UJM 14 (1) 2025
UNNES Journal of Mathematics
http://journal.unnes.ac.id/sju/index.php/ujm

Breast Cancer Prediction Using Artificial Neural Network

Birru Asia Rayani*, Faiza Al Laily Nasron, Neli Septiana Putri, Novita Sari Parapat, Virgania Sari

Department of Mathematics, Faculty of Mathematics and Natural Science,
Universitas Negeri Semarang, Semarang, 50229, Indonesia

Article Info

Article History:
Received …….
Accepted …..
Published …..

Keywords:
breast cancer, Artificial Neural Network, backpropagation algorithm

Abstract

Cancer is one of the medical conditions that causes many deaths in different parts of the world. Based on information obtained from GLOBOCAN, the International Agency for Research on Cancer (IARC) in 2022, there were at least 19.976.499 individuals diagnosed with cancer, and the disease caused death in 9.743.832 people. The detection of breast cancer malignancy relies on the prognosis process, requiring forecasting and automated detection to mitigate diagnostic errors. This facilitates swift and comprehensive analysis of medical data. The study employs the Neural Network, specifically the Artificial Neural Network model, implemented using python and the backpropagation algorithm. Utilizing data from the WDBC Database at the University of Wisconsin, the research achieves a 96,49% accuracy in breast cancer prediction, with an area under curve (AUC) value of 0,992, demonstrating the model's overall efficacy in accurate predictions.

How to cite:

Rayani, B. A., Nasron, F. A. L., Putri, N. S., & Parapat, N. S. 2025. Breast Cancer Prediction Using Artificial Neural Network. *UNNES Journal of Mathematics*. 14 (1): 10-15

*Correspondence Address
E-mail: birruasiarayani@students.unnes.ac.id

p-ISSN 2252-6943
e-ISSN 2460-5859

1. Introduction

Cancer is one of the medical conditions that causes many deaths in different parts of the world (Saini *et al*., 2020). Based on information obtained from GLOBOCAN, the International Agency for Research on Cancer (IARC) in 2022, there were at least 19.976.499 individuals diagnosed with cancer, and the disease caused death in 9.743.832 people. Most individuals suffering from cancer get diagnosed after facing a number of complications. When cancer is detected, it has often reached a more advanced stage of development in the body (Crosby *et al.,* 2022). From a medical perspective, diseases that reach a higher stage tend to be more difficult to manage than those detected at an earlier stage (Binabar *et al*., 2017).

The number of new cases of breast cancer reached 68.858 cases (16,6%) of the total 396.914 new cases of cancer in Indonesia (Rizaty, 2021). Although the causes of breast cancer are still not fully revealed, research has identified several factors that may increase the risk in certain individuals (Obeagu & Obeagu, 2024). Some of these include a family history of similar diseases, increasing age, lack of pregnancy experience or first pregnancy after the age of 30, as well as longer menstrual periods (with earlier first menstruation or later menopause). In addition, hormonal factors have also been identified as one of the elements that may contribute to breast cancer risk (Nugraha *et al*., 2024).

In general, the degree of malignancy of breast cancer is usually detected through the prognosis process. Currently, the determination of breast cancer prognosis factors can be done by molecular biology examination considering that quite a number of patients diagnosed with early stage breast cancer actually show a picture of metastasis at diagnosis (Setiawan, 2023). With the advancement of information technology, especially in the field of artificial intelligence, machine learning techniques have been introduced to improve automated detection capabilities. With the help of this system, there is potential to avoid misdiagnosis that may be made by medical experts, and medical data can be analyzed quickly and in more detail (Wahyuni, 2016).

Machine learning can be applied significantly in the healthcare field, including for diagnosing diseases such as breast cancer, heart disease, and diabetes (Singh *et al.,* 2021). Several previous studies have been developed for breast cancer prediction with machine learning. In a study conducted by Nurelasari (2018), a model for Neural Network was created using a dataset consisting of 9 randomly generated breast cancer attributes, as well as a class that indicates whether the tumor is benign or malignant. The findings of the study showed that optimizing parameters with genetic algorithms successfully improved the predictive ability of breast cancer. The resulting Neural Network model was able to provide higher accuracy than the Neural Network without optimization. The increase recorded in the study was seen from the initial accuracy of the Neural Network algorithm model of 95,42%. However, after optimization using a genetic algorithm, the accuracy of the Neural Network algorithm increased to 96,85%. Therefore, it can be concluded that there was an increase of 1,43% in accuracy.

To evaluate performance, a ROC curve was used which resulted in an AUC (Area Under Curve) value of 0,984 for the Neural Network algorithm model, while the genetic algorithm-based Neural Network algorithm resulted in a value of 0,993, indicating excellent classification. The difference between the two values is 0,009. Thus, it can be concluded that the application of genetic algorithm optimization techniques can improve the accuracy of the Neural Network algorithm.

In addition to genetic algorithms, a classification technique that can be applied in breast cancer prediction is Artificial Neural Network (ANN). ANN is a process of transforming information that has characteristics resembling a neural network. ANN was created as a generalization of the mathematical model of human understanding (Ridho *et al*., 2023). Like the human brain, neural networks also consist of several neurons, and there is a relationship between these neurons. In neural networks this relationship is known as weight. The information is stored at a certain value on the weight. It is then processed by a propagation function that will sum up the values of all future weights. The result of this summation is then compared with information called input sent to neurons with certain arrival weights.

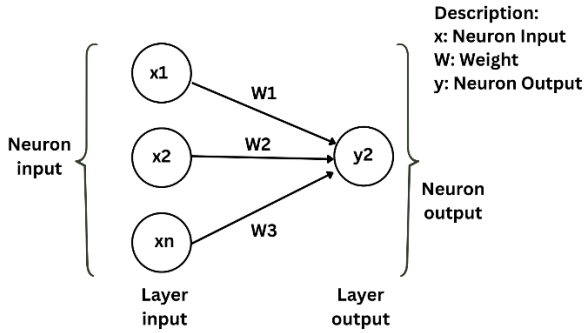

Figure 1 ANN structure

In ANN, the backpropagation algorithm is a supervised learning algorithm that uses the perceptron method to adjust the weights on the hidden layer neurons, based on the resulting output error. Backpropagation is one of the algorithms that is often used in solving complex problems (Annisa, 2023). In ANN the output value is determined by the activation function, which must be continuous, derivative, and have a sigmoid shape (Nasien *et al*., 2022). The ANN approach was chosen because it mimics the way the human nervous system works, enabling identification, prediction, and pattern recognition with a high degree of accuracy.

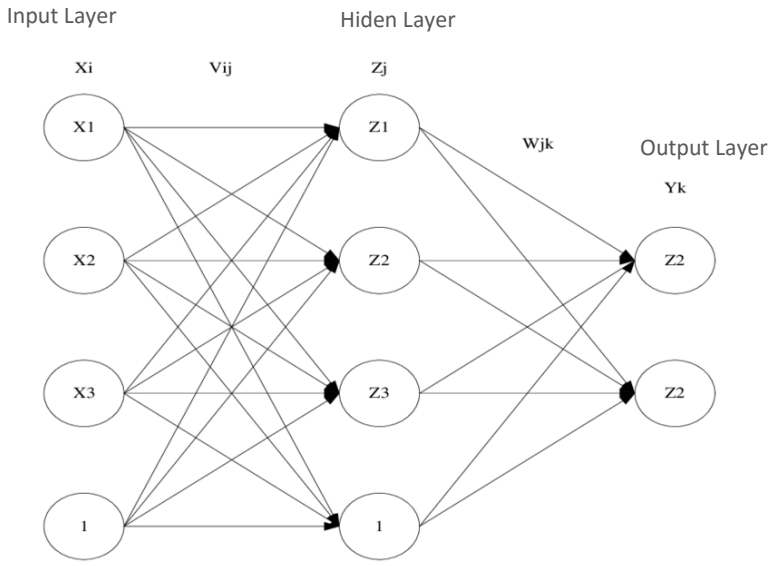

Figure 2 Backpropagation algorithm architecture
(Slamet *et al.,* 2020)

The main objective of this research is to predict or diagnose early breast cancer by applying Artificial Neural Network classification algorithms in the training, learning, and testing processes. This research takes different parameters from Nurelasari's (2018) research with a focus on Wisconsin Diagnostic Breast Cancer (WDBC) data which includes 569 patient clinical cases. In Nurelasari's (2018) research, a Neural Network model was used by optimizing parameters using a genetic algorithm. Meanwhile, this research will use the Artificial Neural Network model by applying the backpropagation algorithm to learn data patterns. It is possible to detect breast cancer early, so that intervention can be carried out more quickly to reduce mortality due to factors other than the nature of the cancer itself. Evaluation of the success of breast cancer diagnosis is done through two types of testing. First, using confusion matrix, and second, through reliability testing using ROC curves. Both methods are used to measure the success rate in predicting and diagnosing breast cancer at an early stage.

2. Method

2.1. Wisconsin Diagnostic Breast Cancer (WDBC) Data Set

The data used in this study was selected from the University of Wisconsin WDBC database (https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic). Each entity in the dataset consists of attributes corresponding to features extracted from digitized cell nucleus images in each tissue. The class associated with each entity provides information about the diagnosis of the tissue, which can be benign or malignant. Each entity has 30 variables assigned to it to represent cell samples. The attributes applied to each cell sample are used to describe instances, where each instance can be classified as a benign or malignant tumor. In the analyzed dataset, there were 212 cases of malignant tumors (37.2%) and 357 cases of benign tumors (62.7%).

2.2. Designing ANN Structure and Configuration

Initially, the data used in the Neural Network is normalized into an explicitly unspecified range, where the data range will be adjusted so that it has a mean of zero and a standard deviation of one. The neural network is implemented using the Python programming language to obtain predictive results for breast cancer classification. The Artificial Neural Network (ANN) model is created by applying the backpropagation algorithm to learn patterns from the data and adjust the weights based on the resulting error. The structure consists of an input layer with 26 neurons and two hidden layers with 16 neurons each in the first and second hidden layers, and an output layer consisting of only one neuron.

2.3. ANN Training, Testing, and Validation

During the training, testing, and validation stages of the artificial neural network, the entire dataset is divided into two parts, the training set which includes 80% of the total data, and the validation set which uses the remaining 20% of the data.

2.4. Desicion Analysis

As the last step, the output of the artificial neural network is generated, and upon exiting the network, it provides a classification result with a value of 0 for benign tumors or 1 for malignant tumors.

3. Results and Discussions

Based on the previous information, this research uses the WDBC dataset with a diagnosis target column consisting of two classes, namely benign and malignant. Before entering into ANN modeling, the data is first cleaned (preprocessing). In this data preprocessing stage, there are several phases carried out, including data cleaning, data transformation, and feature selection. From the feature selection process, 26 features are obtained that are considered important in this study. Experiments conducted using the Artificial Neural Network (ANN) algorithm on the dataset set as training data resulted in the model architecture as shown in Figure 3.

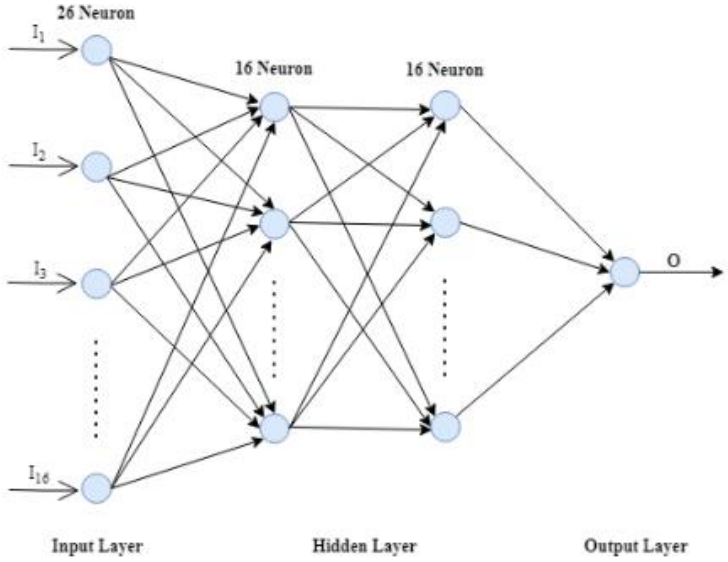

Figure 3 ANN model

In Figure 3, it can be seen that there are 26 input neurons used, 1 output neuron generated, and there are 16 neurons in each hidden layer. After the data is trained using ANN, the results are obtained in the form of a confusion matrix in Figure 4. The result of testing with confusion matrix is a model evaluation using a matrix consisting of true positive (correct prediction results for positive classes), true negative (correct prediction results for negative classes), false positive (negative data prediction results that are considered positive), and false negative (positive data prediction results that are considered negative).

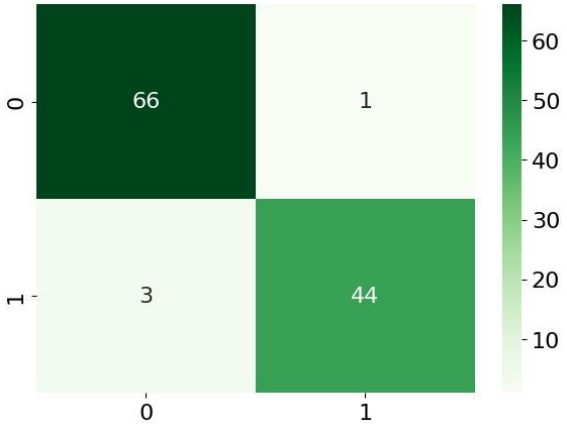

Figure 4 Confusion Matrix

Figure 4 can then be written back into Table 1 as follows.

Table 1 Confusion Matrix interpretation

True Positive (TP)	44
False Positive (FP)	1
True Negative (TN)	66
False Negative (FN)	3

After obtaining the confusion matrix from the training results using Artificial Neural Network (ANN) with the previously mentioned configuration, further analysis is carried out by creating a classification report. Classification report is a summary that provides more detailed information about the performance of the classification model. In the classification report, various evaluation metrics such as precision, recall, F1-score, and support (number of samples in each class) are presented for each class identified by the model. The classification report obtained is presented in Table 2 as follows.

Table 2 Classification Report

	Precission	Recall	F1-score	Support
0	0.96	0.99	0.97	67
1	0.98	0.94	0.96	47
Accuracy			0.96	114
Macro avg	0.97	0.96	0.96	114
Weighted avg	0.97	0.96	0.96	114

Based on Table 2, the accuracy value of the model created is 96,4912%, indicating that the model can predict correctly about 96,49% of the overall training data. Compared to Nurelasari (2018) research that developed a Neural Network model, the ANN model developed in this study has a higher accuracy value of 1,07%. In addition, the model also produces precission, recall, F1-score values above 95%. This shows that the model made has good quality in classifying data. In this study, a loss value of 0,1027 was also obtained, which means that the model is quite good at adjusting to the training data and making the right predictions.

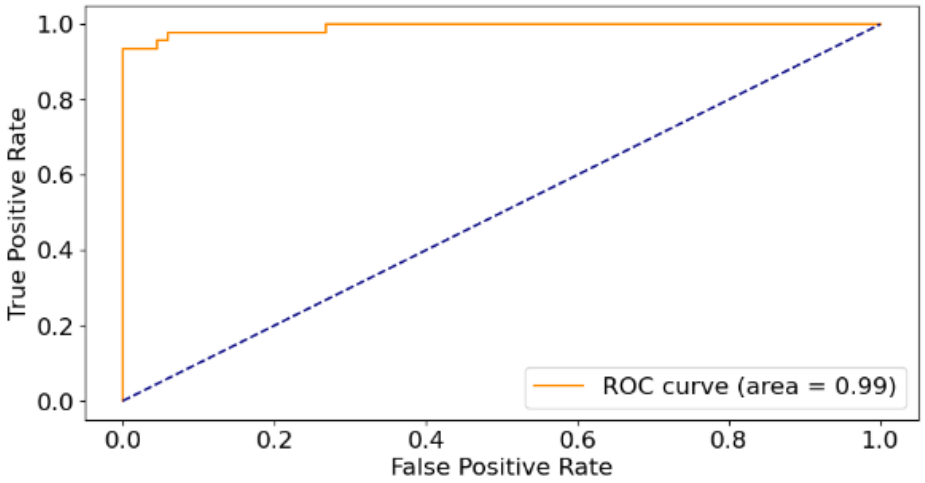

Figure 5 ROC curve

Based on Figure 5, it can be seen that the model in the study has an AUC value of 0,992. Compared to research conducted by Nurelasari (2018), this AUC value is 0,008 higher. This indicates that the area under the ROC curve or the area describing the comparison between the true positive rate and false positive rate at various thresholds reaches a very high value, approaching the perfect area (1,0). This indicates that the model has a very good ability to distinguish between positive and negative classes, and in general the model can be considered to perform very well in the given classification task.

4. Conclusions

From the research conducted, it resulted in a prediction accuracy of 96.49% in breast cancer. For evaluation using the ROC curve to produce the AUC (Area Under Curve) value for the Artificial Neural Network model produces a value of 0.992. So, it can be concluded that the model is generally able to predict well.

The model that has been created can be developed or applied into an application that will assist health practitioners in diagnosing breast cancer. This can improve the accuracy and reliability of diagnosis results, providing easier and more reliable assistance for those involved in the diagnosis process.

References

Annisa, P. R. (2023). *Penggunaan algoritma artificial neural network (ANN) metode backpropagation dalam memprediksi lama studi mahasiswa prodi S-1 Matematika FMIPA UNAND* (Doctoral dissertation, Universitas Andalas).

Binabar, S. W., Stmik, I., & Pekalongan, W. P. (2017). Optimasi parameter K pada algoritma KNN untuk deteksi penyakit kanker payudara. *Jurnal STMIK Widya Pratama*, *12*(2). http://jurnal.stmik-wp.ac.id

Crosby, D., Bhatia, S., Brindle, K. M., Coussens, L. M., Dive, C., Emberton, M., ... & Balasubramanian, S. (2022). Early detection of cancer. *Science*, *375*(6586), eaay9040.

Global Cancer Observatory. (2022). *World fact sheet*. International Agency for Research on Cancer. https://gco.iarc.who.int/media/globocan/factsheets/populations/900-world-fact-sheet.pdf

Nasien, D., Enjeslina, V., Hasmil Adiya, M., & Baharum, Z. (2022). Breast Cancer Prediction Using Artificial Neural Networks Back Propagation Method. *Journal of Physics: Conference Series*, *2319*(1). https://doi.org/10.1088/1742-6596/2319/1/012025

Nugraha, I. G. W., Santoso, A. L., & Hernanda, P. Y. (2024). Analisa faktor genetik terkait estrogen receptor terhadap kejadian kanker payudara pada wanita. In *Prosiding Seminar Nasional COSMIC Kedokteran* (Vol. 2, pp. 157-175)
.

Nurelasari, E. (2018). Penerapan metode neural network berbasis algoritma genetika untuk prediksi penyakit kanker payudara. *Journal SPEED: Sentra Penelitian Engineering dan Edukasi*, *10*.

Obeagu, E. I., & Obeagu, G. U. (2024). Breast cancer: A review of risk factors and diagnosis. *Medicine*, *103*(3), e36905.

Ridho, I. I., Ramadhani, C. F., & Windarto, A. P. (2023). Penerapan Artificial Neural Network dengan Metode Backpropagation Dalam Memprediksi Harga Saham (Kasus: PT. Bank BCA, Tbk). *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)*, *8*(1), 295-303.

Rizaty, M. A. (2021). Ini Jenis Kanker yang Paling Banyak Diderita Penduduk Indonesia. https://databoks.katadata.co.id/datapublish/2021/06/29/ini-jenis-kanker-yang-palingbanyak-diderita-penduduk-indonesia

Saini, A., Kumar, M., Bhatt, S., Saini, V., & Malik, A. (2020). Cancer causes and treatments. *Int J Pharm Sci Res*, *11*(7), 3121-3134.

Setiawan, I. M. A. (2023). Peran Pemeriksaan Imunohistokimia Dalam Diagnosis Dan Prognosis Kanker Payudara. *Cermin Dunia Kedokteran*, *50*(8), 443-446.

Singh, P., Singh, N., Singh, K. K., & Singh, A. (2021). Diagnosing of disease using machine learning. In *Machine learning and the internet of medical things in healthcare* (pp. 89-111). Academic Press.

Slamet, A. H. H., Purnomo, B. H., & Soedibyo, D. W. (2020). Model Jaringan Syaraf Tiruan untuk Prakiraan Harga Komponen Bahan Baku Pakan Unggas di PT XYZ. *Industria: Jurnal Teknologi dan Manajemen Agroindustri*, *9*(2), 151-161.

Wahyuni, E. S. (2016). Penerapan Metode Seleksi Fitur Untuk Meningkatkan Hasil Diagnosis Kanker Payudara. *Jurnal SIMETRIS*, *7*(1).