

Neural Networks Classification for Breast Cancer Analysis

Paquita Putri¹, Intan Nurma Yulita²
{paquitaputri0701@yahoo.com¹, intan.nurma@unpad.ac.id²}

Department of Computer Science, Padjadjaran University, Sumedang 45363, Indonesia^{1,2}

Abstract. Cancer is caused by a lump from a collection of cells that grows and attacks the surrounding tissue. Most breast lumps are benign, but benign breast lumps can increase the risk of developing breast cancer. Early detection can provide better handling. The diagnosis of breast cancer by a doctor is done by analyzing several factors. To help the doctor in diagnosing the data efficiently, this study implemented machine learning. The diagnosis was based on the features of digital image computation in the process of fine-needle aspiration (FNA) of a breast mass. The data came from 569 patients. The study used the neural networks classification method with multilayer perceptron algorithms. The results were obtained that the use of neural networks gave higher accuracy if it compared it in the ZeroR method. Their accuracies were 95.96%, and 62.74%, respectively.

Keywords: Breast Cancer, Neural Networks, ZeroR, Multilayer Perceptron

1 Introduction

Breast cancer is a type of cancer that has a high mortality rate [1]. This cancer is the most common cancer in women throughout the world and also the second most common overall [2]. This cancer is a disease in which breast cells grow out of control and attack the surrounding tissue. These cells usually form tumors that can be seen on x-ray results and felt as a lump [3].

In 2012, nearly 1.7 million cases of new breast cancer were diagnosed [4]. Cancer starts with a tumor. A tumor is a collection of abnormal cells that form lumps and grow. Tumors grow and behave differently, depending on their nature; malignant (cancer) or benign (non-cancerous) [5]. Most breast lumps are benign and not cancerous (malignant tumors). Non-cancerous breast tumors are abnormal growths, but do not spread outside the breast and are not life-threatening. But some benign breast lumps can increase the risk of developing breast cancer. Tumors are said to be malignant (cancer) if cells can grow to attack surrounding tissue or spread (metastasize) to areas of the body. Breast cancer generally attacks women, but men can also get breast cancer [6]. Malignant tumors are cells that can grow attacking the surrounding tissue while benign tumors are cells that grow but can be operated to remove the tumor without damaging the surrounding tissue [7]. The term breast cancer refers to malignant tumors that have developed in cells in the breast.

It is important for everyone to check for a lump in the breast to find out whether the lump is cancerous or not so preventive measures can be taken [8]. When detected in the initial stages, there is a 30% chance that cancer can be treated effectively, but the delay in detecting will make treatment more difficult [9]. In order to optimize cancer prevention and control

efforts, it is necessary to have massive efforts made by all parties both government and society in cancer prevention and control. One effort can be taken by utilizing technology to build applications to detect this cancer early.

When analyzing cancer, the doctor or radiologist observes several factors. Some of these factors become input for the application in making predictions. This prediction can be developed through machine learning. This method provides the ability for computers to carry out learning activities to solve problems independently. The learning process is based on processing data previously owned. This method has been successful in various aspects of life such as sleep disorders [10], education [11], protein grouping [12], transportation [13], and speech processing [14]. One method in machine learning is neural networks. This method is one of the artificial representations of the human brain that always tries to stimulate the learning process in the human brain. The algorithm that is widely used is a multilayer perceptron (MLP). This study compares MLP and zeroR in classifying types of cancer.

This study used data from the University of Wisconsin Hospital [15]. A number of studies have conducted research on this data. The methods that have been used include the ensemble decision tree [16], hybrid of K-means and support vector machines [17], Decision Tree (C4.5), Naive Bayes (NB) and k Nearest Neighbors (k-NN) [18]. Exploration in this study emphasizes MLP and its comparison with zeroR

2 Literature study

This research is based on three main methods, namely neural networks and zeroR.

2.1 Neural Networks

Artificial neural networks are computational systems whose architecture and operations are inspired by knowledge of biological nerve cells in the brain [19]. This method is described as a mathematical and computational model for non-linear approximation functions, classification of cluster data and non-parametric regression or a simulation of a collection of biological neural network models.

This method consists of several neurons and there are connections between neurons as in the human brain. The neuron is an information processing unit that is the basis of the operation of this method [20]. An artificial neural network consists of several non-linear counting elements, each of which is connected through weighting and arranged in parallel. This weighting will later change (adapt) during the training process.

Training needs to be done on this method before it is used to solve problems. The results of the training can be obtained the correct response to the input provided. Problem-solving is done through a learning process from the examples of training provided. Usually, the artificial neural networks are given a set of training patterns consisting of a set of sample patterns. The learning process of artificial neural networks comes from a series of examples of patterns given. The training method that is often used is the guided learning method. During the learning process, the input patterns are presented together with the desired output patterns. The networks will adjust its weight value in response to the input patterns and targets presented.

Weight is a value that defines the level of importance of the relationship between a node with other nodes [21]. The greater the weight of a relationship indicates the increasingly important relationship between the two nodes. Weight is a relationship in the form of real and

integer numbers, depending on the type of problem and the model used. These weights can be determined to be within a certain interval. during the training process, the weights can adjust to the input patterns.

The networks by itself will improve continuously because of the ability to learn. Every time there is a new problem, the networks can learn from the new problem, by rearranging the weight values to adjust the character values.

Each neuron has an internal state called the activation level or activity level which is an input function that is received. Typically a neuron sends its activities to several other neurons as a signal. Noteworthy is that neurons can only send one signal at a time, even though the signal can be transmitted to several other neurons. There are several choices of activation functions used in the backpropagation method, such as binary sigmoid function, and bipolar sigmoid function. Summary of dataset can be seen on Table 1.

Table 1. Summary of dataset features.

No.	Features	Max	Min	Mean	Stdev
1	Radius	28.11	6.98	14.12	3.52
2	Standard error of radius	2.87	0.11	0.40	0.28
3	Largest mean value of radius	36.04	7.93	16.25	4.82
4	Texture	39.28	9.71	19.31	4.29
5	Standard error of texture	4.89	0.36	1.22	0.55
6	Largest mean value of texture	49.54	12.02	25.69	6.14
7	Perimeter	188.50	43.79	91.91	24.29
8	Standard error of perimeter	21.98	0.76	2.86	2.01
9	Largest mean value of perimeter	251.20	50.41	107.13	33.47
10	Area	2501.00	143.50	654.28	351.92
11	Standard error of area	542.20	6.80	40.14	45.28
12	Largest mean value of area	4254.00	185.20	878.58	567.85
13	Smoothness	0.16	0.05	0.10	0.01
14	Standard error of smoothness	0.03	0.00	0.01	0.00
15	Largest mean value of smoothness	0.22	0.07	0.13	0.02
16	Compactness	0.35	0.02	0.10	0.05
17	Standard error of compactness	0.14	0.00	0.03	0.02
18	Largest mean value of compactness	1.06	0.03	0.25	0.16
19	Concavity	0.43	0.00	0.09	0.08
20	Standard error of concavity	0.40	0.00	0.03	0.03
21	Largest mean value of concavity	1.25	0.00	0.27	0.21
22	Concave points	0.20	0.00	0.05	0.04
23	Standard error of concave points	0.05	0.00	0.01	0.01
24	Largest mean value of concave points	0.29	0.00	0.11	0.07
25	Symmetry	0.30	0.11	0.18	0.03
26	Standard error of symmetry	0.08	0.01	0.02	0.01
27	Largest mean value of symmetry	1.25	0.00	0.27	0.21
28	Fractal dimension	0.10	0.05	0.06	0.01
29	Standard error of fractal dimension	0.03	0.00	0.00	0.00
30	Largest mean value of fractal dimension	0.29	0.00	0.11	0.07

One algorithm of ANN is a multilayer perceptron (MLP) [22]. MLP is a feedforward artificial neural network consisting of a number of neurons connected by connecting weights [23]. The neurons are arranged in layers consisting of one input layer, one or more hidden

layers, and one output layer. Every neuron in MLP is interconnected. Each connection has a weight that the value of each weight will be different later [24]. The hidden layer and output layer have additional "input" which is usually called bias. This algorithm can also apply a single hidden layer.

2.2 ZeroR Classifier

ZeroR is the simplest classification method that depends on the target (class) and ignores all predictors (features) [25]. ZeroR classifier only predicts the majority categories (classes). Although there is no predictive power in ZeroR, this method is useful for determining benchmark values for other classification methods [26].

3 Research Methodology

Data came from the University of Wisconsin Hospital. There are 569 data categorized into 212 malignant and 357 benign data [15]. The dataset consists of 30 features based on average values. The descriptions of the attributes are explained in Table 1. The statistics of each feature also included in the table include maximum (max), minimum (min), mean (mean), and standard deviation (stdev) values.

The architecture of MLP is explained in **Figure 1**. MLP had three layers, namely input, hidden, and output layer. MLP in this study used learning rate = 0.3, momentum = 0.2, and batch size = 100. The final stage of the study was an evaluation comparing methods. The mechanism was based on 10 fold cross-validation with parameters in the form of accuracy, error, and root mean squared error (RMSE). A high-performance id obtained if the accuracy is high, and RMSE is low. The error value is inversely proportional to accuracy.

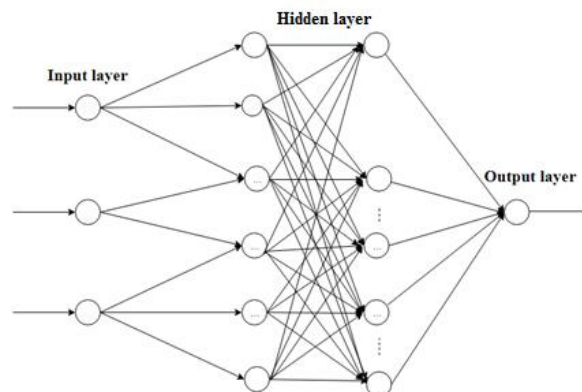


Fig. 1. Architecture of MLP.

4 Results and Discussion

Table 2 shows that zeroR produced 357 instances that were classified correctly and 212 instances that were classified incorrectly. The RMSE of zeroR was 0.48. The MLP produced 546 instances that were classified correctly and 23 instances that were classified incorrectly. The MLP error was lower than zeroR. Such was the case with RMSE from MLP which was lower than zeroR. It indicated that MLP was more appropriate for the classification of breast cancer in this study. The failure of zeroR is seen in Table 3. All data was classified as a benign class by Zero R. In contrast, MLP, which was largely able to classify data according to its actual class, as shown in Table 4.

Table 2. Comparison of Accuracy and RMSE of each method.

Method	Accuracy	Error	RMSE
ZeroR	62.74 %	37.26%	0.48
MLP	95.96 %	4.04%	0.18

Table 3. Confusion matrix of zeroR.

	Classified as	
	Malignant	Benign
Malignant	0	212
Benign	0	357

Table 4. Confusion matrix of MLP.

	Classified as	
	Malignant	Benign
Malignant	200	12
Benign	11	346

5 Conclusion

This study compared MLP and zeroR for the classification of breast cancer. The tests carried out on breast data that has 30 features. Based on Table 2, MLP had a higher performance than zeroR by using 10-fold cross-validation. The accuracy of MLP above 90% showed this method could be used to diagnose breast cancer. Unlike the case with zeroR which classified each data into a benign class. The high accuracy in MLP showed this method was appropriate for the classification in this study. The existence of this diagnosis is expected to provide early information before this cancer attacks a wider area.

References

- [1] Chen, D., Yang, H., Chen, B., Xingfei, Y. U., & Feng, W.: Study on expression of TPS, VEGF of breast cancer patients and relationship with clinical pathological. *Chinese Journal of Biochemical Pharmaceutics*, (5), 121-123 (2015)
- [2] Liu, L., Li, X., & Zhang, J.: Association between the expression of IGF1R and estrogen receptor and efficacy of neoadjuvant chemotherapy in breast cancer patients. *Zhonghua zhong liu za zhi [Chinese journal of oncology]*, 37(11), 833-836 (2015)
- [3] Rheinbay, E., et al.: Recurrent and functional regulatory mutations in breast cancer. *Nature*, 547(7661), 55 (2017)
- [4] DeSantis, C. E., Fedewa, S. A., Goding Sauer, A., Kramer, J. L., Smith, R. A., & Jemal, A.: Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. *CA: a cancer journal for clinicians*, 66(1), 31-42 (2016)
- [5] DeSantis, C. E., Ma, J., Goding Sauer, A., Newman, L. A., & Jemal, A.: Breast cancer statistics, 2017, racial disparity in mortality by state. *CA: a cancer journal for clinicians*, 67(6), 439-448 (2017)
- [6] Cardoso, F., et al.: 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *New England Journal of Medicine*, 375(8), 717-729 (2016)
- [7] Wang, D., Khosla, A., Gargeya, R., Irshad, H., & Beck, A. H.: Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718* (2016)
- [8] Bianchini, G., Balko, J. M., Mayer, I. A., Sanders, M. E., & Gianni, L.: Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nature reviews Clinical oncology*, 13(11), 674 (2016)
- [9] Lauby-Secretan, B., Scoccianti, C., Loomis, D., Benbrahim-Tallaa, L., Bouvard, V., Bianchini, F., & Straif, K.: Breast-cancer screening—viewpoint of the IARC Working Group. *New England journal of medicine*, 372(24), 2353-2358 (2015)
- [10] Yulita, I. N., Purwani, S., Rosadi, R., & Awangga, R. M.: A quantization of deep belief networks for long short-term memory in sleep stage detection. In *2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA) IEEE*. (2017)
- [11] Zhu, X.: Machine teaching: An inverse problem to machine learning and an approach toward optimal education. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*. (2015)
- [12] Yulita, I. N., & Wasito, I.: gCLUPS: Graph clustering based on pairwise similarity. In *2013 International Conference of Information and Communication Technology (ICoICT) IEEE*. (2013)
- [13] Kolouri, S., Park, S. R., Thorpe, M., Slepcev, D., & Rohde, G. K.: Optimal mass transport: Signal processing and machine-learning applications. *IEEE signal processing magazine*, 34(4), 43-59 (2017)
- [14] Yulita, I. N., Hidayat, A., Abdullah, A. S., & Awangga, R. M.: Feature Extraction Analysis for Hidden Markov Models in Sundanese Speech Recognition. *Telkonnika*, 16(5) (2018)
- [15] Bennett, K. P., & Mangasarian, O. L.: Robust linear programming discrimination of two linearly inseparable sets. *Optimization methods and software*, 1(1), 23-34 (1992)
- [16] Lavanya, D., & Rani, K. U.: Ensemble decision tree classifier for breast cancer data. *International Journal of Information Technology Convergence and Services*, 2(1), 17 (2012)
- [17] Zheng, B., Yoon, S. W., & Lam, S. S.: Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Systems with Applications*, 41(4), 1476-1482 (2014)
- [18] Asri, H., Mousannif, H., Al Moatassime, H., & Noel, T.: Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Computer Science*, 83, 1064-1069 (2016)
- [19] Chine, W., Mellit, A., Lughfi, V., Malek, A., Sulligoi, G., & Pavan, A. M.: A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks. *Renewable Energy*, 90, 501-512 (2016)
- [20] Andreatta, M., & Nielsen, M.: Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics*, 32(4), 511-517 (2015)
- [21] Da Silva, I. N., Spatti, D. H., Flauzino, R. A., Liboni, L. H. B., & dos Reis Alves, S. F.: *Artificial neural networks*. Cham: Springer International Publishing. (2017)

- [22] Zhang, Y., Sun, Y., Phillips, P., Liu, G., Zhou, X., & Wang, S.: A multilayer perceptron based smart pathological brain detection system by fractional Fourier entropy. *Journal of medical systems*, 40(7), 173 (2016)
- [23] Pham, B. T., Bui, D. T., Pourghasemi, H. R., Indra, P., & Dholakia, M. B.: Landslide susceptibility assessment in the Uttarakhand area (India) using GIS: a comparison study of prediction capability of naïve Bayes, multilayer perceptron neural networks, and functional trees methods. *Theoretical and Applied Climatology*, 128(1-2), 255-273 (2017)
- [24] Semwal, V. B., Raj, M., & Nandi, G. C.: Biometric gait identification based on a multilayer perceptron. *Robotics and Autonomous Systems*, 65, 65-75 (2015)
- [25] Rajasekaran, T., Jayasheelan, P., & Preethaa, K. S.: Predictive Analysis in Agriculture to Improve the Crop Productivity using ZeroR algorithm. *International Journal of Computer Science and Engineering*, 4(2), 1397-1401 (2016)
- [26] Rani, A. S., & Jyothi, S.: Performance analysis of classification algorithms under different datasets. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom) IEEE*. (2016)