Breast Cancer Classification: Comparison of Bayesian Networks, Multilayer Perceptron, and Boosting Method

Intan Nurma Yulita[1], Shofiyyah Nadhiroh [2]
{intan.nurma@unpad.ac.id[1], shofiyyah.nadhiroh@gmail.com[2]}

Department of Computer Science, Padjadjaran University, Sumedang 45363, Indonesia[1,2]

Abstract. Cancer is a dangerous disease that should not be underestimated. The early stages of this disease are often asymptomatic. Early detection of cancer is an important examination so that the disease does not develop into a serious and dangerous disease. This study detected the presence of cancer through five predictors. This study classified the diagnosis results based on five indicators namely radius, texture, perimeter, area, and smoothness. By using these five indicators, the detection was carried out through a classification mechanism using the boosting method. The result had obtained an accuracy of 93.67%. The accuracy was higher than other classification methods such as Bayesian Networks and multilayer Perceptron. Both of them only obtained an accuracy of 89.63%, and 92.79%, respectively. It showed that the ensemble method mechanism of boosting had proven to be more effective in classifying the presence or absence of breast cancer.

Keywords: Breast Cancer, Classification, Boosting, Bayesian Networks, Multilayer Perceptron

1 Introduction

The number of cancer patients all over the world significantly keeps increasing [1]. International Agency for Research for Cancer reports that World Health Organization (WHO) estimates around 18.1 million new cancer with the mortality rate of 9.6 million [2]. The cancer issue drives WHO predicts Cancer will be the most primary death cause in this century [3]. 1 out of 5 men and 1 out of 6 women suffer from cancer on earth. In addition to that, 1 out of 8 men and 1 out of 11 women die from having cancer. This information was obtained after researchers analyze data from 185 countries in the world through deeply focusing on 36 kinds of cancer.

In general, cancer is split into two classifications which are benign and malignant [4]. While cancer is benign, cancer is detected without having the ability to spread out and damage other tissue around. However, cancer is disruptive, spread out and damage other tissues around. Malignant cancer possesses a severe impact if it is not cured immediately. So early diagnosis of cancer is very crucial for the salvation of the cancer patient itself [5]. Through knowing the severity level of cancer the patient suffers from as early as possible [6], it can be determined what treatment should be done since cancer is able to cause death [7].

Early detection of breast cancer can be done through data mining [8]. It is a process for finding useful knowledge or information from large-scale data [9]. Methods in data mining have been widely implemented for a number of cases such as classification [10], association

MSCEIS 2019, October 12, Bandung, Indonesia

DOI 10.4108/eai.12-10-2019.2296544

rules [11], clustering [12], and forecasting [13]. Therefore, this study adopts data mining for the early detection of breast cancer.

Data mining consists of two stages of learning namely supervised and unsupervised [14]. This study classifies data into categories so that the learning process is supervised. The classification is divided into single and multiple classification models (ensemble method). The ensemble method can improve the accuracy of the single classification model by combining predictions made by several classifiers. One of the ensemble method algorithms is boosting. The method runs an iterative procedure to change training data adaptively with a focus on data that is difficult to classify. This procedure makes the performance of this method more optimal. Thus, this study compares the single and multiple classification models. The single classification model is represented by Bayesian networks and multilayer perceptron.

2 Literature study

The method in this study consisted of Bayesian networks, neural networks, and MLP. Both are explained in Sections 2.1, 2.2 and 2.3.

2.1 Bayesian networks

Bayesian network is a graphical model that encodes probabilistic relationships between interesting variables [15]. Bayesian networks can show the probability of a relationship between related and unrelated events. Bayesian network generalizations can represent and solve decisions under uncertainty called influence diagrams [16]. Bayesian networks can be used to calculate the probability of the presence of various symptoms of the disease [17]. In processing a Bayesian Network, it is built with Conditional Probability. It estimates the probability of an event B if an event A has occurred, denoted P (A). It calculates the estimation of a data set to enter a certain class based on the inference of existing data. The basic equation of the Bayes theorem is :

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \quad (1)$$

Bayesian networks can make probabilistic decision making (inference). It is to predict the value of a variable that cannot be known directly by using the values of other variables that are already known.

2.2 Neural networks

Neural networks are information processing systems that have characteristics similar to biological neural networks [18], which are neural networks in the human brain. The characteristics of neural networks are determined by several things, namely:

1. Network architecture
   Architecture is a form of the pattern of relationships between neurons. In neural networks, neurons are arranged in layers. The arrangement of neurons in layers and their relationships is called network architecture. Neural networks can be classified into two

types, single layer, and multilayer. In a single layer network, neurons can be grouped into two parts, namely input units (units of input) and output units (units of output). Whereas in multilayer networks, in addition to input and output units there are also hidden units (hidden units).

2. Learning
   The learning algorithm is a method used to determine the weight of the relationship [19]. The purpose of neural network training is to find the weights contained in each layer. There are two types of training in neural network systems, namely supervised and unsupervised learning. In a supervised learning process, neural networks are trained by providing data called training data or training vectors. Then given to the neural network so that the neural network can modify the weights to try to find similarities between the output results generated by the neural network with the desired output results. After the training process is complete, the neural network is then given an input value and will produce an output.
3. Activation Function
   The activation function is a function to produce output.
4. Neural Network Training
   In addition to minimizing errors in the output generated by the network, another goal of neural network training is to strike a balance between the ability to respond to input patterns used correctly in the training process [20].

2.3 MLP

MLP is one type of algorithm for neural networks. Learning this algorithm is done by backpropagation. Determination of the optimal weight will lead to the correct prediction results [21]. In MLP, the Sigmoid standard function is used wherein the amount of weighting from a number of inputs and biases are entered into the activation level through the transfer function to produce output, and units are arranged in a feed-forward topology layer called the Feed Forward Neural Network (FFNN).

When there is more than one hidden layer, the output of the hidden layer is entered into the next hidden layer and separate weights are used for addition to each subsequent layer. MLP consists of an input layer, one or more hidden layers, and an output layer [22]. Each node in MLP is a processing unit. Each node has several inputs and an output. Each node combines several input values, performs calculations, and generates output values (activations). In each node, there are two functions, namely a function to combine input and an activation function to calculate the output.

Backpropagation works through an iterative process using training data, comparing the predicted value of the network with each data contained in the training data. In each process, the weight of relations in the network is modified to minimize the value of Mean Squared Error (MSE) between the predicted value of the network with the actual value. The modification of the neural network relation is done in the reverse direction, from the output layer to the first layer of the hidden layer so that this algorithm is called backpropagation.

2.4 Boosting

The boosting method is one of the machine learning algorithms. It combines a number of weak classifiers to produce a strong classifier [23]. The formation process is done iteratively. At each iteration, the weak classifier is trained on data. If a classification failure occurs, the weight of the data will increase in the next iteration. It is intended that the classification process of next weak classifiers focuses on the data fails. The final stage combines the results of each iteration with its weight. In general, the steps of this method are:

1. Determine the initial weight of each data that is w [i] = 1 / n where n is the amount of data, and i = 1, 2, ..., n.
2. j is the number of iterations, then for m = 1, 2, ... M, At each iteration:
   a. Build a single tree based on data with a weight of w [i].
   b. Calculate the misclassification rate.
   c. Calculate the value of a [m], which is the evaluation parameter in iteration m
   d. Assign a new weight for each data. If the data is properly classified, then there is no change in weight. However, if data fails to be classified, the weight is updated. The new weight is the multiplication between the old weight and a [m].
3. The final prediction is a combination of predictions for each iteration. The class of data is based on the largest prediction value compared to other classes.

3 Research Methodology

The research methodology is explained in **Figure 1**. The data used came from the University of Wisconsin Hospitals. There are two types of variables in this study, the independent variable, and the dependent variable [24]. Meanwhile, the total data processed in the classification of breast cancer was 569 data with 5 independent variables and 1 dependent variable. Table 1 shows the variables used in the study and their statistic values.

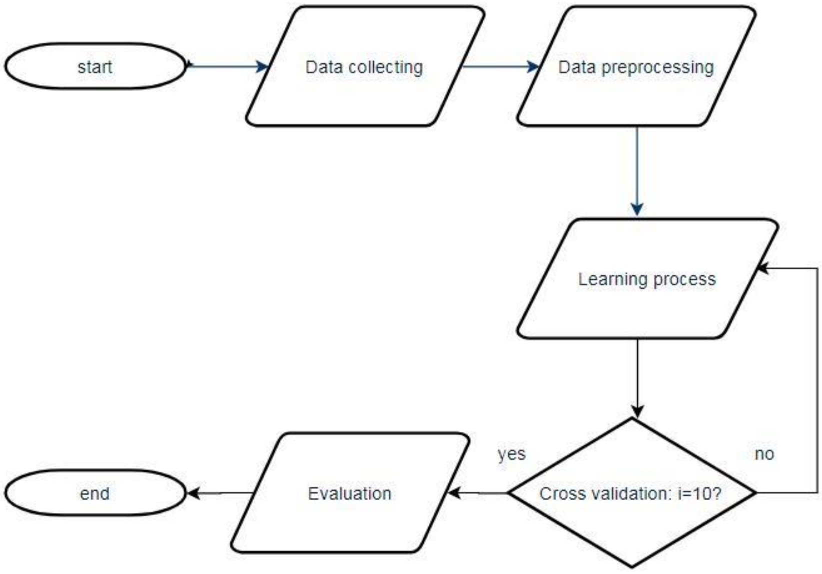

Fig. 1. Research methodology.

This study applied 10 fold cross-validation for the testing mechanism. The evaluation parameters were accuracy and root mean square error (RMSE). The learning rate in MLP in this study was 0.3 for momentum = 0.2. The base classifier of boosting based on decision stump with maximum iteration was 10. The batch size of the three methods in this study was 100.

Table 1. Summary of dataset features.

No.	Features	Max	Min	Mean	Stdev
1	Area	2501.00	143.50	654.28	351.92
2	Perimeter	188.50	43.79	91.91	24.29
3	Radius	28.11	6.98	14.12	3.52
4	Smoothness	0.16	0.05	0.10	0.01
5	Texture	39.28	9.71	19.31	4.29

4 Results and Discussion

The results of using the Bayesian Networks algorithm were 510 instances that are classified correctly and 59 instances that were classified incorrectly. The RMSE from Bayesian networks was 0.28. MLP in this study succeeded in 528 correct instances and 41 incorrect instances. The RMSE of MLP was 0.24. The smallest RMSE obtained by boosting was 0.23. on the other hand, boosting got the highest accuracy. It shows that boosting had the best performance, as shown in Table 2.

The confusion matrix of the three methods is shown in Table 3, 4, and 5. For the malignant class, the best performance in classifying it was by boosting, while the lowest was by Bayesian networks. As for the benign class, the best performance in classifying it was by boosting, while the lowest was by MLP.

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

Method	Accuracy (%)	RMSE
Bayesian networks	89.63	0.28
Multilayer perceptron	92.79	0.24
Boosting	93.67	0.23

Table 3. Confusion matrix of Bayesian networks.

	Classified as	
	Malignant	Benign
Malignant	168	44
Benign	15	342

Table 4. Confusion matrix of multilayer perceptron.

	Classified as	
	Malignant	Benign
Malignant	187	25
Benign	16	341

Table 5. Confusion matrix of boosting.

	Classified as	
	Malignant	Benign
Malignant	188	24
Benign	12	345

5 Conclusion

This study compared Bayesian networks, MLP, and boosting in classifying types of breast cancer. The testing using data as many as 510 instances based on 10 fold cross-validation. Based on research results, the highest performance was obtained by boosting while the lowest was by Bayesian networks.

References

[1] Holleczek, B., et al.: On-going improvement and persistent differences in the survival for patients with colon and rectum cancer across Europe 1999–2007–results from the EUROCARE-5 study. European Journal of Cancer, 51(15), 2158-2168 (2015)
[2] Bray, F., et al.: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians, 68(6), 394-424 (2018)
[3] Torre, L. A., et al.: Global cancer statistics, 2012. CA: a cancer journal for clinicians, 65(2), 87-108 (2015)
[4] Röhrich, M., et al.: Methylation-based classification of benign and malignant peripheral nerve sheath tumors. Acta neuropathologica, 131(6), 877-887 (2016)
[5] Li, Q. K., et al.: An integrated proteomic and glycoproteomic approach uncovers differences in glycosylation occupancy from benign and malignant epithelial ovarian tumors. Clinical proteomics, 14(1), 16 (2017)
[6] Shah, C.,et al.: The impact of early detection and intervention of breast cancer-related lymphedema: a systematic review. Cancer medicine, 5(6), 1154-1162 (2016)
[7] Ghoncheh, M., Pournamdar, Z., & Salehiniya, H.: Incidence and mortality and epidemiology of breast cancer in the world. Asian Pac J Cancer Prev, 17(S3), 43-46 (2016)
[8] Uttley, L., et al.: Building the evidence base of blood-based biomarkers for early detection of cancer: a rapid systematic mapping review. EBioMedicine, 10, 164-173 (2016)

[9] Jothi, N., & Husain, W.: Data mining in healthcare–a review. Procedia Computer Science, 72, 306-313 (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] Kaur, M., & Kang, S.: Market Basket Analysis: Identify the changing trends of market data using association rule mining. Procedia computer science, 85, 78-85 (2016)
[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] Martínez-Álvarez, F., Troncoso, A., Asencio-Cortés, G., & Riquelme, J.: A survey on data mining techniques applied to electricity-related time series forecasting. Energies, 8(11), 13162-13193 (2015)
[14] Yulita, I. N., Fanany, M. I., & Arymurthy, A. M.: Fuzzy Clustering and Bidirectional Long Short-Term Memory for Sleep Stages Classification. In 2017 International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT) (pp. 11-16). IEEE. (2017)
[15] Zhang, J., Cormode, G., Procopiuc, C. M., Srivastava, D., & Xiao, X.: Privbayes: Private data release via bayesian networks. ACM Transactions on Database Systems (TODS), 42(4), 25 (2017)
[16] Cai, B., Huang, L., & Xie, M.: Bayesian networks in fault diagnosis. IEEE Transactions on Industrial Informatics, 13(5), 2227-2240 (2017)
[17] Xu, S., Thompson, W., Ancoli-Israel, S., Liu, L., Palmer, B., & Natarajan, L.: Cognition, quality-of-life, and symptom clusters in breast cancer: Using Bayesian networks to elucidate complex relationships. Psycho-oncology, 27(3), 802-809 (2018)
[18] Da Silva, I. N., et al.: Artificial neural networks. Cham: Springer International Publishing. (2017)
[19] Tkáč, M., & Verner, R.: Artificial neural networks in business: Two decades of research. Applied Soft Computing, 38, 788-804 (2016)
[20] Deb, C., Eang, L. S., Yang, J., & Santamouris, M.: Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks. Energy and Buildings, 121, 284-297 (2016)
[21] 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)
[22] Ramchoun, H., Idrissi, M. A. J., Ghanou, Y., & Ettaouil, M.: Multilayer Perceptron: Architecture Optimization and Training. IJIMAI, 4(1), 26-30 (2016)
[23] Yao, T., Pan, Y., Li, Y., Qiu, Z., & Mei, T.: Boosting image captioning with attributes. In Proceedings of the IEEE International Conference on Computer Vision (pp. 4894-4902) (2017)
[24] Bennett, K. P., & Mangasarian, O. L.: Robust linear programming discrimination of two linearly inseparable sets. Optimization methods and software, 1(1), 23-34 (1992)