

Breast Cancer Diagnosis using Artificial Neural Networks with Extreme Learning Techniques

Chandra Prasetyo Utomo

YARSI E-Health Research Center
Faculty of Information Technology
YARSI University,
Jakarta, Indonesia

Aan Kardiana

YARSI E-Health Research Center
Faculty of Information Technology
YARSI University,
Jakarta, Indonesia

Rika Yuliwulandari

YARSI Genetics Research Center
Faculty of Medicine
YARSI University, Jakarta,
Indonesia

Abstract—Breast cancer is the second cause of dead among women. Early detection followed by appropriate cancer treatment can reduce the deadly risk. Medical professionals can make mistakes while identifying a disease. The help of technology such as data mining and machine learning can substantially improve the diagnosis accuracy. Artificial Neural Networks (ANN) has been widely used in intelligent breast cancer diagnosis. However, the standard Gradient-Based Back Propagation Artificial Neural Networks (BP ANN) has some limitations. There are parameters to be set in the beginning, long time for training process, and possibility to be trapped in local minima. In this research, we implemented ANN with extreme learning techniques for diagnosing breast cancer based on Breast Cancer Wisconsin Dataset. Results showed that Extreme Learning Machine Neural Networks (ELM ANN) has better generalization classifier model than BP ANN. The development of this technique is promising as intelligent component in medical decision support systems.

Keywords—breast cancer; artificial neural networks; extreme learning machine; medical decision support systems

I. INTRODUCTION

The out of control development of cells in an organ is called tumors that can be cancerous. There are two kinds of tumors, benign and malignant. Benign or non-cancerous tumors are not spreading and are not life intimidating. In the other hand, malignant or cancerous tumor are expanding and life threatening [1]. Malignant breast cancer is defined when the growing cells are in the breast tissue. Breast cancer is the second overall cause of mortality among women and the first cause of dead among them between 40 and 55 ages [2].

Regular breast cancer diagnosis followed by appropriate cancer treatment can reduce the unwilling risk. It is suggested to do tumor evaluation test every 4-6 weeks. Based on that reason, benign and malignant detection based on classification features become very important [3].

Careful diagnosis in early detection has been proven to lessen the dead rate because of breast cancer [4]. Depend on the expertise, mistakes can be made by medical professionals while identifying a disease. With the help of technology such as data mining and machine learning, diagnosis can be more accurate (91.1%) when related to a diagnosis made by an experienced doctor (79.9%) [5].

ANN is one of the best artificial intelligence techniques for common data mining tasks, such classification and regression problems. A lot of research showed that ANN delivered good accuracy in breast cancer diagnosis. However, this method has several limitations. First, ANN has some parameters to be tuned in the beginning of training process such as number of hidden layer and hidden nodes, learning rates, and activation function. Second, it takes long time for training process due to complex architecture and parameters update process in each iteration that need expensive computational cost. Third, it can be trapped to local minima so that the optimal performance cannot be guaranteed. Numerous efforts had been attempted to get the solutions of neural networks limitations. Huang and Babri [6] proved that Single Hidden Layer Neural Networks (SFLN) with tree steps extreme learning process that called ELM can solve that problems.

In this paper, we revealed the implementation of artificial neural networks with extreme learning techniques in breast cancer diagnosis. The dataset used for experiments was Breast Cancer Wisconsin Dataset that was obtained from the University of Wisconsin Hospital, Madison from Dr. William H. Wolberg [7]. We compared the performance of ELM with conventional BP ANN with gradient descent based learning algorithms. Sensitivity, specificity, and accuracy were used as performance measurements. Results showed that ELM ANN generally produced better result than BP ANN.

The rest of this paper is organized as the following. Section 2 is dedicated as literature review. In this section, brief review of previous works in breast cancer diagnosis are presented. In Section 3, the concept, mathematical model, and training process of extreme learning machine are explained. In Section 4, experiments, results, and analysis are provided. Finally, conclusions and future works are given in Section 5.

II. LITREATURE REVIEW

The uses of classification systems in medical diagnosis, including breast cancer diagnosis, are growing rapidly. Evaluation and decision making process from expert medical diagnosis is key important factor. However, intelligent classification algorithm may help doctor especially in minimizing error from unexperienced practitioners [3].

Several techniques have been deployed to predict and recognize meaningful pattern for breast cancer diagnosis. Ryua

[8] developed data classification method, called isotonic separation. The performances were compared against support vector machines, learning vector quantization, decision tree induction, and other methods based on two-breast cancer data set, sufficient and insufficient data. The experiment results demonstrated that isotonic separation was a practical tool for classification in the medical domain.

Hybrid machine learning method was applied by Sahar [9] in diagnosing breast cancer. The method hybridized a fuzzy-artificial immune system with k-nearest neighbour algorithm. The hybrid method delivered good accuracy in Wisconsin Breast Cancer Dataset (WBCD). They believe it can also be tested in other breast cancer diagnosis problems.

Comprehensive view of automated diagnostic systems implementation for breast cancer detection was provided by Ubeyli [10]. It compared the performances of multilayer perceptron neural network (MLPNN), combined neural network (CNN), probabilistic neural network (PNN), recurrent neural network (RNN) and support vector machine (SVM). The aim of that works was to be a guide for a reader who wants to develop this kind of systems.

Numerous combinations and hybrid systems used neural networks as a component. However, since almost all of the employed neural networks are conventional gradient descent BP ANN, the novel or hybrid method still suffered the neural networks drawbacks that were mentioned in the previous section.

III. EXTREME LEARNING MACHINE

Huge efforts had been attempted to solve the weaknesses of BP ANN. Huang and Babri [6] demonstrated that single hidden layer feedforward neural networks (SLFN) with at most m hidden nodes was capable to estimate function for m different vectors in training dataset.

Given m instances in $\mathbf{D} = \{(\mathbf{x}^{(k)}, \mathbf{t}^{(k)}) \mid \mathbf{x}^{(k)} \in \mathbf{R}^n, \mathbf{t}^{(k)} \in \mathbf{R}^p, k = 1, \dots, m\}$ as training dataset where $\mathbf{x}^{(k)} = [x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}]^T$ as features and $\mathbf{t}^{(k)} = [t_1^{(k)}, t_2^{(k)}, \dots, t_p^{(k)}]^T$ as target. A SLFN with M number of hidden nodes, activation function $g(x)$ in hidden nodes, and linear activation function in output nodes is mathematically wrote as:

$$\sum_{i=1}^M \beta_i g_i(\mathbf{x}^{(k)}) = \sum_{i=1}^M \beta_i g(\mathbf{w}_i \cdot \mathbf{x}^{(k)} + b_i) = \mathbf{o}^{(k)}, \quad k = 1, \dots, m \quad (1)$$

where

$\mathbf{w}_i \in \mathbf{R}^n$ is the weights between the input nodes and the i -th hidden node

$$\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T, \quad (2)$$

$\beta_i \in \mathbf{R}^p$ is the weights between the i -th hidden node and the output nodes

$$\boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{ip}]^T, \quad (3)$$

$\mathbf{w}_i \cdot \mathbf{x}^{(k)}$ is the inner product of \mathbf{w}_i and $\mathbf{x}^{(k)}$,

b_i is the bias of the i -th hidden node,

$\mathbf{o}^{(k)} \in \mathbf{R}^p$ is the output of neural network for k -th vector.

SLFN can approximate m vectors means that there exist $\mathbf{w}_i, \boldsymbol{\beta}_i$, and b_i , such that:

$$\|\mathbf{o}^{(k)} - \mathbf{t}^{(k)}\| = 0 \quad (4)$$

$$\sum_{i=1}^M \beta_i g(\mathbf{w}_i \cdot \mathbf{x}^{(k)} + b_i) = \mathbf{t}^{(k)}, \quad k = 1, \dots, m \quad (5)$$

Equation (5) can be written as:

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T}, \quad (6)$$

where

$\mathbf{H} \in \mathbf{R}^{m \times M}$ is the hidden layer output matrix of the neural networks.

$$\mathbf{H} = \begin{bmatrix} g(w_1 \cdot x^{(1)} + b_1) & \dots & g(w_M \cdot x^{(1)} + b_M) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot x^{(m)} + b_1) & \dots & g(w_M \cdot x^{(m)} + b_M) \end{bmatrix} \quad (7)$$

$\boldsymbol{\beta} \in \mathbf{R}^{M \times p}$ is the weights between hidden and output layers

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1^T \\ \vdots \\ \boldsymbol{\beta}_M^T \end{bmatrix}, \quad (8)$$

$\mathbf{T} \in \mathbf{R}^{m \times p}$ is the target values of m vectors in training dataset

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}^{(1)T} \\ \vdots \\ \mathbf{t}^{(m)T} \end{bmatrix}, \quad (9)$$

In the traditional gradient descent based learning algorithm, weights \mathbf{w}_i which was connecting the input layer and hidden layer and biases b_i in the hidden nodes were needed to be initialized and tuned in every iteration. This was the main factor which often made training process of neural became time consuming and the trained model may not reach global minima.

Huang [11] proposed minimum norm least-squares solution of SLFN which didn't need to tune those parameters. Training SLFN with fixed input weights \mathbf{w}_i and the hidden layer biases b_i was similar to find a least square solution $\hat{\beta}$ of the linear system $\mathbf{H}\beta = \mathbf{T}$:

$$\begin{aligned} & \|\mathbf{H}(\mathbf{w}_1, \dots, \mathbf{w}_M, b_1, \dots, b_M,)\hat{\beta} - \mathbf{T}\| = \\ & \min_{\beta} \|\mathbf{H}(\mathbf{w}_1, \dots, \mathbf{w}_M, b_1, \dots, b_M,)\beta - \mathbf{T}\|. \end{aligned} \quad (10)$$

The smallest norm least squares solution of that linear system was

$$\hat{\beta} = \mathbf{H}^{\dagger}\mathbf{T} \quad (11)$$

where \mathbf{H}^{\dagger} was the *Moore-Penrose generalized inverse* of matrix \mathbf{H} . This solution had three important properties which were minimum training error, smallest norm of weights, and unique solution which is $\hat{\beta} = \mathbf{H}^{\dagger}\mathbf{T}$.

The above minimum norm least-square solution for SLFN was called extreme learning machine (ELM). Given m instances in training dataset $\mathbf{D} = \{(\mathbf{x}^{(k)}, \mathbf{t}^{(k)}) \mid \mathbf{x}^{(k)} \in \mathbf{R}^n, \mathbf{t}^{(k)} \in \mathbf{R}^p, k = 1, \dots, m\}$, activation function $g(x)$, and number of hidden node M . The training process of ELM is the the following:

1. Randomly set input-hidden layer weights \mathbf{w}_i and bias $b_i, i = 1, \dots, M$.
2. Compute the matrix of hidden layer output \mathbf{H}
3. Compute the hidden-output layer weights $\hat{\beta}$ for $\hat{\beta} = \mathbf{H}^{\dagger}\mathbf{T}$ where $\mathbf{T} = [\mathbf{t}^{(1)}, \dots, \mathbf{t}^{(m)}]$.

Based on that definition, there are three main differences between BP ANN and ELM ANN. First, BP ANN needs to tuning several parameters, such as number of hidden nodes, learning rates, momentum, and termination criteria. On the other hand, ELM ANN is a simple tuning free algorithm. The only one to be defined is number of hidden nodes. Second, BP ANN works only for differentiable activation functions in hidden and output nodes while ELM ANN can use both differentiable and undifferentiable activation functions.

Finally, BP ANN get trained model which has minimum training error so that there is a possibility to finish in local minima. On the other hand, ELM ANN get trained model which has minimum training error and smallest norm of weight so that it can produce better generalization model and reach global minima [12].

IV. EXPERIMENTS, RESULTS, AND ANALYSIS

This section discussed about experimental design, generated results, and analytical process in order to get valid conclusion.

A. Experiments

The experiments consisted of three main steps, which were data gathering, data preprocessing, and performance evaluating. The dataset used in this experiment was Breast

Cancer Wisconsin Dataset obtained from the University of Wisconsin Hospital, Madison from Dr. William H. Wolberg [7]. The data has 699 instances with 10 attributes plus the class attributes. The class distribution are 65.5% (458 instances) for benign and 34.5% (241 instances) for malignant. The attribute information can be seen in TABLE I.

TABLE I. ATTRIBUTE INFORMATION

#	Attributes	Domain
1	Sample Code Number	id number
2	Clump Thickness	1 – 10
3	Uniformity of Cell Size	1 – 10
4	Uniformity of Cell Shape	1 – 10
5	Marginal Adhesion	1 – 10
6	Single Epithelial Cell Size	1 – 10
7	Bare Nuclei	1 – 10
8	Bland Chromatin	1 – 10
9	Normal Nucleoli	1 – 10
10	Mitoses	1 – 10
11	Class	2: benign; 4: malignant

In the second step, the raw dataset was preprocessed to produce well-from data that suitable for training and testing process. The first attribute, sample code number, was removed because it was not relevant to the diagnosis. The next nine attributes were normalized into [-1, 1] and used as predictor. The last attribute was transformed to 0 (benign) and 1 (malignant) such that it can be properly fitted to the standard BP ANN and ELM ANN implementation.

The method in this experiment was k-fold crossvalidation with $k = 5$. This means, the data were randomly divided into 5 partitions. There were 5 experiments. In the each experiment, a partition was used as testing data and the rest partitions were treated as training data.

The standard performance measurement for classification problem was accuracy. However, since the class distribution was not balanced, it was important to use specificity and sensitivity as supplementary measurements. In addition, to minimize the effect of random generated weights in BP ANN and ELM ANN, each experiment was run three times and the average results were noted.

$$accuracy = \frac{True\ Positive + True\ Negative}{All\ Testing\ Data} \quad (13)$$

$$sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (14)$$

$$specificity = \frac{True\ Negative}{False\ Positive + True\ Negative} \quad (15)$$

B. Results and Analysis

With 5-fold crossvalidation method and each experiment were run three times, there will be 15 experiments in total. The whole steps had been done in computer with Intel® Core™ i3, 4096MB RAM, and Windows 7 OS. The results of ELM are given in TABEL II.

TABLE II. ELM PERFORMANCES

Experiments	Running	Sensitivity	Specificity	Accuracy
#1	1	0,902	0,965	0,942
	2	0,882	0,977	0,942
	3	0,882	0,965	0,934
#2	1	0,975	0,958	0,963
	2	0,950	0,979	0,971
	3	1,000	0,969	0,978
#3	1	0,909	0,988	0,956
	2	0,964	0,963	0,963
	3	0,964	0,975	0,971
#4	1	0,957	1,000	0,985
	2	0,957	0,989	0,978
	3	0,957	1,000	0,985
#5	1	0,957	0,967	0,964
	2	1,000	0,956	0,971
	3	0,957	0,956	0,956

In order to compare the ELM ANN performances, training and testing with BP ANN were conducted with identical experimental design. The performances of BP ANN can be seen in TABLE III.

TABLE III. BP ANN PERFORMANCES

Experiments	Running	Sensitivity	Specificity	Accuracy
#1	1	0,875	0,975	0,934
	2	0,750	0,988	0,891
	3	0,804	0,963	0,898
#2	1	0,855	0,987	0,927
	2	0,871	0,973	0,927
	3	0,903	0,973	0,941
#3	1	0,825	0,949	0,897
	2	0,860	0,987	0,934
	3	0,825	0,975	0,912
#4	1	0,790	0,975	0,898
	2	0,825	0,975	0,912
	3	0,807	0,975	0,905
#5	1	0,892	1,000	0,949
	2	0,892	1,000	0,949
	3	0,877	1,000	0,942

To have clear view between the performances of ELM ANN and BP ANN, the results were transformed to graphical charts. In each performance measurement, the average values were computed in each experiment. Fig 1 shows the comparison of average sensitivity rates between BP ANN and ELM ANN. The comparison of specificity rates were given in Fig 2 while accuracy rates can be seen in Fig 3.

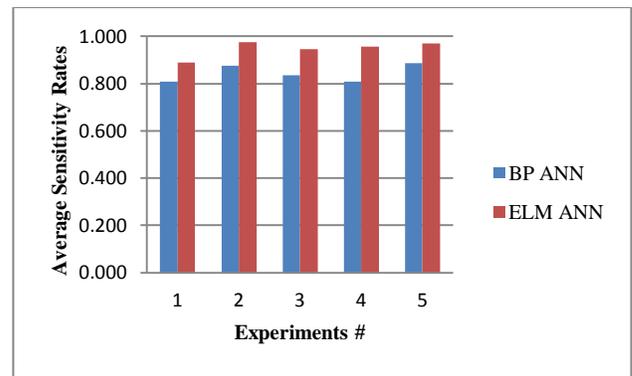


Fig. 1. Average sensitivity rates of BP ANN dan ELM

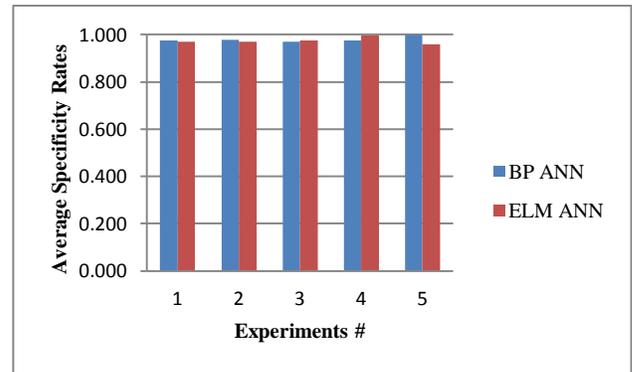


Fig. 2. Average specificity rates of BP ANN dan ELM

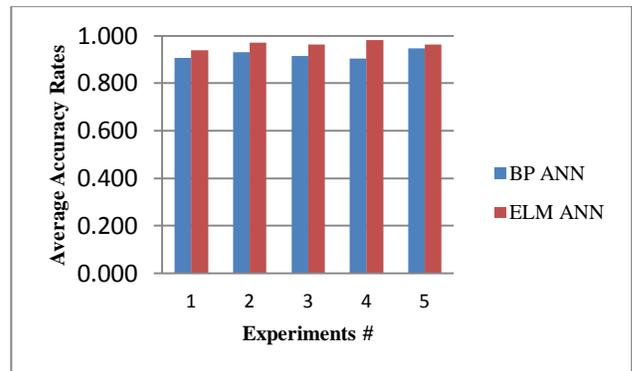


Fig. 3. Average accuracy rates of BP ANN dan ELM

Based on Fig 1 and Fig 3, we can see that ELM ANN was superior compared to BP ANN. ELM ANN has better performances in term of sensitivity and accuracy in all experiments.

However, in term of specificity, BP ANN has better performance in three experiments which were experiment #1, #2, and #5.

To get general conclusion, overall comparison need to be computed. In each performance measurement, commulative average rates were matched. Fig 4 displays the whole sensitivity, specificity, and accuracy average rates between BP ANN and ELM ANN. Result showed that, generally ELM ANN were better than BP ANN.

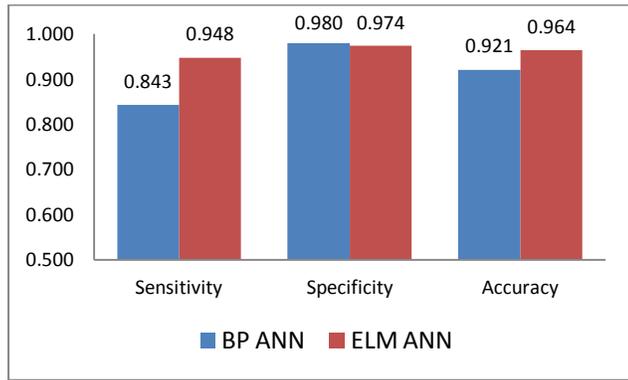


Fig. 4. Overall average rates between BP ANN dan ELM

V. CONCLUSION AND FUTURE WORKS

The performances of ELM ANN were generally better than BP ANN in breast cancer diagnosis. Although the specificity rate was slightly lower than BP ANN, it can be clearly seen that ELM ANN remarkably improved the sensitivity and accuracy rates. Based on these results, we can conclude that ELM ANN has better generalization model than BP ANN in diagnosing breast cancer based on Breast Cancer Wisconsin Dataset.

There are some necessary works to be done in near future. First, it is important to communicate the model with domain expert. The hybrid of ELM ANN with Decision Tree or any other technique that can produce meaningful knowledge representation will be promising. Second, to make intelligent diagnosis tool that can be used by end user, it is necessary to develop interactive user interface.

The development of interfaces in mobile, desktop, or web application may be useful. Third, there are new cases added regularly in the hospital. Developing intelligent diagnosis systems that can not only learn from available data in repositories but also from newly available data will be required.

REFERENCES

- [1] T. S. Subashini, V. Ramalingam, and S. Palanivel, "Breast mass classification based on cytological patterns using RBFNN and SVM," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5284–5290, Apr. 2009.
- [2] TheAmericanCancerSociety, "What is Breast Cancer."
- [3] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3240–3247, Mar. 2009.
- [4] D. West, P. Mangiameli, R. Rampal, and V. West, "Ensemble strategies for a medical diagnostic decision support system: A breast cancer diagnosis application," *Eur. J. Oper. Res.*, vol. 162, no. 2, pp. 532–551, Apr. 2005.
- [5] R. W. Brause, "Medical Analysis and Diagnosis by Neural Networks," in *Proceeding ISMDA '01 Proceedings of the Second International Symposium on Medical Data Analysis*, 2001, pp. 1–13.
- [6] G. B. Huang and H. A. Babri, "Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions," *IEEE Trans. Neural Netw.*, vol. 9, no. 1, pp. 224–9, Jan. 1998.
- [7] W. H. Wolberg and O. L. Mangasarian, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology," in *Proceedings of the National Academy of Sciences, U.S.A., Volume 87, December 1990*, pp. 9193–9196.
- [8] Y. U. Ryu, R. Chandrasekaran, and V. S. Jacob, "Breast cancer prediction using the isotonic separation technique," *Eur. J. Oper. Res.*, vol. 181, no. 2, pp. 842–854, Sep. 2007.
- [9] S. Sahan, K. Polat, H. Kodaz, and S. Güneş, "A new hybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis," *Comput. Biol. Med.*, vol. 37, no. 3, pp. 415–23, Mar. 2007.
- [10] E. D. Übeyli, "Implementing automated diagnostic systems for breast cancer detection," *Expert Syst. Appl.*, vol. 33, no. 4, pp. 1054–1062, Nov. 2007.
- [11] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, Dec. 2006.
- [12] C. P. Utomo, "The Hybrid of Classification Tree and Extreme Learning Machine for Permeability Prediction in Oil Reservoir," *Int. J. Comput. Sci. Issues*, vol. 10, no. 1, pp. 52–60, 2013.