C3

by Adi Wibowo

Submission date: 03-Apr-2023 09:10AM (UTC+0700)
Submission ID: 2054023889
File name: 14_6_26.pdf (710.38K)
Word count: 8061
Character count: 43796

Journal of Engineering Science and Technology
Vol. 14, No. 6 (2019) 3437 - 3456

OPTIMIZATION OF BACKPROPAGATION USING NGUYEN-WIDROW AND STIMULUS-SAMPLING FOR BREAST CANCER CLASSIFICATION AND FEATURE SELECTION

ADI WIBOWO[1,*], DWI SETIANTO[1], PANJI WISNU W.[1], WISNU JATMIKO[2]

[1]Department of Informatics, Diponegoro University
Tembalang Campus Jl. Prof. Soedarto SH Tembalang, Semarang, Central Java, Indonesia
[2]Faculty of Computer Science, Universitas Indonesia, Indonesia
Kampus UI Depok, Depok, West Java, Indonesia 16424
*Corresponding Author: bowo,adi@live.undip.ac.id

Abstract

Breast cancer is one of the leading causes of death in the world for women. Early detection and diagnosis can be done to reduce the cancer death rate. MicroRNA is known as a biomarker for breast cancer and with the help of artificial neural network technology, made it possible to perform a classification process for early detection. Backpropagation neural network has good performance in classification, however, still has a drawback related to its long training time. This research is conducted to classify breast cancer (whether a cell is cancer or normal and whether it is before or after metastatic stage) based on microRNA profiles using backpropagation with Nguyen-Widrow and Stimulus-Sampling algorithm optimization. In this paper, three breast cancer datasets are used to compare the classification performances. Furthermore, some alternatives microRNA features set are obtained using feature selection methods and compare the accuracy values. The results show that the combination of Nguyen-Widrow and Stimulus-Sampling algorithm produces the best backpropagation performance based on the accuracy, sensitivity, specificity, and AUC value as well as reducing the running time. The combination of Nguyen-Widrow and Stimulus-Sampling algorithm proved to be able to increase the performance of the method. Nguyen-Widrow algorithm provides weights value initialization that is not too small or too large so that the convergence process can be enhanced. Meanwhile, the use of Stimulus-Sampling improves the performance of backpropagation by strengthening the output unit, which give the smallest error value.

Keywords: Artificial neural network, Breast cancer, MicroRNA, Nguyen-widrow, Stimulus-sampling.

1. Introduction

Breast cancer is the second most common cause of death because of cancer among women [1] with 2,088,849 cases in the world in 2018, and 626,679 of them were declared dead [2]. The number of female deaths due to breast cancer cases occupies the highest position after cervical cancer [3]. One effort that can be done to reduce the cancer death rate is to do early detection. Early detection of cancer can not only reduce cancer deaths but also improve the patient's quality of life [4]. Women who have been detected breast cancer early (early stages) and then given adequate treatment have a higher percentage of survival [5]. Even a late diagnosis and improperly treatment could lead to a metastatic stage where cancer spread to lymph nodes and distant organs [6].

One way to do early detection of breast cancer is to classify the microRNA profile. Each cell, whether cancer cell or normal, has different microRNA expression profile. MicroRNA is a single-stranded RNA form, and about 50% of the recorded microRNAs are found to be related to cancer and are located in fragile places, which are areas where a genome is associated with cancer [7]. As a biomarker, microRNA can indicate an outstanding development of cancer. This is because it plays an essential role in almost all cell activities, even abnormal activities [7]. A microRNA profile consists of several features. A study had proved that some features showed the most significant differences between normal cells and breast cancer cells [8]. Another study by Khasburrahman et al. [9] even compared some feature sets of MicroRNA using feature selection; the features were selected using Weka's Wrapper Subset Eval Attribute Evaluator and then were classified using Multi-Layer Perceptron and Naïve Bayes. Madhavan et al. [10] proved that MicroRNA could be used to recognize whether a cell is a cancer cell or not even in the earliest stage by reviewing the application of MicroRNA as a biomarker. A study by Yan et al. [11] reported that aberrant expression of microRNA profile also involved in an advanced stage of breast cancer.

The microRNA profile classification process can be achieved using an artificial neural network that can be applied using a computer. Backpropagation has excellent performance in generalization and good learning ability, making it suitable for object classification with many features [12]. Backpropagation as a method of classification in the medical field has been considered as a powerful method [13]. An experimental classification of the Wisconsin Breast Cancer Diagnostic (WBCD) dataset with 30 features showed the accuracy of backpropagation outperformed the k-NN and SVM methods [14]. Supervised machine learning such as backpropagation, Deep Learning, Naïve Bayes, and Decision Tree had been applied and successfully classified cancer cells based on microRNA profiles [15]. A study applied the deep belief network to classify breast cancer based on WBCD and produced an accuracy of 99.68%, the construction used backpropagation with Liebenberg Marquardt learning function [16]. Other studies had also utilized microRNA for the classification of ovarian cancer, breast cancer, and lung cancer using the backpropagation algorithm with the highest accuracy obtained was 83.33% without adding any optimization function [17]. The performance of the backpropagation algorithm had also been compared with the Naïve Bayes algorithm, with breast cancer case studies, showed that the backpropagation algorithm produced better accuracy than Naïve Bayes [9].

However, backpropagation still has a major drawback, specifically the high of training time [18]. This backpropagation processing capability is determined by the relationships pattern between neurons and the initialization of weights value. The large input units and hidden units affect the training time process. Moreover, the initial weight value that is distant from the actual weight value will also slow down the training process.

Various studies have attempted to overcome this weakness by implementing optimization algorithms such as Genetic Algorithm (GA) [19], Ant Colony Optimization (ACO) [20], Particle Swarm Optimization (PSO) [21], Ant Lion optimizer (ALO) [22], Social Spider Optimization (SSO) [23], Artificial Bee Colony (ABC) [24], etc. However, most of the above optimization required a more training time compared to the standard algorithm. Aisyah et al. [25] used the Nguyen-Widrow optimization algorithm to shorten the training time, but there was no significant improvement of the accuracy value. Then Gorunescu and Belciug [26] used the Stimulus-Sampling algorithm to their study of five diseases data. This addition could improve accuracy, but the training time needed is not presented. Both Nguyen-Widrow and Stimulus-Sampling algorithm had been successfully implemented to backpropagation to classify some medical datasets [26, 27]. The combination of these optimization algorithms was interesting since both Nguyen-Widrow and Stimulus-Sampling algorithm were implemented inside the standard backpropagation algorithm.

In this paper, we combined Nguyen-Widrow and Stimulus-Sampling algorithm to the backpropagation method to produce a classification model with better performance and faster training time. The Nguyen-Widrow algorithm was used to produce faster training process by modifying the weight initialization [28], while the Stimulus-Sampling algorithm gave rewards to each output unit in the next training step [26]. Three breast cancer datasets from TCGA Research Network, WBCD dataset and the selected breast cancer datasets [29] are used to compare the classification performances. The WBDC dataset is used for comparison with previous studies for Stimulus-Sampling algorithm, the TCGA datasets (A) and the selected breast cancer datasets (B) are datasets that use microRNA biomarkers to classify breast cancer. The A dataset is used to classify normal cell and breast cancer cell classes, while the B dataset contains classes of breast cancer before the phase of metastasis and class of breast cancer after the metastasis phase. Furthermore, we explored subsets of features from the original dataset using feature selection methods (Correlation matrix, SVM-RFE, Univariate) to obtain the possibility of higher accuracy value compare to a feature set that is a result of medical analysis.

2. Materials and Methods

2.1. Data collection

This paper uses three breast cancer datasets. First, the Dataset A was obtained from the GDC Data Portal National Cancer Institute through the site http://gdc.cancer.gov/. It was TCGA data generated by TCGA Research Network and was ready for high-level analyses. TCGA data can be used to analyse the correlations of various factors of clinicopathology associated with cancer initiation, progression, and invasion [30, 31]. The data was divided into two classes, namely 100 normal class data and 100 breast cancer class data. According to research by

Iorio et al. [8], MicroRNA features used were hsa-miR-10b, hsa-miR-21, hsa-miR-125b-1, hsa-miR-125b-2, hsa-miR-145, and hsa-miR-155. Secondly, we used the dataset B that obtained from a study by Rosenfeld et al. [29]. This data collection contained 48 microRNA data, divided into two classes, 45 breast cancer data in metastasis stage and three breast cancer data before in metastasis stage. Thirdly, based on research by Gorunescu and Belciug [26], Wisconsin Breast Cancer Diagnostic (WBCD) dataset that applied in the previous is utilized to compare the algorithm. In addition, some feature selection methods were used to obtained some new features set and compared the accuracy values to previous datasets.

2.2. Backpropagation neural network

Backpropagation is the most famous method to train an artificial neural network because it has a simple mathematical analysis and good representation capabilities [14]. Backpropagation trains the network to get a convergence between the ability of the net to recognize the patterns used during training and the strength of the system to provide the correct response to input patterns that are similar to the models used during training [9]. With multilayer perceptron as the neural network structure, the three main activities of backpropagation are to feed-forward from the input layer to the output layer, calculate the error, and update the weights of the output layer to the input layer [32]. The steps of backpropagation learning algorithm are as follows:

- Initialize the v_{ji} weights for each input (x_i, $i = 1, 2, ..., n$) to each hidden unit and w_{kj} weights for each hidden unit (z_i, $i = 1, 2, ..., p$) to each output unit (y_i, $i = 1, 2, ..., m$).
- Calculate output in each hidden unit (z_j) using Eqs. (1) and (2).

$$z\_net_j = v_{j0} + \sum_{i=1}^{n} x_i v_{ji} \tag{1}$$

$$z_j = f(z\_net_j) = \frac{1}{1+e^{-z\_net_j}} \tag{2}$$

- Calculate the output in each output unit (y_k) using Eqs. (3) and (4).

$$y\_net_k = w_{k0} + \sum_{j=1}^{p} z_j w_{kj} \tag{3}$$

$$y_k = f(y\_net_k) = \frac{1}{1+e^{-y\_net_k}} \tag{4}$$

- Calculate the factor δ in each unit output and calculate the rate of change in the weight of each hidden unit (Δw_{kj}) using Eqs. (5) and (6).

$$\delta_k = (t_k - y_k) y_k (1 - y_k) \tag{5}$$

$$\Delta w_{kj} = \alpha \, \delta_k \, z_j \tag{6}$$

- Calculate the factor δ in each hidden unit and calculate the rate of change in weight for each input unit (Δv_{ji}) using Eqs. (7) to (9).

$$\delta\_net_j = \sum_{k=1}^{m} \delta_k w_{kj} \tag{7}$$

$$\delta_j = \delta\_net_j z_j (1 - z_j) \tag{8}$$

$$\Delta v_{ji} = \alpha\, \delta_j\, x_i \tag{9}$$

- Calculate the new weight using Eqs. (10) and (11).

$$w_{kj}(new) = w_{kj}(old) + \Delta w_{kj} \tag{10}$$

$$v_{ji}(new) = v_{ji}(old) + \Delta v_{ji} \tag{11}$$

- Update the epoch value and calculate the Mean Squared Error (MSE) using the Eq. (12).

$$MSE = \frac{\sum_{k=1}^{n}(t_k - y_k)^2}{n} \tag{12}$$

2.3. Nguyen-widrow algorithm

Initial weight value will influence how fast a network reaches its convergence. The general procedure used on backpropagation in initializing weight value is random initialization. The method of initial weight initialization using Nguyen-Widrow can accelerate learning time compared to random initialization. Nguyen-Widrow optimization initializes the weight of the input unit to the hidden unit. Weight values are initialized within a certain range by considering the number of hidden units [28]. The Nguyen-Widrow (β) scale factor is calculated by Eq. (13) then weights initialization is calculated by Eqs. (14) and (15).

$$\beta = 0.7(p)^{1/n} \tag{13}$$

$$||v_{ji}|| = \sqrt{v_{j1}(old)^2 + \cdots + v_{jn}(old)^2} \tag{14}$$

$$v_{ji} = \frac{\beta\, v_{ji}(old)}{||v_{ji}||} \tag{15}$$

2.4. Stimulus-sampling algorithm

The stimulus-sampling theory states that certain stimulus-response relationships can be learned in learning, but the whole learning process remains continuous and is the accumulation of these stimulus-response discrete values [33]. This theory is applied to the backpropagation algorithm because it is expected to improve the performance of backpropagation. This application involves competition between output units to produce the best output, strengthen the response of these neurons by giving 'reward' [26]. At each step, the stimulus-sampling algorithm will strengthen or weaken the response of the neurons in the output layer, depending on their performance from the previous step.

This approach assigns a counter c on each neuron output. Counter c is initialized with value 0, then each c_j ($j = 1, 2, ..., q$) changes its value according to the learning result, gets a value of 1 if the output of neuron j produces a response that is closest to the output target and 0 for the output of other neurons. Each output unit will receive the s_j stimulus according to its performance with the s_j value calculated by Eq. (16) at the end of each epoch. The value of the neuron output for the epoch is then multiplied by the value of s_j as in Eq. (17).

$$s_j = \frac{c_j}{\sum_{j=0}^{q} c_j} \tag{16}$$

$$y_k = f\left(y_{net_k}\right) s_j = \frac{1}{1+e^{-y\_net_k}} s_j \quad (17)$$

2.5. Feature selection

Nowadays, the domain features of datasets used for classification or pattern recognition had expanded into very high dimension [34]. Mostly, this could lead to overfitting problems and high execution times, since there were many redundant, noisy and irrelevant features [35]. Applying feature selection over the dataset by eliminating noisy features or capturing the most important features could enhance the classification performance, reduce the execution time and help in understanding data [36]. SVM-RFE is one of the most powerful feature selection algorithms [37] and has successfully used in gene selection, producing higher accuracy, 98% [38]. Another method for feature selection in gene selection is univariate selection. Univariate selection proved to be able to perform very well in gene selection, even outperformed the multivariate approach [39]. Some studies by Malik and Mishra [40] and Reif et al. [41] used RapidMiner to conduct feature selection because it contains modules for features selection methods and machine learning, these studies used one of the simplest, i.e., correlation matrix. Correlation matrix used statistical technique to calculate how strong the pairs of attributes were related [42].

2.6. Evaluation methods

The evaluation methods used to test this experiment were accuracy, sensitivity, specificity and execution time. These values were also used as the fitness functions along with MSE value. First, we calculated the value of true positive, true negative, false positive, and false negative. True positive is the number of data that is correctly identified as the positive class. True negative is the number of data that is correctly identified as negative class. False-positive is the number of data that is incorrectly identified as the positive class. False-negative is the number of data that is incorrectly identified as a negative class.

Accuracy is a value that shows how similar value is predicted to the actual value. The formula is as in Eq. (17). Sensitivity is a value that shows how many positive conditions are correctly classified as positive. The formula is as in Eq. (18). Specificity is a value that shows how many negative conditions are correctly classified as negative. The formula is as in Eq. (19).

$$Accuracy = \frac{\sum \text{True positive} + \sum \text{True negative}}{\sum \text{Total Population}} \quad (17)$$

$$Sensitivity = \frac{\sum \text{True positive}}{\sum \text{Condition positive}} \quad (18)$$

$$Specificity = \frac{\sum \text{True negative}}{\sum \text{Condition negative}} \quad (19)$$

3. Results and Discussion

In this study, there were four scenarios. The first scenario aimed to get the best value of training parameters (learning rate and a number of hidden units). The second scenario aimed to get the best model for classification of breast cancer based on MicroRNA profile. In the third scenario, the performance of the best model from the second scenario was compared to the performance of the model if another dataset was used. This other dataset was not microRNA profile, but WBCD Dataset

was used in the previous research by Gorunescu and Belciug [26]. The last scenario, feature selection methods were used to obtain some alternatives features set and the accuracy values were compared. This study was done on a system running Windows 10 OS powered with Intel Core i7 processor with 4 GB RAM.

The first scenario was done using the standard backpropagation method with feedforward network architecture as shown in Fig. 1. In this scenario, the combination of the learning rate and the number of hidden units were used. A study has stated that the best number of hidden units usually ranges between the number of output units and the number of input units [26].

In this case, there were 2 unit outputs and 6 unit inputs for dataset A and 48 unit inputs for dataset B, so the range of the number of hidden units used in this scenario were 2 to 6 for dataset A and 2 to 48 for dataset B. Whereas, for the learning rate (alpha), the initial experiment was carried out using values of 0.001, 0.01, 0.1, 0.5, and 0.9. The proposed approach (backpropagation with Nguyen-Widrow and Stimulus-Sampling algorithm optimization) was in following steps:

- Initialize the v_{ji} weights for each input (x_i, $i = 1, 2, ..., n$) to each hidden unit and w_{kj} weights for each hidden unit (z_i, $i = 1, 2, ..., p$) to each output unit (y_i, $i = 1, 2, ..., m$) using Nguyen-Widrow algorithm as in Eqs. (13) to (15). Initialize counter c_j ($j = 1, 2, ..., q$) to 0.
- Calculate output in each hidden unit using Eqs. (1) and (2).
- Calculate the output in each unit output using Eqs. (3) and (4).
- Calculate the factor δ in each unit output and calculate the rate of change in the weight of each hidden unit using Eqs. (5) and (6).
- Calculate the factor δ in each hidden unit and calculate the rate of change in weight for each input unit using Eqs. (7) to (9).
- Calculate the new weight using Eqs. (10) and (11).
- Update the epoch value and calculate the Mean Squared Error (MSE) using the Eq. (12) for each output unit.
- Update counter c_j ($j = 1, 2, ..., q$) based on error value of each output unit. Iterate 1 for output unit with the smallest error value.
- At the end of each epoch, value of the output unit is then multiplied by the value of s_j as in Eq. (16) and (17).
- The iteration is terminated when the stop condition is reached.

The results of first experiment are shown in Table 1. For dataset A, the alpha value of 0.001 still gave a low accuracy value, which was 69.29%. This happened because the alpha value, which was too small would made the learning rate slow and required more iterations, therefore, eventhough the maximum number of epochs had been met, the network had not reached the specified error target.

At the next alpha value, the accuracy value experienced a significant increase and began to stabilize when the alpha value was 0.1. Therefore, further testing was carried out using a range of alpha values from 0.1 to 0.9 with an increase of 0.1.

Whereas, for dataset B, the accuracy value seemed stable, so no further testing needed. The highest accuracy value, 87.50%, was obtained from learning rate 0.1.

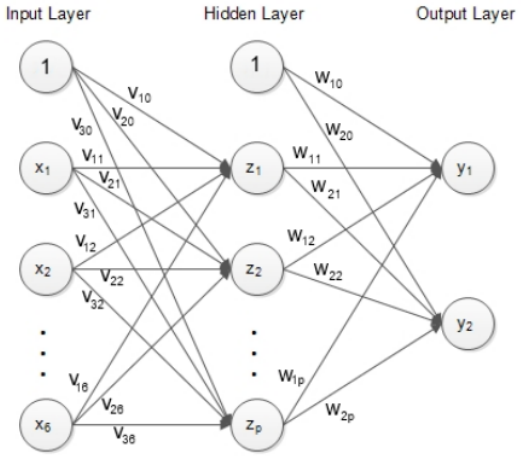

Fig. 1. Backpropagation network architecture.

Table 1. Test Results with alpha value 0.001, 0.01, 0.1, 0.5, and 0.9.

Alpha	Accuracy	
	Dataset A	Dataset B
0.001	69.29%	75.00%
0.01	96.56%	84.72%
0.1	97.77%	87.50%
0.5	98.59%	79.17%
0.9	98.15%	77.08%

The combination of testing using hidden unit ranges 2-6 and alpha values 0.1 - 0.9 means that it required 45 times of testing. Graphs of the effect of learning rate (alpha) on the accuracy values as shown in Fig. 2 and graphs of the effect of the number the hidden unit for the accuracy value is shown in Fig. 3.

Based on Fig. 2, started when the alpha value was 0.1, the accuracy value increased until it reached its peak when the value was 0.4. The accuracy value then decreased until the alpha value was 0.8 then rose again when the alpha value was 0.9. Figure 3 shows that the graph of the effect of the number of hidden units on the accuracy value was fluctuated with the highest point achieved when using four hidden units.

The highest average accuracy value was produced by the number of 4 hidden units, which was 98.68%. The highest average accuracy value was produced with the learning rate of 0.4, which was 98.80%. The number of hidden units and the best learning rate values were then used in the second scenario.

Based on Table 2, it can be seen that there was no significant effect of changing the number of hidden units. Adding more hidden units only made it costs more training time. The best accuracy, 83.33%, the value was produced with 8 hidden units.

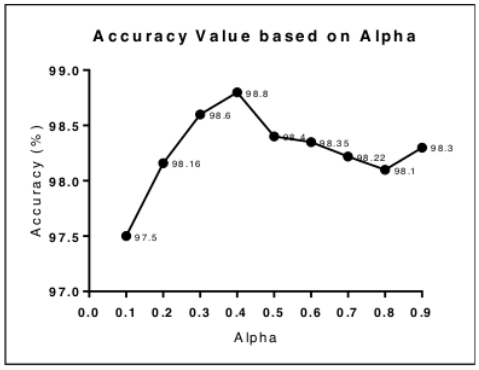

Fig. 2. Effect of alpha value on accuracy value (dataset A).

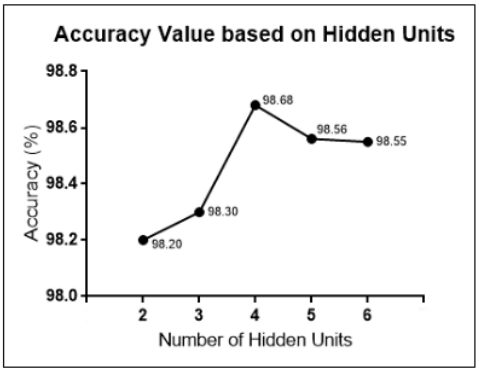

Fig. 3. Effect of number of hidden units on accuracy values (dataset A).

Table 2. Testing results with the number of hidden units 2-48 (dataset B).

Number of hidden units	Accuracy
2	75.00%
4	75.00%
8	83.33%
12	79.17%
16	81.94%
20	81.94%
24	80.56%
28	79.17%
32	77.78%
36	77.78%
40	79.17%
44	81.94%
48	77.78%

The second scenario test included four experiments, namely the experiment to test the performance of standard backpropagation method (BP), backpropagation method with Nguyen-Widrow optimization algorithm (BP-NW), backpropagation method with Stimulus-Sampling optimization algorithm (BP-SS) and backpropagation method with Nguyen-Widrow and Stimulus-Sampling optimization algorithms (BP-NW-SS). In this scenario, the network architecture was created based on the results of the first scenario. Each experiment used 0.4 learning rate, 4 hidden units for dataset A and 8 hidden units for dataset B, maximum epoch 500, error target 0.001, and K-Fold Cross Validation process with $K = 10$ for dataset A and $K = 3$ for dataset B. Each testing in this paper was carried out 10 times, the values in the tables are the average values. The testing results of the second scenario are shown in Tables 3 and 4.

According to Figs. 4 and 5, it can be seen that the BP-NW-SS method produced the best value of accuracy, sensitivity, and specificity, namely 99.05%, 98.60%, and 99.60%, while the fastest execution time was produced by the BP-NW method, which was 94.36 seconds. Similar to the results for dataset A, based on Table 4, the BP-NW-SS method using dataset B produced the best value of accuracy, sensitivity, and specificity, namely 85.00%, 96.19%, and 13.33%, while the fastest execution time was produced by the BP-NW method, which was 16.29 seconds.

Table 3. Second scenario testing results (dataset A).

Method	Accuracy	Sensitivity	Specificity	Time (s)	MSE
BP	98.45%	98.00%	98.90%	102.34	0.0025542
BP-NW	98.90%	98.20%	99.60%	94.36	0.0024475
BP-SS	98.90%	98.30%	99.50%	97.28	0.0023676
BP-NW-SS	99.05%	98.60%	99.60%	96.55	0.0023494

Table 4. Second scenario testing results (dataset B).

Method	Accuracy	Sensitivity	Specificity	Time (s)	MSE
BP	83.33%	94.29%	6.67%	26.126	0.0009658
BP-NW	83.33%	95.24%	6.67%	16.295	0.0009507
BP-SS	84.17%	96.19%	6.67%	16.314	0.0009364
BP-NW-SS	85.00%	95.24%	13.33%	16.374	0.0009338

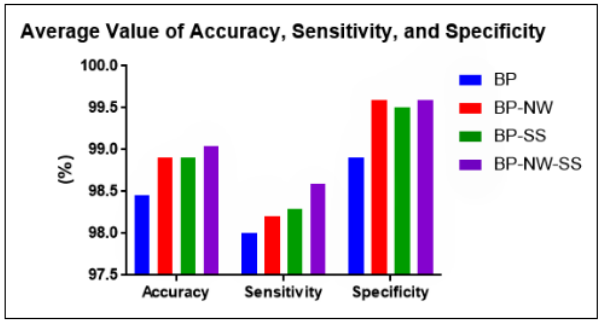

Fig. 4. Comparison of average value of accuracy, sensitivity, and specificity (dataset A).

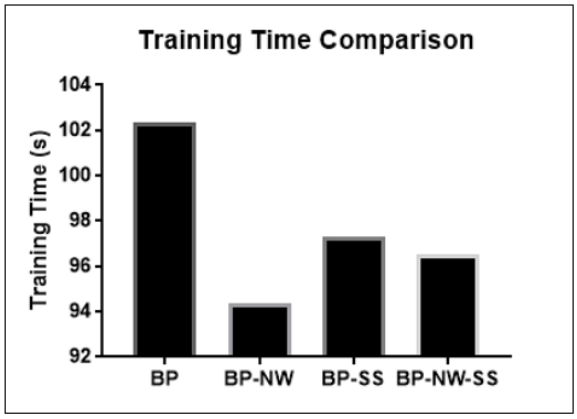

Fig. 5. Comparison of execution time for each method (dataset A).

For dataset B, which contained fewer data than dataset A, we tried to conduct an additional experiment without using K-Fold Cross-Validation. This experiment produced all accuracy, sensitivity, and specificity values of 100% and still the optimization algorithm Nguyen-Widrow and Stimulus-Sampling could reduce the training time.

The level of sensitivity and specificity of each model were then summarized in the form of a ROC (Receiver Operating Characteristic) Curve. The ROC curve can show the ability of a method to classify data into each class within a certain threshold range. Figures 6 and 7 visualize the ROC curve comparison. Each ROC curve was then compared using the AUC (Area Under the Curve) value. The AUC value was obtained by calculating the area under the ROC curve.

The higher the AUC value, the more the method can distinguish data between class one and other classes [43]. Based on the ROC curve in Fig. 6 for Dataset A, the AUC values of the BP, BP-NW, BP-SS, and BP-NW-SS methods were respectively 0.9850, 0.9875, 0.9890, and 0.9912. Then in Fig. 7 for Dataset B, the AUC values of the BP, BP-NW, BP-SS, and BP-NW-SS methods were respectively 0.6785, 0.7050, 0.7611, and 0.8570. The BP-NW-SS method produced the highest AUC value in both datasets A and B, which showed the best ability to classify MicroRNA profiles among the other methods.

A one-way ANOVA technique with posthoc Tukey and Wilcoxon signed-ranked test were chosen to compare the performance of BP-NW-SS with the standard BP statistically. The test showed the statistically significant difference based on the running time, Tukey HSD p-value = 0.0010053 (p-value < 0.05) and based on the accuracy, Tukey HSD p-value = 0.002066 (p-value < 0.05).

The Wilcoxon signed-rank test also indicated a significant difference that BP-NW-SS method produced better performance than standard BP algorithm with Asymp. Sig. (2-tailed) value was 0.005 (p-value < 0.05) based on running time and 0.026 (p-value < 0.05) based on accuracy values.

To ensure the effectiveness of the use of Nguyen-Widrow and Stimulus-Sampling, we conducted an additional experiment using WBCD Dataset as used in the previous research by Gorunescu and Belciug [26]. In this experiment, we did not use the same experimental parameters and setting, but the result of BP-SS was still following the previous study by Gorunescu and Belciug [26]. From Table 5, we can see that the BP-NW-SS still could produce the best performance and faster training time. Then in Table 6, the accuracy values were compared to Dataset A, but not Dataset B, since the predicted classes of WBCD dataset and Dataset A were the same. It can be seen that dataset A could produce better performances than WBCD dataset.

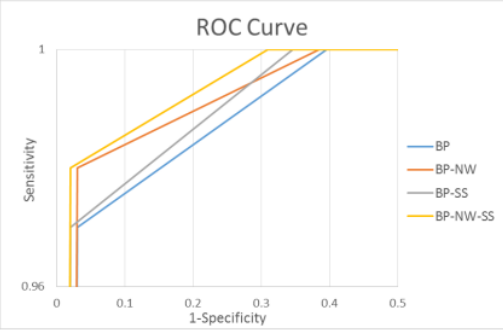

Fig. 6. ROC curves (dataset A).

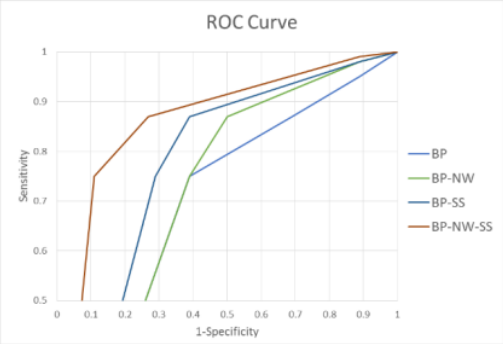

Fig. 7. ROC curves (dataset B).

Table 5. Testing results using WBCD dataset.

Method	Accuracy	Sensitivity	Specificity	Time (s)	MSE
BP	94.33%	94.09%	94.98%	192.94	0.0009989
BP-NW	95.04%	94.75%	95.33%	171.12	0.0009986
BP-SS	95.30%	94.81%	96.22%	173.03	0.0009985
BP-NW-SS	96.46%	95.28%	97.64%	172.53	0.0009983

Table 6. Comparison of accuracy value of WBCD dataset and dataset A.

Method	WBCD	Dataset A
BP	94.33%	98.45%
BP-NW	95.04%	98.90%
BP-SS	95.30%	98.90%
BP-NW-SS	96.46%	99.05%

In this study, there were two main focuses. Firstly, breast cancer classification is based on microRNA profile was conducted as a technique to perform early detection. Secondly, a classification model with better performance and faster training time were developed by implementing Nguyen-Widrow and the Stimulus-Sampling optimization algorithm to the standard backpropagation. Compared to the standard backpropagation, it could be seen that the Nguyen-Widrow algorithm and the Stimulus-Sampling algorithm proved to be able to increase the level of accuracy, sensitivity, and specificity and reduce the training time. This decrease in excess time showed that the use of optimization algorithms could accelerate the convergence during the training process. Another previous study by Aljarah et al. [44] that applied an optimization to multilayer perceptron was conducted in 2018, the optimizer was called Whale Optimization Algorithm (WOA) where some search agents were initialized. Each search agent represents a candidate of multilayer perceptron network, then the network with the smallest MSE was chosen.

This optimization was proved to be able to maintain the accuracy value with average of 97.31% for breast cancer dataset and speed up the convergence [44]. Differ from WOA, in this study, Nguyen-Widrow algorithm provided weights value initialization that was not too small or too large so that the convergence process could be enhanced. Meanwhile, the use of Stimulus-Sampling improved the performance of backpropagation by strengthening the output unit, which gave the smallest error value. Consequently, the error value kept decreasing in each epoch. The results were in accordance with the results of study by Gorunescu and Belciug and Belciug [26] where BP-SS could outperform BP with 95.3% accuracy, they were the first who proposed Stimulus-Sampling approach in their study. They didn't use other optimization algorithm to shorten the running time resulted in no different running time between the proposed method and 3-MLP.

The highest accuracy value, 99.05%, from this research was obtained using backpropagation method with Nguyen-Widrow and Stimulus-Sampling algorithm optimization. This surprisingly high value indicated both excellent classification method and dataset. The performance of the classification method had been compared. Then, to ensure the statement of the previous sentence, we conducted a feature selection over the original dataset of microRNA from the GDC Data Portal National Cancer Institute with 1881 features. Using SVM-RFE and Univariate Selection methods, some datasets with different feature combination was obtained. We also used an operator from RapidMiner, the correlation matrix, to obtain the best features set. These three methods were chosen because SVM-RFE, Univariate Selection, and Correlation had successfully used in gene selection, even univariate selection could outperform the multivariate approach [37, 38, 45]. Then the feature sets were classified using BP-NW-SS and the results are shown in Table 7. This

testing results used the same dataset (Dataset A), but with different features depending on the feature selection method used.

Table 7. Testing results using different feature sets.

Method	Features	Accuracy
Correlation matrix	hsa-mir-21, hsa-mir-200a, hsa-mir-7705, hsa-mir-141, hsa-mir-139, hsa-mir-10b	98.50%
SVM-RFE	hsa-mir-379, hsa-mir-625	71.50%
Univariate selection	hsa-let-7c, hsa-mir-100, hsa-mir-10b, hsa-mir-143, hsa-mir-148a, hsa-mir-182, hsa-mir-183, hsa-mir-21, hsa-mir-22, hsa-mir-375	98.33%

For each selection feature method, the experiment was executed 10 times to produce 10 features. Then, features that showed up in all 10 experiments were selected as the most significant features, the results are as in Table 7 column Features. Based on Table 7, feature set obtained from features correlation (RapidMiner tool) produced the highest accuracy value, 99.05%. This value still couldn't outperform the accuracy when using dataset A (96.46%) with features obtained from the research by Iorio et al. [8] that used in this research (i.e., hsa-miR-10b, hsa-miR-21, hsa-miR-125b-1, hsa-miR-125b-2, hsa-miR-145, and hsa-miR-155). Therefore, it was proved that the features of dataset A were very potential as biomarkers in breast cancer detection.

However, 98.50% was a high number too and the dataset could become an alternative. Compared to Dataset A, both of them have six features, even two of them are the same (i.e., hsa-mir-21 and hsa-mir-10b) and the other four are different (i.e., hsa-mir-200a, hsa-mir-7705, hsa-mir-141, and hsa-mir-139). Uhlmann et al. [46] in their study stated that hsa-mir-200 family had a tumor-suppressor function that regulated the viability, EGF-driven cell invasion, and cell cycle progression in breast cancer. It was also known that hsa-mir-200 regulated the expression of SIRT1 and EMT-like transformation in mammary apithelial cells in samples of breast cancer patient [47]. To our knowledge, the functionality of hsa-mir-7705 had never been reported to be related to breast cancer, but a study confirmed that it had association with bladder urothelial carcinoma patient survival [48]. Based on study by Neves et al. [49], hsa-mir-141 played an important role since it acted as metastasis suppressor genes and regulated by DNA methylation that suggested epigenetic regulation in aggressive breast cancer cell lines. Recent study stated that hsa-mir-139 involved in anti-metastatic and anti-oncogenic activity in human body, it also induced apoptosis, inhibited migration and invasion in breast cancer cells [50]. Eventually, these features indeed had significant role in the progression and prognosis of breast cancer.

As for the case of metastasis stage, the highest accuracy value was 85.00% using BP-NW-SS method and 48 MicroRNA features. Based on medical studies by Lan et al. [51], 4 features were selected from features of Dataset B (i.e., hsa-mir-196a, hsa-mir-205, hsa-mir-29b, hsa-mir-34b). We also used feature selection method to obtain some feature set combinations. These feature sets were classified using BP-NW-SS and the results are shown in Table 8. This testing results used the same dataset (dataset B), but with different features depended on the feature selection method used.

From Table 8, we can see that only accuracy value obtained using a feature set from medical studies by Neves et al. [49] could outperform the accuracy value of dataset B, 2.5% higher. While features set from RapidMiner correlation value, SVM-RFE, and Univariate Selection could not reach a higher value. Generally, a dataset with less features takes shorter of running time.

However, fewer or more features do not guarantee high accuracy since there might be some features with missing values, no related to the disease, or insignificant correlation. Therefore, datasets containing features with strong correlation and significant role to disease identification are more reliable.

Lan et al. [51] stated that only microRNA features with an aberrant expression that is closely associated with the pathogenesis and progression of human malignancies. Some features are specific to the cancer cells [52]. This indicated that the feature selection involving the medical aspect and proved to be associated with the certain disease could produce a better feature set.

Table 8. Testing results using different feature set for metastasis stage case.

Method	Features	Accuracy
Selected feature from Eades et al. [47]	hsa-mir-196a, hsa-mir-205, hsa-mir-29b, hsa-34b	87.50%
Correlation matrix	hsa-mir-214, hsa-mir-92b, hsa-mir-130a, hsa-196a, hsa-mir-345, hsa-181a	83.33%
SVM-RFE	hsa-mir-196a, hsa-mir-29c, hsa-mir-92	79.67%
Univariate selection	hsa-mir-122a, hsa-mir-130a, hsa-mir-181a, hsa-mir-181b, hsa-mir-192, hsa-mir-200a, hsa-mir-210, hsa-mir-214, hsa-mir-375, hsa-mir-99a	83.33%

4. Conclusions

The backpropagation method with Nguyen-Widrow and Stimulus-Sampling algorithm optimization showed the best performance with AUC value 0.9912 and 0.8570. This denoted that this model had excellent ability to classify breast cancer, both normal or breast class and before or after metastasis stage class, based on microRNA profiles. The features of hsa-miR-10b, hsa-miR-21, hsa-miR-125b-1, hsa-miR-125b-2, hsa-miR-145, and hsa-miR-155 proved to be very potential as a breast cancer biomarker, however, feature selection methods could be used to generate alternative feature set combination. The combination of Nguyen-Widrow and Stimulus-Sampling algorithm proved to be able to increase the value of accuracy, sensitivity, and specificity and reduce the execution time in the backpropagation method. Further research can be focused on extracting more information about each feature in the dataset and finding out whether the order of the six microRNA features when entering the input layer can affect the performance of the method used.

Acknowledgements

This work was supported by the DRPM-RISTEKDIKTI, The Republic of Indonesia for PASCA DOCTOR Research Grant 2018 under Grant No. 101-185/UN7.P4.3/PP/2018.

Nomenclatures

c	Counter variable
n	Number of input unit
m	Number of output unit
p	Number of hidden unit
s	Stimulus value given to output unit
t_k	Target value of output unit y_k
v_{ji}	Weight value from input unit x_i to hidden unit z_j
w_{kj}	Weight value from hidden unit z_j to output unit y_k
x_i	Input unit

Greek Symbols

β	Nguyen-Widrow scale factor
δ	Eerror factor

Abbreviations

ABC	Artificial Bee Colony
ACO	Ant Colony Optimization
ALO	Ant Lion Optimizer
AUC	Area Under the Curve
BP	Backpropagation
BP-NW	Backpropagation with Nguyen-Widrow algorithm optimization
BP-NW-SS	Backpropagation with Nguyen-Widrow and Stimulus-Sampling algorithm optimization
BP-SS	Backpropagation with Stimulus-Sampling algorithm optimization
EMT	Epithelial to Mesenchymal Transition
GA	Genetic Algorithm
GDC	Genomic Data Commons
K-NN	K-Nearest Neighbor
MLP	Multilayer Perceptron
MSE	Mean Squared Error
PSO	Particle Swarm Optimization
ROC	Receiver Operating Characteristic
SSO	Social Spider Optimization
SVM	Support Vector Machine
SVM-RFE	Support Vector Machine - Recursive Feature Elimination
WAO	Whale Optimization Algorithm
WBCD	Wisconsin Breast Cancer Diagnostic

References

1. DeSantis, C.E.; Fedewa, S.A.; Sauer, A.G.; Kramer, J.L.; Smith, R.A.; and Jemal, A. (2016). Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. *CA: A Cancer Journal for Clinicians*, 66(1), 31-42.
2. International Agency for Research on Cancer (IARC). (2018). Fact sheets by cancer. Retrieved November 8, 2018, from http://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet.pdf.

3. World Health Organization (WHO). (2014). Cancer country profiles. Retrieved January 8, 2018, from http://www.who.int/entity/cancer/country-profiles/idn_en.pdf.
4. Cardoso, F.; Harbeck, N.; Fallowfield, L.; Kyriakides, S.; and Senkus, E. (2012). Locally recurrent or metastatic breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, 23(S7), vii11-vii19.
5. Gadgil, A.; Sauvaget, C.; Roy, N.; Muwonge, R.; Kantharia, S.; Chakrabarty, A.; Bantwal, K.; Haldar, I.; and Sankaranarayanan, R. (2017). Cancer early detection program based on awareness and clinical breast examination: Interim results from an urban community in Mumbai, India. *The Breast*, 31, 85-89.
6. Weigelt, B.; Peterse, J.L.; and van't Veer, L.J. (2005). Breast cancer metastasis: Markers and models. *Nature Reviews Cancer*, 5(8), 591-602.
7. Pal, M.K.; Jaiswar, S.P.; Dwivedi, V.N.; Tripathi, A.K.; Dwivedi, A.; and Sankhwar, P. (2015). MicroRNA: A new and promising potential biomarker for diagnosis and prognosis of ovarian cancer. *Cancer Biology & Medicine*, 12(4), 328-341.
8. Iorio, M.V.; Ferracin, M.; Liu, C.G.; Veronese, A.; Spizzo, R.; Sabbioni, S.; Magri, E.; Pedriali, M.; Fabbri, M.; Campiglio, M.; Menard, S.; Palazzo, J.P.; Rosenberg, A.; Musiani, P.; Volinia, S.; Nenci, I.; Calin, G.A.; Querzoli, P.; Negrini, M.; and Croce, C.M. (2005). MicroRNA gene expression deregulation in human breast cancer. *Cancer Research*, 65(16), 7065-7070.
9. Khasburrahman, K.; Wibowo, A.; Waspada, I.; Hashim, H.B.; and Jatmiko, W. (2017). Comparison of diagnostics set and feature selection for breast cancer classification based on microRNA expression. *Proceedings of the 1st International Conference on Informatics and Computational Sciences (ICICoS)*. Semarang, Indonesia, 165-169.
10. Madhavan, D.; Cuk, K.; Burwinkel, B.; and Yang, R. (2013). Cancer diagnosis and prognosis decoded by blood-based circulating microRNA signatures. *Frontiers in Genetics*, 4, Article 116, 13 pages.
11. Yan, L.-X.; Huang, X.-F.; Shao, Q.; Huang, M.-Y.; Deng, L.; Wu, Q.-L.; Zeng, Y.-X.; and Shao, J.-Y. (2008). MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. *RNA*, 14(11), 2348-2360.
12. Jiang, L.; Zhang, J.; Xuan, P.; and Zou, Q. (2016). BP neural network could help improve pre-miRNA identification in various species. *BioMed Research International*, Article 9565689, 11 pages.
13. Paliwal, M.; and Kumar, U.A. (2009). Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications*, 36(1), 2-17.
14. Mert, A.; Kilic, N.; Bilgili, E.; and Akan, A. (2015). Breast cancer detection with reduced feature set. *Computational and Mathematical Methods in Medicine*, ID 265138, 11 pages.
15. Waspada, I.; Wibowo, A.; and Meraz, N.S. (2017). Supervised machine learning model for microrna expression data in cancer. *Jurnal Ilmu Komputer Dan Informasi*, 10(2), 108-115.
16. Abdel-Zaher, A.M.; and Eldeib, A.M. (2016). Breast cancer classification using deep belief networks. *Expert Systems with Applications*, 46, 139-144.

17. Anjarwati; Wibowo, A.; Adhy, S.; Kusumaningrum, R. (2018). MicroRNA expression profile selection for cancer staging classification using backpropagation. *Journal of Physics Conference Series,* 1025(1), 012090.
18. Saki, F.; Tahmasbi, A.; Soltanian-Zadeh, H.; and Shokouhi, S.B. (2013). Fast opposite weight learning rules with application in breast cancer diagnosis. *Computers in Biology and Medicine*, 43(1), 32-41.
19. Huang, H.-X.; Li, J.-C.; and Xiao, C.-L. (2015). A proposed iteration optimization approach integrating backpropagation neural network with genetic algorithm. *Expert Systems with Applications*, 42(1), 146-155.
20. Sebayang, A.H.; Masjuki, H.H.; Ong, H.C.; Dharma, S.; Silitonga, A.S.; Kusumo, F.; and Milano, J. (2017). Optimization of bioethanol production from sorghum grains using artificial neural networks integrated with ant colony. *Industrial Crops and Products*, 97, 146-155.
21. Zhang, Y.; Wang, S.; and Ji, G. (2015). A comprehensive survey on particle swarm optimization algorithm and its applications. *Mathematical Problems in Engineering*, Article ID 931256, 38 pages.
22. Mirjalili, S. (2015). The ant lion optimizer. *Advances in Engineering Software*, 83, 80-98.
23. Mirjalili, S.Z.; Saremi, S.; and Mirjalili, S.M. (2015). Designing evolutionary feedforward neural networks using social spider optimization algorithm. *Neural Computing and Applications*, 26(8), 1919-1928.
24. Karaboga, D.; Gorkemli, B.; Ozturk, C.; and Karaboga, N. (2014). A comprehensive survey: artificial bee colony (ABC) algorithm and applications. *Artificial Intelligence Review*, 42(1), 21-57.
25. Aisyah, S.; Harahap, M.; Siregar, A.M.H.; and Turnip, M. (2018). Optimization of training backpropagation algorithm using nguyen widrow for angina ludwig diagnosis. *Journal of Physics: Conference Series*, 1007(1), 012050.
26. Gorunescu, F.; and Belciug, S. (2016). Boosting backpropagation algorithm by stimulus-sampling: Application in computer-aided medical diagnosis. *Journal of Biomedical Informatics*, 63, 74-81.
27. Shaikhina, T.; Lowe, D.; Daga, S.; Briggs, D.; Higgins, R.; and Khovanova, N. (2015). Machine learning for predictive modelling based on small data in biomedical engineering. *IFAC-PapersOnLine*, 48(20), 469-474.
28. Nguyen, D.; and Widrow, B. (1990). Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. *Proceedings of the International Joint Conference on Neural Network*. San Diego, California, United States of America, 21-26.
29. Rosenfeld, N.; Aharonov, R.; Meiri, E.; Rosenwald, S.; Spector, Y.; Zepeniuk, M.; Benjamin, H.; Shabes, N.; Tabak, S.; Levy, A.; Lebanony, D.; Gorey, Y.; Silberschein, E.; Targan, N.; Ben-Ari, A.; Gilad, S.; Sion-Vardy, N.; Tobar, A.; Feinmesser, M.; Kharenko, O.; Nativ, O.; Nass, D.; Perelman, M.; Yosepovich, A.; Shalmon, B.; Polak-Charcon, S.; Fridman, E.; Avniel, A.; Bentwich, I.; Bentwich, Z.; Cohen, D.; Chajut, A.; and Barshack, I. (2008). MicroRNAs accurately identify cancer tissue origin. *Nature Biotechnology*, 26(4), 462-469.
30. Chandrashekar, D.S.; Bashel, B.; Balasubramanya, S.A.H.; Creighton, C.J.; Ponce-Rodriguez, I.; Chakravarthi, B.V.S.K.; and Varambally, S. (2017).

UALCAN: A portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia*, 19(8), 649-658.

31. Xu, J.; Zhao, J.; and Zhang, R. (2016). Four microRNAs signature for survival prognosis in colon cancer using TCGA data. *Scientific Reports*, 6, 8 pages.
32. Nahato, K.B.; Harichandran, K.N.; and Arputharaj, K. (2015). Knowledge mining from clinical datasets using rough sets and backpropagation neural network. *Computational and Mathematical Methods in Medicine*, Article ID 460189, 13 pages.
33. Estes, W.K. (1950). Toward a statistical theory of learning. *Psychological Review*, 57(2), 94-107
34. Pohjalainen, J.; Räsänen, O.; and Kadioglu, S. (2015). Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits. *Computer Speech & Language*, 29(1), 145-171.
35. Amarnath, B.; and Balamurugan, S.A.A. (2016). Review on feature selection techniques and its impact for effective data classification using uci machine learning repository dataset. *Journal of Engineering Science and Technology (JESTEC)*, 11(11), 1639-1646.
36. Nagpal, A.; and Singh, V. (2018). Identification of significant features using random forest for high dimensional microarray data. *Journal of Engineering Science and Technology (JESTEC)*, 13(8), 2446-2463.
37. Yan, K.; and Zhang, D. (2015). Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sensors and Actuators B: Chemical*, 212, 353-363.
38. Guyon, I.; Weston, J.; Barnhill, S.; and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1-3), 389-422.
39. Lai, C.; Reinders, M.J.T.; van't Veer, L.J.; and Wessels, L.F.A. (2006). A comparison of univariate and multivariate gene selection techniques for classification of cancer datasets. *BMC Bioinformatics*, 7, 10 pages.
40. Malik, H.; and Mishra, S. (2014). Feature selection using RapidMiner and classification through probabilistic neural network for fault diagnostics of power transformer. *Proceedings of the Annual IEEE India Conference (INDICON)*. Pune, India, 1-6.
41. Reif, M.; Shafait, F.; Goldstein, M.; Breuel, T.; and Dengel, A. (2014). Automatic classifier selection for non-experts. *Pattern Analysis and Applications*, 17(1), 83-96.
42. RapidMiner. (2018). Correlation matrix. Retrieved December 3, 2018, from https://docs.rapidminer.com/8.0/studio/operators/modeling/correlations/correlation_matrix.html.
43. Sokolova, M.; and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427-437.
44. Aljarah, I.; Faris, H.; and Mirjalili, S. (2018). Optimizing connection weights in neural networks using the whale optimization algorithm. *Soft Computing*, 22(1), 1-15.

45. Zuber, V.; and Strimmer, K. (2009). Gene ranking and biomarker discovery under correlation. *Bioinformatics*, 25(20), 2700-2707.

46. Uhlmann, S.; Zhang, J.D.; Schwäger, A.; Mannsperger, H.; Riazalhosseini, Y.; Burmester, S.; Ward, A.; Korf, U.; Wiemann, S.; and Sahin, Ö. (2010). miR-200bc/429 cluster targets PLCgamma1 and differentially regulates proliferation and EGF-driven invasion than miR-200a/141 in breast cancer. *Oncogene*, 29(30), 4297-4306.

47. Wades, G.; Yao, Y.; Yang, M.; Zhang, Y.; Chumsri, S.; and Zhou, Q. (2011). miR-200a regulates SIRT1 expression and epithelial to mesenchymal transition (EMT)-like transformation in mammary epithelial cells. *Journal of Biological Chemistry*, 286(29), 25992-26002.

48. Hu, Y.; Cheng, C.; Hong, Z.; and Shi, Z. (2017). Independent prognostic miRNAs for bladder urothelial carcinoma. *Oncology Letters*, 14(3), 3001-3005.

49. Neves, R.; Scheel, C.; Weinhold, S.; Honisch, E.; Iwaniuk, K.M.; Trompeter, H.I.; Niederacher, D.; Wernet, P.; Santourlidis, S.; and Uhrberg, M. (2010). Role of DNA methylation in miR-200c/141 cluster silencing in invasive breast cancer cells. *BMC Research Notes*, 3, 7 pages.

50. Zhang, H.-d.; Sun, D.-w.; Mao, L.; Zhang, J.; Jiang, L.-h.; Li, J.; Wu, Y.; Ji, H.; Chen, W.; Wang, J.; Ma, R.; Cao, H.–x.; Wu, J.-z.; and Tang, J.-h. (2015). MiR-139-5p inhibits the biological function of breast cancer cells by targeting Notch1 and mediates chemosensitivity to docetaxel. *Biochemical and Biophysical Research Communications*, 465(4), 702-713.

51. Lan, H.; Lu, H.; Wang, X.; and Jin, H. (2015). MicroRNAs as potential biomarkers in cancer: opportunities and challenges. *BioMed Research International*, Article ID. 125094, 17 pages.

52. Pigati, L.; Yaddanapudi, S.C.S.; Iyengar, R.; Kim, D.-J.; Hearn, S.A.; Danforth, D.; Hastings, M.L.; and Duelli, D.M. (2010). Selective release of microRNA species from normal and malignant mammary epithelial cells. *PloS One*, 5(10), 13 pages.

C3

ORIGINALITY REPORT

12%	%	12%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

1%

★ Gabriel Eades, Yuan Yao, Muhua Yang, Yongshu Zhang, Saranya Chumsri, Qun Zhou. "miR-200a Regulates SIRT1 Expression and Epithelial to Mesenchymal Transition (EMT)-like Transformation in Mammary Epithelial Cells", Journal of Biological Chemistry, 2011

Publication

Exclude quotes	Off	Exclude matches	< 1%
Exclude bibliography	Off		