Available *online* at : http://bit.ly/InfoTekJar

**InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan**

ISSN (Print) 2540-7597 | ISSN (Online) 2540-7600

# Performance Enhancement and Accuracy of Artificial Neural Networks Using Particle Swarm Optimization for Breast Cancer Prediction

*Jimmy Nganta Ginting, Ronsen Purba, Erwin Setiawan Panjaitan*

*Program Study Master Of Information Technology, STMIK Mikroskil Medan*
*Jl. M.H Thamrin No.140 Kel, Pusat Ps., Kec. Medan Kota, Kota Medan, Sumatera Utara 20212*

KEYWORDS

*Breast Cancer, Backpropagation Algorithm, Particle Swarm Optimization*.

CORRESPONDENCE

Phone: 0823-6903-6008

E-mail: ngantaiting@gmail.com

**A B S T R A C T**

Breast cancer is the one of leading causes of death among the women in many parts of the world. According to Global Cancer Observatory (GCO) data from WHO (2018) show that approximately 58,256 (16,7%) cancer cases were found in Indonesia out of a total of 348,809 cancer cases. The number of breast cancer patients throughout the world reached 42.1 per 100,000 population on average death rate of 17 per 100,000 inhabitants.Various ways have been used to find effective methods in the early detection of breast cancer. A prediction of breast cancer in early stage is very important in the medical world, which allows them to develop strategic programs that will help diagnose and reduce mortality rates from breast cancer. Performance enhancement and accuracy of artificial neural networks using particle swarm optimization is an effective solution for breast cancer prediction. The accuracy result was found 70% for training data and 96.1% for 30% prediction in this study. Previous studies only used the backpropagation algorithm to predict breast cancer and the result was 94.17%. Compared with previous study, there is an increase of 1.93% in combining Backpropagation with Particle Swarm Optimization.

## INTRODUCTION

Prediction is the act of forecasting what will happen in the future. Prediction is central to medicine as preventive and therapeutic interventions are prescribed or recommended onimplicit or explicit expectations about future health outcomes/ The medical model in which interventions are targeted to the induividual or to risk groups rather than given to population at large is known as Precision medicine [1].Prediction of breast cancer is critical in medical world because it allows them to develop strategic program that will help to decrease the affected [26].

Breast cancer is the leading cause of cancer death in women around the world. According to Global Cancer Observatory (GCO) data from WHO in 2018 show that approximately 58,256 (16,7%)cancer cases were found in Indonesia out of a total of 348,809 cancer cases. The number of breast cancer patients throughout the world reached 42.1 per 100,000 population on average death rate of 17 per 100,000 inhabitants [16]

Patricio, et al., (2018) has conducted the study to predict the presence of breast cancer based on four parameters namely Resistin, Glucose, Age and Body Mass Index (BMI)[24]. By using of three methods comparison namely logistic regression, random forests, and support vector machines.It was found that the support vector machines method is better than the other methods. The sensitivity and specificity of the support vector machines method were obtained 88% and 90%, respectively. In other hand, Jaisankar & Victorseelan (2019) have conducted studies using a comparison of statistical discriminant analysis and artificial neural network model for the prediction of breast cancer. The study reveals that higher accuracy is provided by Neural Network analysis (94.17%) than Discriminant analysis (77.6%) in terms of prediction[16]. The study has been conducted by Bayrak, Kirci, & Ensari (2019) in early predicting of breast cancer using Artificial Neural Network (ANN) and Support Vector Method (SVM). The study reveals that 96 % was obtained for accuracy of SVM Method and 95 % for ANN Method [7].

Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way the biological nervous system such as brain process information [25]. ANN combined with backpropagation algorithm most commonly used for prediction study due to their universal approximation capabilities and flexible structure that allow to capture complex non liniers behaviours. The limitations of using *Backpropagaiton* are the existence of temporary, local minima resulting from the saturation behaviour of the activation function, and the slow rates of convergence [32]. Therefore, with the presence of *Particle Swarm Optimization* (PSO), the limitations of *backpropagation* can be solved. *Particle Swarm Optimization* (PSO) as a determination of the initial weight

https://doi.org/10.30743/infotekjar.v5i1.2939

value on the input attribute (attribute weighting) in predicting a decision, thereby speeding up the process of iteration or calculation of trainer data.

Based on the background that has been explained above, The problem formulation of this paper is how to find a way for improving backpropagation algorithm performance and accuracy to obtain more accurate prediction results. Where if using the backpropagation dataset will be randomly predicted so that the iteration level is relatively slow, the time of the training data is long to get the output. *Particle swarm optimization is used to solve the limitations of Backpropagaiton algorithm* by weighting the attributes first with PSO before making a prediction of breast cancer using the backpropagation algorithm.Therefore, we get faster iteration process and more leads to output results.

The aim of this research was to improving of *artificial neural network* performance with backpropagation algorithm by *particle swarm optimization model. We plan that combaining* backpropagation algorithm with *Particle Swarm Optimization* dapat could improve the performance and accuracy in breast cancer diagnosis

## METHOD

Research methodology is the description of tbe specific procedures or techniques used to identify, select, process and analyze information. This research is started by input the whole of dataset. The type of data used in this study is secondary data that has been used in previous studies. The data obtained from the *Uci Repository Machine Learning* https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra.

The data has been used in previous studies for prediction of breast cancer. The dataset has 116 medical records with 9 attributes and 1 classification diveded into 2 healthy and cancer categories. The criteria for 9 attributes can be seen in table 1.

Tabel 1 The criteria of dataset attribute

| No | Attribute Names | Range | Attribute Information |
|---|---|---|---|
| 1 | *Age (year)* | 24 – 89 year | Age of the patient |
| 2 | *Body Mass Index (BMI)* | 18,67 kg/m$^2$ – 38,57 kg/m$^2$ | Idea weigh |
| 3 | *Glucose* | 60 mg – 201 mg | Glucose level |
| 4 | *Insulin* | 2 unit – 59 unit | The liquid to stabilize blood sugar levels |
| 5 | *Homeostasis* | 0,467 $^o$C – 25,050 $^o$C | Body temperature |
| 6 | *Leptin* | 4,311 ng/mL – 83,48 mg/mL | A hormone produced by the fat cells in human body. |
| 7 | *Adiponectin* | 1,656 ug/ml – 22,432 ug/ml | Protein hormone |
| 8 | *Resistin* | 3,21 mg/dL – 55,21 mg/dL | Protein that could caused obesity |
| 9 | *Monocyte Chemoattractant Protein -1 (MCP-1)* | 63,61 pg/mL – 1698,44 pg/mL | Their roles in the immune system |

Figure 1 shows the flowchart of this study. The flowchart is represents a workflow or process of this research. The research began with seeking dataset record of breast cancer patients, the next step is data normalization, and atribute weight using particle swarm optimization (PSO) modeling. After that forecasting the result of the calculation using backpropagation algorithm.

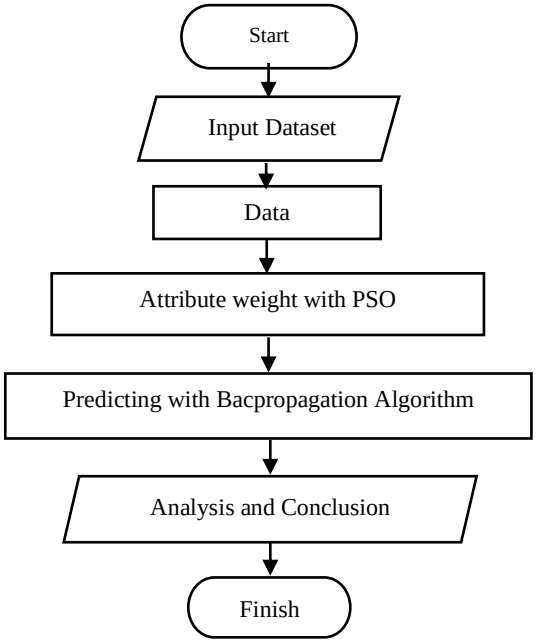


Figure 1. Research Flow Chart

## RESULTS AND DISCUSSION

The initial step in this research is input data. The class for the normal state is assigned the class label 1 and the class with abnormal state (breast cancer) is assigned the class label 2. The next step is data normalization. The aim is to set the value measured on the uniform scale or range, namely 0-1. After finding the similar scale, the next step is attribute weight dataset that has been normalized with particle swarm optimization algorithm. The final weight value was obtained as a result. Then, the next step is predict the data obtained by backpropagation algorithm.

This research used 116 medical records data of which the number of normal (healthy) patients are predicted around 52 and 64 patients with breast cancer. The datasets contain 9 attributes and one classification of which are devided into 2 category namely healthy control and patient with breast cancer class. This method is according to Jaisankar & Victorseelan (2019) study that divided the data become 70% of the data were used

for training and 30 percent for testing model. Basedon present research, The training is 70% (81 medical record data) of which from 116 data are taken 37 data as category of healthy control and 44 data as category of patient (breast cancer). For testing models are used 30% (35 medical records data) of which category of healthy control is 15 and 20 data as patient.The first taken is data of 30% and then data of 70%.
The results of 70 % and 30 % data of accuracy prediction are shown in Table 2.

Table 2. The result data of 70% and 30% accuracy prediction

| No | Iteration | BestCost | Accuracy(%) |
|---|---|---|---|
| 1 | 100 | 6.3280e-07 | 94 |
| 2 | 200 | 2.3770e-28 | 97 |
| 3 | 300 | 2.0090e-57 | 97 |
| 4 | 400 | 7.3178e-89 | 94 |
| 5 | 500 | 3.5816e-120 | 100 |
| 6 | 600 | 1.6248e-152 | 100 |
| 7 | 700 | 3.9719e-189 | 94 |
| 8 | 800 | 8.1412e-216 | 94 |
| 9 | 900 | 8.3387e-260 | 94 |
| 10 | 1000 | 2.4456e-291 | 97 |
| **Average** | | | **96.1** |

Based on Table 2, we can make accuracy percentages plot to percent. The graphic is shown in Figure 2.

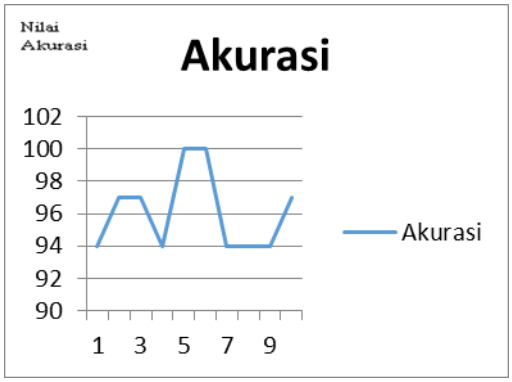


**Figure 2 Graphic Accuracy**

## CONCLUSIONS

Based on the research result was obtained the values of accuracy using Particle Swarm Optimization for improving of performance and accuracy Artificial Neural Network for breast cancer prediction namely 96.1% (30 % predicting data and 70% training data). The best of accuracy value was obtained at iteration of 500 and 600 namely 100% Meanwhile, the mean accuracy values is 96.1 %. It shows that there is an increase when two of algorithm were combined.This present result is better than compared with the result of Jaisankar dan Victorseelan (2019) study of which the result was obtained using *Artificial Neural Network* (*ANN*) on *Backpropagation algorithm* namely 94,17%. It shows that there is an increase around 1.93 %.

Table 3. Comparison of Accuracy

| Backpropagation | Combination of Backpropagation & PSO |
|---|---|
| Accuracy = 94,17% | Akurasi 96,1% |

## ACKNOWLEDGMENT

I also express my gratitude for the help of Mr. Ronsen Purba and Mr. Erwin Setiawan Panjaitan who participated in completing this research. have given you the time and knowledge that you have. As well as families who have provided financial support and provided encouragement until the completion of this research.

## REFERENCES

[1] A. Cecile J.W.Janssens, & ForikeK. Martens., 2018. Prediction Research. *University Medical Center, Department of Clinical Genetics/EMGO Institute for Health and Care Research, Amsterdam, The Netherlands*.
[2] A. K. Santra, C.& Josephine Christy.,2012. Genetic Algorithm and Confusion Matrix for Document Clustering. *International Journal of Computer Science Issues*, Vol. 9, Issue 1, No 2.
[3]Arman Hakim Nasution & Yudha Prasetyawan.,2008. Perencanaan & Pengendalian Produksi. Graha Ilmu, hal 29-37.
[4] Breast Cancer Hospital Authority.,2017. Kanker Payudara.https://www21.ha.org.hk/ smartpatient/EM/MediaLibraries/EM/Diseases/Cancer/Breast%20Cancer/Cancer-Breast-Cancer Indonesian.pdf?ext=.pdf
[5] Badrudin Hadibrata.,2010. Data Mining. *https://www.academia.edu/7284965/Dosen_Badrudin_Hadibrata_S.T_M.Kom.*Academia edu.
[6] Daniel T.Larose.,2005.*Discovering Knowledge In Data An Introduction To Data Mining*. Director of Data Mining Central Connecticut State University
[7] Ebru Aydindag Bayrak, Pinar Kirci & Tolga Ensari., 2019. Comparison Of Machine Learning Methods For Breast Cancer Diagnosis. *Department Of Engineering Sciences anf Computer Engineering.*
[8] Elah Nurlela &M. Sukrisno Mardiyanto.,2019. Pemilihan Atribut Pada Algoritma C4.5 Menggunakan Particle Swarm Optimization Untuk Meningkatkan Akurasi Prediksi Diagnosis Penyakit Liver.*Jurnal PILAR Nusa Mandiri* Vol. 15, No. 2
[9] Eko Prasetyo.,2012. Data Mining Konsep dan Aplikasi menggunakan Matlab. ANDI, hal 1-7.
[10]Elyse J. Watkins, DHSc, PA-C, DFAAPA., 2019. Overview Of Breast Cancer.JAAPA Journal of the American Academy of Pas.
[11] Fajar Astuti Hermawati.,2013. Data Mining. ANDI, hal 23-29.
[12] Gunawan Wibisono, Arief Hermawan., 2019. Faktor-Faktor Penentu Gejala Penyakit Kanker Payudara Dengan Pendekatan Jaringan Syaraf Tiruan. Universitas Teknologi Yogyakarta. JASIEK, Vol. 1, No. 1, PP 1-6.


https://doi.org/10.30743/infotekjar.v5i1.2939

[13] Han, J., Kamber, M. & Pei, J., 2012. *Data Mining Concepts and Techniques*. Third ed. USA: Morgan Kaufmann.

[14] Han-Byoel Lee & Wonshik Han.,2014. Unique Features of Young Age Breast Cancer and Its Management.*Journal Of Breast Cancer*, pISSN 1738-6756 eISSN 2092-9900.

[15] Herdianto, 2013, Prediksi Kerusakan Motor Induksi Menggunakan Metode Jaringan Syaraf Tiruan Backpropagation. *Thesis Universitas Sumatra Utara Medan*.

[16] International Agency For Research On Cancer., 2018. World Health Organization. *The Global Cancer Observatory*.

[17] Jui-Fang Chang, Chi-Ming Kuan, & Yu-Wen Lin., 2009. Forecasting Exchange Rates by Genetic Algorithms Based Back Propagation Network Model. *International Conference on Intelligent Information Hiding and Multimedia Signal Processing.*

[18] Santosa, Budi & Willy, Paul. 2011. Metoda Metahuristik, Konsep Dan Impelementasi. Surabaya: Graha Ilmu.

[19] Siti Rahmiati Pratiwi, Efri Widianti, & Tetti Solehati.,2017. Gambaran Faktor-Faktor yang Berhubungan dengan Kecemasan Pasien Kanker Payudara dalam Menjalani Kemoterapi. Universitas Padjadjaran Bandung.

[20] Massimo Buscema, 1998. Back Propagation Neural Networks. *See discussions, stats, and author profiles for this publication at: https://www.researchgate .net/publication/13731614*.Substance Use & Misuse, 33(2), 233–270.

[21] Mahendra.,2013. Perbedaan: precision, recall & accuracy.*universitas brawijaya fakultas of computer science*.

[22] Muhammad Ali Ridla., 2018. Particle Swarm Optimization Sebagai Penentu Nilai Bobot Pada Artificial Neural Network Berbasis Bacpropagation Untuk Prediksi Tingkat Penjualan Minyak Pelumas Pertamina. *Jurnal Ilmiah Informatika*. Volume 3 No 1.

[23] Nyce, Charles (2007), Predictive Analytics White Paper(PDF), American Institute for Chartered Property Casualty Underwriters/Insurance Institute of America,p. 1

[24] Patricio, et al., 2018. Using Resistin Glucose Age And BMI To Predict The Presence Of Breast Cancer. *BMC cancer*, pp 01-08

[25] Parag Singhal & Saurav Pareek., 2018. Artificial Neural Network for Prediction Of Breast Cancer. *Proccedings Of The Second International Conference On I-SMAC (Iot In Social, Mobile , Analytics And Cloud)*. ISBN:978-1-5386-1442-6.

[26] R, Jaisankar. & B, Victorseelan., 2019. A Comparison Of Statistical Discriminant Analysis And Artificial Neural Network Model For The Prediction Of Breast Cancer. *International Journal Of Scientific Research In Mathematical and Statistical Sciences*, Volume 6, pp. 284-289.

[27] Retno Putri, I., 2015.Optimasi Metode Adaptive Fuzzy K-Nearest Neighbor Dengan Particle Swarm Optimization Untuk Klasifikasi Status Sosial Ekonomi Keluarga . Universitas Brawijaya.

[28] Ricard Konig.,2018. Predictive Techniques And Methods For Decision Support In Situation With Poor Data Quality. Science And Technology, Licentiate Thesis Unversity Boras.

[29] Riska Yanu Fa'rifah & Zulfiqar Busrah.,2017. Backpropagation Neural Network Untuk Optimasi Akurasi Pada Prediksi Financial Distress Perusahaan. JURNAL INSTEK, Volume 2 Nomor 2.

[30] Siang , JJ., 2005. Artificial Neural Network & Pemograman Menggunakan Matlab. Yogyakarta: Andi Offset.

[31] Y. A. Lesnussaa, C. G. Mustamua, F. Kondo Lembanga, & M. W. Talakuaa.,2018. Application Of Backpropagation Neural Networks In Predicting Rainfall Data In Ambon City. International Journal Of Artificial Intelegence Research,ISSN: 2579-7298 Volume 2, No 2.

[32] Y.H. Zweiri, J.F. Whidborne, & L.D. Seneviratne., 2003. A three-term backpropagation algorithm. Elsevier Science, Neurocomputing 50 (2003) 305–318

https://elib.unikom.ac.id/files/disk1/396/jbptunikompp-gdl-andriansya-19792-7-babii--i.pdf. Andriansya. Jaringan Syaraf Tiruan.BabII