System Algorithm

The first step is to establish parameters such as populations, generations, mutation, and crossover in the genetic algorithm method process. Then form a rule-based on fuzzy logic. Crossovers and mutations will of course be determined at (0.8) and (0.01). During the process, the binary string is applied to fuzzy rules. The final output will apply the defuzzification method of fuzzy logic. The MAE method will be applied to the results of the defuzzification because it will only produce positive (1) or negative (-1) whereas if the result shows the number 0 then the defuzzification process will be ignored.

3.2. Testing and Dataset

The dataset used for this study is a dataset downloaded from the UCI dataset because there is a data center for Machine Learning and Intelligent Systems as well as many previous studies using the same data for small research scales [16].

4. Result

Overall data from the UCI dataset includes 699 samples and each will be divided into 350 training data, 209 validation data, and 140 for testing data. The data will be entered into the rules algorithm so that it becomes as as show in Table 3.

Table 3. Rule Base Genetic-Fuzzy System

Attributes	IF									THEN Class
	i	ii	iii	iv	v	vi	vii	viii	ix	
Rule 1	1-2,4-6	2,4,6	1-2,4-5	1-3,6	2,5-6	1	2,5,7	1,3-4	1,3	-1
Rule 2	1-4,6-7	3-4,6	2-4,7	1-2,5-6	1-3,6	1-2	7	1,3-4,6-7	1,4	-1
Rule 3	2-3,5-7	4-7	2,4-7	1-2,4-5,7	1,4,6	1,3-7	1,4,6	1,3-6	1,3,6-7	1

Based on the data in Table 3, it was found that the results of the accuracy rate are 96.35%.

5. Conclusions

In this research, the level of accuracy produced is 96.35%. It is very important to get a high level of accuracy in terms of disease detection. Genetic-Fuzzy System has that advantage, mainly because fuzzy rules can be extracted from classification techniques when compared to other methods. But of course, this level of accuracy can be further improved if using the combination of other algorithm methods.

References

- [1] M. N. Afifah, "10 Penyebab Kanker Payudara dan Cara Mencegahnya," Kompas, 2021. [Online]. Available: <https://health.kompas.com/read/2021/10/06/200100168/10-penyebab-kankerpayu-dara-dan-cara-mencegahnya>. [Accessed: October 2021].
- [2] Anonymous, "Kanker," WHO, 2019. [Online]. Available: <https://www.who.int/indonesia>. [Accessed: March 2021].
- [3] J. Han, and M. Kamber, *Data Mining Concept and Techniques*. USA: Morgan Kaufmant, 2006.
- [4] N. Maleki, Y. Zeinali, and S. T. A. Niaki, "A k-NN Method for Lung Cancer Prognosis with the Use of A Genetic Algorithm for Feature Selection," *Expert Systems with Applications*, vol. 164, 2021. doi: 10.1016/j.eswa.2020.113981.
- [5] D. F. Taino, M. G. Ribeiro, and G. F. Roberto, "Analysis of Cancer in Histological Images: Employing an Approach Based on Genetic Algorithm," *Pattern Analytic Application*, vol. 24, pp. 483–496, 2021. doi: 10.1007/s10044-020-00931-3
- [6] P. Chandrakar, K. Shrivasa, A. Kumar, and S. Neelam, "Design of a Novel Ensemble Model of Classification Technique for Gene-Expression Data of Lung Cancer with Modified Genetic Algorithm," *Endorsed Transactions on Pervasive Health and Technology*, vol. 7, no. 25, 2021.
- [7] S. Jijitha and T. Amudha, "Breast Cancer Prognosis Using Machine Learning Techniques and Genetic Algorithm: Experiment on Six Different Datasets," in *Evolutionary Computing and Mobile Sustainable Networks*. Singapore: Springer, 2019.
- [8] R. Roger, S. Lincoln, A. S. Adriel, M. Petrucio, M. Débora, and C. Aura, "Combining Genetic Algorithms and SVM for Breast Cancer Diagnosis Using Infrared Thermography," *Sensors*, vol. 21, no. 14, pp. 1424-8220, 2021.
- [9] W. Wu, H. Huang, Z. Zhou, and Z. Liu, *A Genetic Algorithm-Based Routing Service for Simulation Grid Multi-focus Image Fusion Based on SOFM Neural Networks and Evolution Strategies*. Berlin Heidelberg: Springer, 2005.
- [10] E. M. Golafshani, A. Behnood, S. S. Hosseinikebria, and M. Arashpour, "Novel metaheuristic-Based Type-2 Fuzzy Inference System for Predicting the Compressive Strength of Recycled Aggregate Concrete," *Journal of Cleaner Production*, vol. 320, 2021. doi: 10.1016/j.jclepro.2021.128771.
- [11] M. Parsajoo, A. S. Mohammed, S. Yagiz, D. J. Armaghani, and M. Khandelwal, "An evolutionary Adaptive Neuro-Fuzzy Inference System for Estimating Field Penetration Index of Tunnel Boring Machine in Rock Mass," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 13, no. 6, 2021. doi: 10.1016/j.jrmge.2021.05.010.
- [12] M. E. Zayed, J. Zhao, W. Li, A. H. Elsheikh, and M. A. Elaziz, "A Hybrid Adaptive Neuro-Fuzzy Inference System Integrated with Equilibrium Optimizer Algorithm for Predicting the Energetic Performance of Solar Dish Collector," *Energy*, vol. 235, 2021. doi: 10.1016/j.energy.2021.121289.
- [13] J. M. Keller, D. Liu, and D. B. Fogel, *Fundamental of Computational Intelligence*. Canada: Press Wiley, 2016.
- [14] J. Luis, G. Arroyo and B. G. Zapirain, *Computer in Biology and Medicine: Algorithm for Segmentation*. USA: Springer, 2019.
- [15] C. H. Tan, M. S. Tan, M. S. W. Chang, K. S. Yap, H. J. Yap, and S. Y. Wong, "Genetic Algorithm Fuzzy Logic for Medical Knowledge-Based Pattern Classification," *Journal of Engineering Science and Technology*, vol. 13, pp. 242-258, 2018.
- [16] T. M. Mitchell, *Machine Learning*. USA: McGraw Hill, 1997.