

# Implementation of Modified Backpropagation with Conjugate Gradient as Microarray Data Classifier with Binary Particle Swarm Optimization as Feature Selection for Cancer Detection

Muhammad Naufal Mukhbit Amrullah<sup>[1]\*</sup>, Adiwijaya<sup>[2]</sup>, Widi Astuti<sup>[3]</sup>

Bachelor of Informatics, Faculty of Informatics<sup>[1], [2], [3]</sup>

Telkom University

Bandung, Indonesia

nmukhbit@student.telkomuniversity.ac.id<sup>[1]</sup>, adiwijaya@telkomuniversity.ac.id<sup>[2]</sup>, astutiwidi@telkomuniversity.ac.id<sup>[3]</sup>

**Abstract**— Cancer is one of the deadliest diseases in the world that needs to be handled as early as possible. One of the methods to detect the presence of cancer cells early on is by using microarray data. Microarray data can store human gene expression and use it to classify cancer cells. But one of the challenges of using microarray is its vast number of features, not proportional to its small number of samples. To resolve that problem, dimensionality reduction is needed to reduce the number of features stored in microarray data. Binary Particle Swarm Optimization (BPSO) is one of the methods to reduce dimensionality of microarray data that can increase classification performance. Although when combined with Backpropagation, BPSO still shows a relatively low performance. In this research, Modified Backpropagation with Conjugate Gradient is used to classify data that has been reduced with BPSO. The average accuracy result of BPSO+CGBP is 86.1%, giving it an improvement compared to BPSO+BP which averaged to 80.8%.

**Keywords**— cancer, microarray, binary particle swarm optimization, backpropagation, conjugate gradient

## I. INTRODUCTION

Cancer is one of the deadliest diseases in the world that responsible for the loss of 9.6 million lives in 2018. Until present time, there is no method of treatment that can effectively cure cancer. However, if cancer can be detected early, patients with cancer can be treated to prevent cancer deaths [1].

One technology that can detect the presence of cancer cells early on is microarray data. Microarray data collects and stores human gene expressions in an array of float numbers [2]. The data then can be used to classify the genes whether it is cancerous or not [3]. With early detection of cancer tissue, infected patients can be treated more quickly and can reduce the chance of the cancerous tissue to spread.

The nature of microarray data is its large number of features, which can reach tens of thousands per sample, with only a small number of samples. Not all of these features have relevant information for the classification process [4]. Therefore,

dimensionality reduction is needed to reduce the number of features in microarray data. One method of dimensionality reduction algorithm that can be used is Binary Particle Swarm Optimization (BPSO). In previous studies, BPSO was proven to be able to produce accuracy that was almost twice greater than Information Gain, which is 99.5% compared to 54.5% [5]. But BPSO still gives a pretty bad result when combined with Backpropagation, with an accuracy of 44% when used for cancer data classification [6].

Based on a research conducted in 2015, Modified Backpropagation with Conjugate Gradient can improve accuracy in cancer data when compared to classification using classic Backpropagation [7]. Therefore, this study will implement Modified Backpropagation with Conjugate Gradient to classify the data that has been reduced with BPSO and compare the result with classic Backpropagation.

## II. LITERATURE REVIEW

### A. Microarray Data

The growth of cancer in the cells can cause genetic aberrations. One widely studied technique to detect those aberrations are microarray data, as it can perform rapid genetic analysis [8]. Microarray data is a technology that can collect and store thousands of extracted human gene expressions in an array of numbers at the same time [9]. Each data in microarray is stored as float numbers that represent the composition of gene expression in the human body [2]. The data then can be compared on genomic-scale with other samples to search biological relevance such as cancer [8].

The data is observed with DNA Microarray, a microscope slide that stores the extracted mRNA from human cells. The genes will be arranged in specific positions on the glass plate and will be reacted with cDNA which has been labeled with fluorescent dye. The results of this reaction will produce bright colors which can be translated into the data that stored in microarray [10].

### B. Binary Particle Swarm Optimization

After the data has gone through the pre-processing stage, the number of features in the data will be reduced with Binary Particle Swarm Optimization (BPSO). Binary Particle Swarm Optimization (BPSO) is a modified version of the original Particle Swarm Optimization (PSO) which works by imitating the movement of a flock of birds or fish school. In PSO, a particle is defined as a part of the population that represents a solution to the current problem. The movement of birds will be simulated on the population until a solution that fulfills the satisfying condition is found [11]. Each particle will move based on velocity value that will be updated based on the particle that has the best fitness.

Many problems occur in discrete space, and that encouraged Kennedy and Eberhart (1997) to modify PSO into a binary form [12]. In binary space, particles will define their change in coordinates by turning the bits in the particles on or off. Velocity will be defined as the probability that a bit will get a value of 1. In PSO, velocity is defined as follows [13]:

$$v_{id}^{new} = w \times v_{id}^{old} + c_1 r_1 (pb_{id}^{old} - x_{id}^{old}) + c_2 r_2 (gb_d^{old} - x_{id}^{old}) \quad (1)$$

where  $r_1$  and  $r_2$  are random numbers between 0 or 1,  $c_1$  is the cognitive learning factor,  $c_2$  is the social learning factor, and  $w$  is the inertia weight. Velocity and position of the  $i$ -th particle in the search space dimension  $d$  are defined as  $v_{id}$  and  $x_{id}$ , respectively. The previous best position of the  $i$ -th particle in the search space dimension  $d$  is  $pb_{id}$ , while  $gb_d$  is the best position in the search space dimension  $d$  [11, 14].

In BPSO, velocity is updated in the same fashion as PSO. However, because the values of  $x_{id}$ ,  $pb$ , and  $gb$  can only be 0 or 1, velocity is defined as the probability of the bits in the position vector to get a value of 1. This means that the velocity must be limited to the range [0.0, 1.0], which can be obtained by applying the sigmoid function that can modify the velocity value. The sigmoid function can be formulated as follows:

$$S(v_{id}) = \frac{1}{1+e^{-v_{id}}} \quad (2)$$

Once the velocity is found, the  $x_{id}$  position can be updated with the following rule:

$$x_{id}^{new} = \begin{cases} 1, & S(v_{id}^{new}) > rand() \\ 0, & S(v_{id}^{new}) \leq rand() \end{cases} \quad (3)$$

where  $rand()$  is a function that generates a random value in the range [0.0, 1.0] and  $S(v_{id}^{new})$  is the sigmoid function in equation (2) [9, 11].

### C. Backpropagation Neural Network

Backpropagation is an algorithm to train Artificial Neural Networks (ANN). Backpropagation is used as a method to update the weights and biases found in each neuron in the network. There are three main stages for training ANN with Backpropagation, namely forward propagation, backward propagation, and weight update [15]. The forward propagation stage is used to calculate the predicted output, backward propagation is used to calculate the loss of output to find the

gradient of each weight and bias, and the weight update is used to update the weight and bias based on the gradient. Loss will be calculated with Mean Squared Error which defined as follows:

$$f(w) = \frac{1}{NJ} \sum_{n=1}^N \sum_{j=1}^J (t_{nj} - y_{nj}(w))^2 \quad (4)$$

where  $t_{nj}$  is the output target and  $y_{nj}(w)$  is the predicted output from  $n$ -th data and  $j$ -th class, respectively.  $N$  is defined as the number of data and  $J$  is defined as the number of classes [7, 16].

### D. Modified Backpropagation with Conjugate Gradient

Conjugate Gradient (CG) is an optimization method that aims to minimize a function. CG works by searching in conjugate directions whose value is orthogonal [17]. Two nonzero vectors ( $d_i$  and  $d_j$ ) can be said to be orthogonal if the inner product is 0, which can be formulated as follows [7]:

$$d_i^T d_j = 0 \quad (5)$$

Because CG searches in orthogonal directions, CG can converge to solutions quickly [7]. This allows CG to train neural networks faster and requires less memory than classic Backpropagation [16]. CG will minimize the objective function in equation (4) by calculating the weight which will be updated by:

$$w_{k+1} = w_k + \alpha_k d_k \quad (6)$$

$$d_{k+1} = -g_{k+1} + \beta_k d_k \quad (7)$$

where  $\alpha$  and  $\beta$  are momentum parameters to avoid the solution being stuck at local optimum. The parameter  $\alpha$  can be searched using the line search technique [17]. Meanwhile, the parameter  $\beta$  can be found using one of the following methods:

a) Powell-Beale:

$$\beta_{k+1} = \frac{g_{k+1}^T (g_{k+1} - g_k)}{d_k^T (g_{k+1} - g_k)} \quad (8)$$

b) Fletcher-Reeves:

$$\beta_{k+1} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} \quad (9)$$

c) Polak-Ribiere:

$$\beta_{k+1} = \frac{g_{k+1}^T (g_{k+1} - g_k)}{g_k^T g_k} \quad (10)$$

where  $\beta_{k+1}$  and  $g_{k+1}$  are the values of  $\beta$  and  $g$  in the current iteration, while  $g_k$  is  $g$  in the previous iteration.

The Conjugate Gradient Algorithm for training Backpropagation Neural Networks can be explained as follows [7]:

1. Network initiation, which consists of network architecture, initiation of weight and bias values with small random numbers, and determination of termination conditions
2. While the termination conditions are not met, perform steps 2-12

Forward Propagation:

3. Calculate the output of hidden layer  $y_j$  ( $j=1,2,\dots,p$ ).

$$y\_net_j = v_{j0} + \sum_{i=1}^n x_i v_{ji} \quad (11)$$

$$y_j = f(y\_net_j) = \frac{1}{1+e^{-y\_net_j}} \quad (12)$$

4. Calculate the output of output layer  $z_k$  ( $k=1,2,\dots,m$ ).

$$z\_net_k = w_{k0} + \sum_{j=1}^p y_j w_{kj} \quad (13)$$

$$z_k = f(z\_net_k) = \frac{1}{1+e^{-z\_net_k}} \quad (14)$$

Backward Propagation:

5. Calculate the error factor of the output layer's output based on the error of the actual and predicted results.

$$\delta_k = (t_k - z_k) f'(z\_net_k) = (t_k - z_k) z_k (1 - z_k) \quad (15)$$

6. Calculate the error factor of the hidden layer's output based on the error of the previous layers.

$$\delta\_net_j = \sum_{k=1}^m \delta_k w_{kj} \quad (16)$$

$$\delta_j = \delta\_net_j f'(y\_net_j) = \delta\_net_j y_j (1 - y_j) \quad (17)$$

7. Calculate the gradient on the output unit.

$$g_{k+1} = \frac{1}{N} \sum_{n=1}^m \delta_{nk} z_{nk} \quad (18)$$

8. Calculate the gradient on the hidden unit.

$$g_{j+1} = \frac{1}{N} \sum_{n=1}^p \delta_{nj} z_{nj} \quad (19)$$

9. Calculate the parameter  $\beta$  for all the neurons in the hidden and output unit with equation (8), (9), and (10).

10. Calculate the direction for all neurons in the hidden and output unit.

$$d_{t+1} = -g_{t+1} + \beta_t d_t \quad (20)$$

For initial direction:

$$d_1 = -g_1 \quad (21)$$

11. Calculate the parameter  $\alpha$  for all neurons in the hidden and output unit.

Weight update:

12. Weight will be updated with the following formula:

$$w_{t+1} = w_t + \alpha_{t+1} d_{t+1} \quad (22)$$

### III. RESEARCH METHOD

#### A. System Overview

This research aims to build a system with classic Backpropagation and Modified Backpropagation with Conjugate Gradient as classifier method and Binary Particle Swarm Optimization (BPSO) as dimensionality reduction method. Both classification results will be compared and analyzed.

The classification system will be built in several stages. The first stage is pre-processing, where microarray data will be normalized into intervals [0.0, 1.0]. After the data has been normalized, then the data will be divided into training data and testing data with K-Fold Cross Validation. The next stage is dimensionality reduction that will be done using the BPSO method. After the features in the microarray data are reduced, the data will be used to train the model using two methods, namely classic Backpropagation and Modified Backpropagation with Conjugate Gradient. After the model is trained, the model will be used to classify the test data. The accuracy of the two classification results will be calculated and will be compared with one another. The flowchart of the system can be seen on Figure 1.

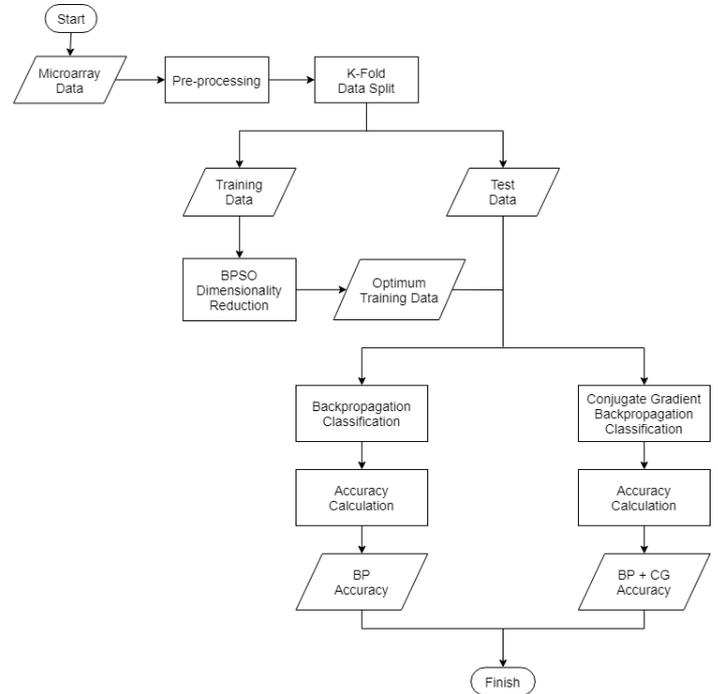


Fig. 1. System Flowchart

#### B. Dataset

This research will use five cancer datasets that originates from Kent-Ridge Biomedical Data Set Repository, namely Breast Cancer, Colon Tumor, Lung Cancer, Ovarian Cancer, and Prostate Cancer [18]. The specifications of the datasets can be seen on Table I.

TABLE I. DATASET SPECIFICATION

Data	Feature Size	Sample Size	Number of Classes
Breast Cancer	24481	97	2
Colon Tumor	2000	62	2
Lung Cancer	12533	181	2
Ovarian Cancer	15154	253	2
Prostate Cancer	12600	136	2

TABLE II. BPSO PARAMETER TESTING CONDITIONS

	Cognitive Learning ( $c_1$ )	Social Learning ( $c_2$ )	Inertia Weight ( $w$ )
Combination 1	1	1	0.1
Combination 2	1	2	0.1
Combination 3	1	1	0.5
Combination 4	1	2	0.5
Combination 5	2	1	0.1
Combination 6	2	2	0.1
Combination 7	2	1	0.5
Combination 8	2	2	0.5

C. Pre-processing

The pre-processing stage is the stage to refine the data that will be used for classification. In this stage, the data will go through a normalization process. Normalization will map data from its original form to data in a range such as [0.0, 1.0] to ensure that the data being tested can be compared with one another [19].

The method that will be used to normalize the data is Min Max Normalization. Min Max Normalization will perform linear normalization on the data. Therefore, the relationship between data will remain the same as the original data [20]. To perform Min Max Normalization, the data must be converted with the following formula:

$$v' = \frac{v - \min_A}{\max_A - \min_A} \quad (23)$$

where  $v$  is the currently observed value and  $A$  is an attribute with the range of  $[\min_A, \max_A]$  that will be mapped to a new range of  $[\text{new\_min}_A, \text{new\_max}_A]$  [21].

D. Data Split with K-Fold Cross Validation

To divide the data into training data and test data, the K-Fold Cross Validation method will be used. This method will divide the data into k-parts, each of which will function as training and test data. For each subset in k, the data will act as test data, while the rest of the data will act as training data.

In this study, the k value used is k=5, meaning that the data will be divided into five subset parts, each of which will be training data and test data. The ratio of training data to test data for each i in k is 4:1, or 80%:20%.

E. Feature Selection with Binary Particle Swarm Optimization

The first stage in feature selection with BPSO is the determination of parameter values. The population size is tested with several values, namely 10, 20, 30, and 50, so that the population can cover small, medium, and large numbers [22]. Other parameters that must be initiated are cognitive learning ( $c_1$ ), social learning ( $c_2$ ), and inertia weight ( $w$ ). There are eight test combinations to test the value of cognitive learning ( $c_1$ ), social learning ( $c_2$ ), and inertia weight ( $w$ ), which can be seen in Table II.

In BPSO, the particles will seek direction based on the particle with the best fitness in the population. A particle is considered to have the best fitness when the particle has the smallest cost value. Since the purpose of feature selection is to improve classification performance, the best fitness can be obtained from particles with the smallest error rate and number of features. The cost function will be adopted from the research conducted by Vieira et al. in 2013 that adjusted for minimization process [23, 24]:

$$f(u_i) = \alpha(1 - P) + (1 - \alpha) \left( \frac{N_f}{N_t} \right) \quad (24)$$

where  $P$  is the performance value of the classification model,  $N_f$  is the number of selected features on the particle, and  $N_t$  is the number of original features. The constant  $\alpha$  is the weight value that determines the importance of  $P$  and  $N_f$ .

F. Classification with Classic Backpropagation

The first step in classic Backpropagation is to initiate a network that includes three main layers, namely the input layer, hidden layer, and output layer. Each layer has a fixed number of neurons, each of which has a weight and bias values. The weight and bias values in the network will be initiated with a random number. The number of neurons in the input layer is the number of features that the data has, while the number of neurons in the output layer is the number of classes in the data. The number of neurons in the hidden layer was tested with a value range of 5-100 with intervals of 5, to determine the effect of the number of neurons on the accuracy and the optimal number of neurons.

Tests will also be conducted on the number of hidden layers in the network. The number of hidden layers will be tested with a value of 1 to 10, with an interval of 1. The number of neurons that will be used in this test refers to the best results obtained in the previous test.

During the training process, each neuron will update its weight and bias based on the gradient and learning rate. The learning rate values will also be tested, with values of 0.05, 0.1, and 0.5 to determine the amount of learning rate that could

produce the best accuracy.

G. Classification with Conjugate Gradient Backpropagation

The Conjugate Gradient method changes the way Backpropagation updates the weight and bias values. The learning rate constant is no longer required for Conjugate Gradient Backpropagation. Instead, the learning rate constant is substituted with  $\alpha$  and  $\beta$  which will be used to find the direction value.

The main stages performed on Conjugate Gradient Backpropagation is the same as classic Backpropagation, beginning with defining the network. The number of neurons in the hidden layer will be tested with the same testing conditions as classic Backpropagation, with a value range of 5-100 with an interval of 5. To update the weight, three  $\beta$  search methods are used, namely Powell-Beale, Fletcher-Reeves, and Polak-Ribiere, to find out the results of each method.

H. Accuracy Calculation

The test data classification results will be compared to the actual class of the data to get the accuracy value. Accuracy will be calculated with confusion matrix that represented in Table III.

TABLE III. CONFUSION MATRIX

		Predicted class	
		Class = true	Class = false
Actual class	Class = true	True Positive	False Negative
	Class = false	False Positive	True Negative

The table above has four different terms in four different cells. The term that starts with true means the model successfully predicted the results of the classification according

to the original class. The label will be true positive (TP) when the original class is cancer positive and true negative (TN) when the original class is cancer negative. Conversely, when a label starts with false, it means that the model cannot successfully predict the original class in the microarray data. False positive (FP) occurs when the original class is cancer positive and false negative (FN) occurs when the original class is cancer negative. The accuracy value will be obtained by:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{25}$$

IV. RESULT AND DISCUSSION

Tests conducted in this study aim to find the best parameter values for each method and to compare the classification results obtained from classic Backpropagation method and Modified Backpropagation with Conjugate Gradient method. Both scenarios will be carried out by applying the Binary Particle Swarm Optimization feature selection. The data that will be used for classification can be seen in Table I.

A. The effect of the number of neurons

In this testing stage, testing is based on the network architecture. Tests were conducted on the number of neurons in the hidden layer to determine the effect of number of neurons on the performance of the classification results. The network was tested with 5-100 neurons with intervals of 5 for classification with classic Backpropagation and Modified Backpropagation with Conjugate Gradient. The data tested below have not been feature-selected. The training will stop when the difference in cost value in the last five iterations is less than 0.0001 with a maximum of 1000 iterations.

Based on the test results in Figure 2, the accuracy does not show any pattern of change from the 5-100 neurons experiment. The accuracy values tend to be in the same range but with a

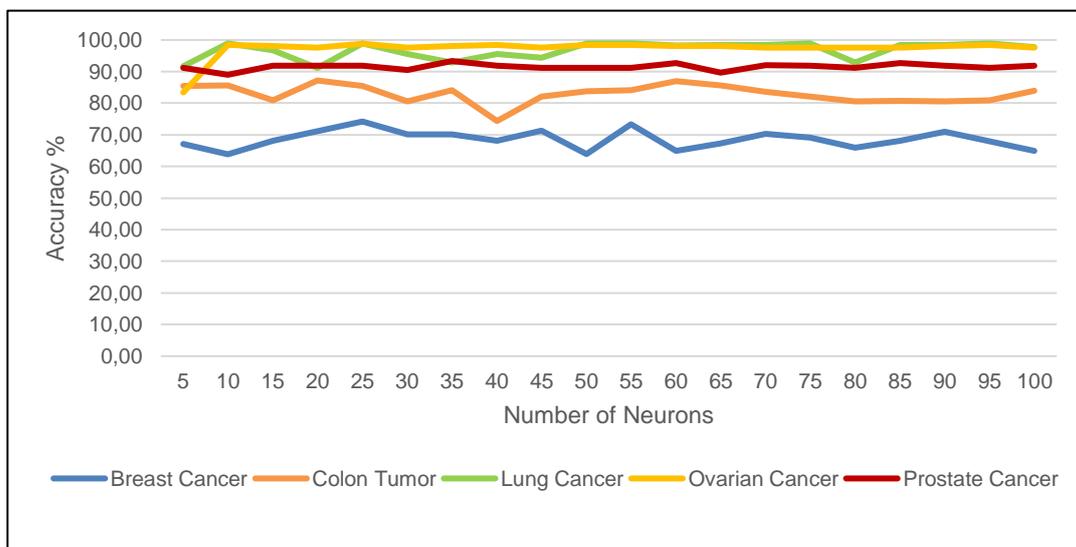


Fig. 2. The Effect of Number of Neurons to Accuracy on Classic Backpropagation

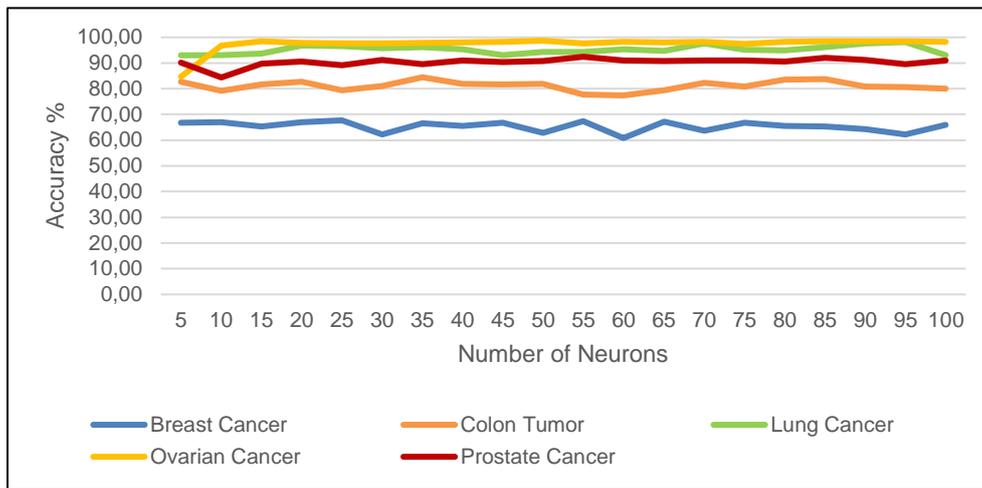


Fig. 3. The Effect of Number of Neurons to Accuracy on Conjugate Gradient Backpropagation

little fluctuation. The most stable results were obtained in the Ovarian Cancer and Prostate Cancer data, with an average of 96.91% and 91.52%, respectively. The Prostate Cancer data appeared to be the most stable in the test, while the Ovarian Cancer data appeared to have the accuracy increased when the number of neuron is 10 and remained stable in the rest of the tests.

A fluctuation happened in the results of Breast Cancer, Colon Tumor, and Lung Cancer data. Breast Cancer data experienced the greatest decrease in accuracy when the number of neurons was 60 with a decrease of 8.42% from the 73.26% accuracy obtained when the neurons were 55. In the Colon Tumor data, the greatest decrease in accuracy was obtained when the number of neurons was 40, with an accuracy value of 74.36%. This value is the smallest value out of all number of neurons tests in the Colon Tumor data, which on average always get an accuracy above 80%. Lung Cancer data showed a drastic decrease when there were 20 and 80 neurons, with the largest decrease occurred when the number of neurons was 80, which got an accuracy of 92.79% from 98.89% when the

number of neurons was 75.

The test results with Conjugate Gradient Backpropagation method shown in Figure 3 show more stable results than the test using classic Backpropagation. Colon Tumor and Lung Cancer data that fluctuate in the previous test look much more stable on the test with Conjugate Gradient Backpropagation. Fluctuation still slightly occurs in Breast Cancer data, but it happens in a low range, with the smallest value of 60.81% when the neurons are 60 and the largest value is 67.7% when the neurons are 25. The Ovarian Cancer data has something in common with previous tests, that they both experienced a drastic increase when the number of neurons was 10.

*B. The effect of the number of hidden layers*

In this test, the classification is tested based on the number of hidden layers in the network. This test is done 10 times, with the number of hidden layers tested is 1 to 10. The test is carried out using classic Backpropagation and Conjugate Gradient Backpropagation. Classification will stop when the difference

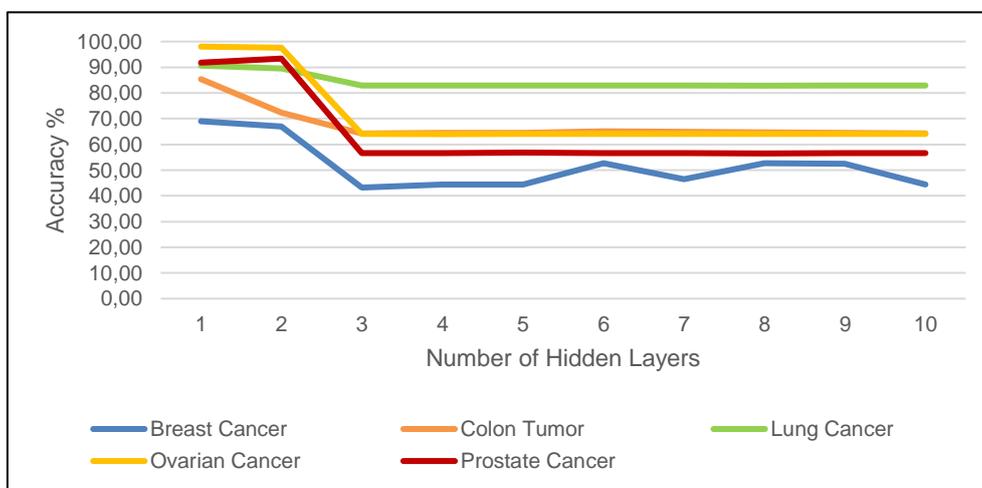


Fig. 4. The Effect of Number of Hidden Layers to Accuracy on Classic Backpropagation

in cost value in the last five iterations is less than 0.0001 with a contrast to the previous test which decreased when the hidden

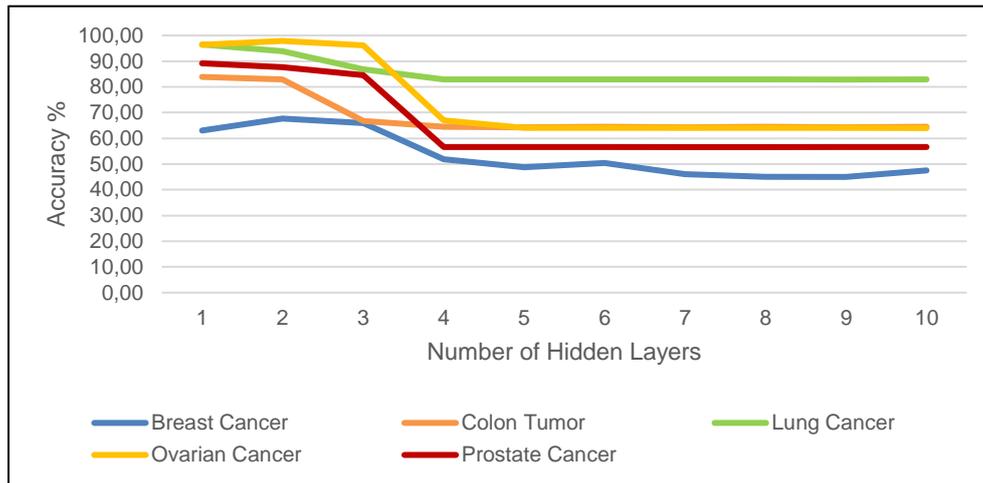


Fig. 5. The Effect of Number of Hidden Layers to Accuracy on Conjugate Gradient Backpropagation

maximum of 1000 iterations.

In the test results carried out with classic Backpropagation which can be seen in Figure 4, one similarity occurs in all data, that the largest average accuracy obtained when the number of hidden layer is 1 and experiencing a fairly drastic decrease in accuracy when the number of hidden layers is 3. The most drastic decrease occurred in Ovarian Cancer and Prostate Cancer data, where Ovarian Cancer data decreased from 97.63% to 64.02% and Prostate Cancer data decreased from 93.39% to 56.61%. Almost all data get flat result when the number of hidden layers is greater than 3, except for Breast Cancer data which has a slight increase when the hidden layers are 6, 8, and 9.

The test results using Conjugate Gradient Backpropagation can be seen in Figure 5. As can be seen from the figure, the test results are not much different from the test results using classic Backpropagation. There is a difference in Ovarian Cancer and Prostate Cancer data, which in this test has the accuracy decreased drastically when the hidden layer numbered 4, in

layer was 3. The greatest average was obtained when the number of hidden layers is 2, with an average of 86.02%, in contrast to the previous test which got the best average when the number of hidden layers is 1.

C. The effect of BPSO parameters

In this stage, the parameters at BPSO are tested to measure their performance of producing the highest accuracy. In the first test, the values tested were cognitive learning ( $c_1$ ), social learning ( $c_2$ ), and inertia weight ( $w$ ). The tested values of  $c_1$  and  $c_2$  is 1 and 2, by combining these values so that all conditions can be tested. While the  $w$  values tested are 0.1 and 0.5.

In the test results in Table IV, the accuracy tends to be higher when the cognitive learning parameter ( $c_1$ ) is 1. This can be seen in the average accuracy which shows a greater value when  $c_1 = 1$ . The average accuracy value when  $c_1 = 1$  is 83.49%, while the average accuracy value when  $c_1 = 2$  is only 81.49%.

TABLE IV. PARAMETER TESTING ON BPSO

Cognitive Learning ( $c_1$ )	Social Learning ( $c_2$ )	Inertia Weight ( $w$ )	Accuracy (%) on Data					Average (%)
			Breast Cancer	Colon Tumor	Lung Cancer	Ovarian Cancer	Prostate Cancer	
1	1	0.1	56,68	83,97	97,78	98,01	81,64	83,62
1	2	0.1	64	85,64	98,89	98,42	83,78	86,15
1	1	0.5	58,63	66,28	98,34	98,01	88,28	81,9
1	2	0.5	62	64,74	97,78	98,03	89,02	82,31
2	1	0.1	54,32	66,79	97,22	97,23	88,28	80,77
2	2	0.1	61,63	63,72	98,33	98,43	83,15	81,05
2	1	0.5	56,58	67,69	97,78	98,82	87,46	81,67
2	2	0.5	66,89	66,15	98,89	98,01	82,43	82,47

TABLE V. POPULATION SIZE TESTING ON BPSO

Population Size	Accuracy (%) on Data					Average (%)
	Breast Cancer	Colon Tumor	Lung Cancer	Ovarian Cancer	Prostate Cancer	
10	68,05	66,41	98,89	97,64	85,98	83,39
20	67,11	67,69	97,78	97,65	86,77	83,4
30	62	64,74	97,78	98,03	89,02	82,31
50	67,79	65,9	97,79	98,42	89,76	83,93

The accuracy value also shows a fairly large number when the social learning ( $c_2$ ) is 2. This can be seen in the average accuracy which always has a greater value when  $c_2$  is 2 compared to when  $c_2$  is 1. When averaged, the accuracy with  $c_2 = 2$  is 82.99%, while the accuracy when  $c_2 = 1$  is only 81.99%.

The same thing can also be found on Breast Cancer, Lung Cancer, and Ovarian Cancer tests. In Breast Cancer data, the accuracy when  $c_2 = 1$  is always less than 60%, while when  $c_2 = 2$ , the accuracy can reach more than 60%. If averaged, the accuracy when  $c_2 = 1$  is 56.55%, while when  $c_2 = 2$  the average accuracy increases to 63.63%. However, in Lung Cancer and Ovarian Cancer data, this does not occur in every case. For example, in Lung Cancer data, the accuracy when  $c_2 = 1$  is greater than the accuracy when  $c_2 = 2$ , when  $c_1$  and  $w$  are 1 and 0.5, respectively. In Ovarian Cancer data, this pattern does not apply when the values of  $c_1 = 2$  and  $w = 0.5$ .

The difference of inertia weight ( $w$ ) does not show a significant change in the results' accuracy. The average accuracy when  $w = 0.1$  is 82.89%. This value is not too much of a difference from the average accuracy when  $w = 0.1$ , which averaged at 82.09%. However, the accuracy is seen to be higher in most cases when  $w = 0.1$  is combined with a cognitive learning of 1. This can be seen in the Breast Cancer, Colon Tumor, and Prostate Cancer data, where the accuracy with  $w = 0.1$  combined with  $c_1 = 1$  always show better results than the accuracy of  $w = 0.5$ . In contrast,  $w = 0.5$  shows a higher value when combined with a cognitive learning of 2.

The next BPSO parameter test is the population size. The

population size can affect the size of the search space of the swarm. The greater the number of populations tested, the greater the search space of the system. In this test, the number of tested populations is 10, 20, 30, and 50. The aim is to test the system against small to large search spaces.

The data in Table V shows the test results based on the population size. The effect of the population size on the accuracy is different for each type of data. In Breast Cancer and Lung Cancer data, the best accuracy is obtained in the smallest number of populations in the test, which is 10. It is the opposite of Ovarian Cancer and Prostate Cancer data, which get the best accuracy when the population is 50. The only exception is Colon Tumor data that gets the highest accuracy when the population is 20.

Although the population size of 50 on Breast Cancer, Colon Tumor, and Lung Cancer data did not produce the highest accuracy, the accuracy obtained was also not the smallest. In Breast Cancer and Colon Tumor data, the smallest value is obtained when the population was 30. Meanwhile, in Lung Cancer data, the smallest value is obtained when the population size is 20 and 30, which produced the same value of 97.78%.

D. The effect of learning rate constant

In classic Backpropagation classification process, updating the weight and bias still requires a learning rate constant that determines how much learning Backpropagation will do. In this stage, the learning rate values are tested with values of 0.05, 0.1, and 0.5. Classification is done with 100 iterations.

TABLE VI. LEARNING RATE TESTING ON CLASSIC BACKPROPAGATION

Testing method	Learning rate	Accuracy (%) on Data					Average (%)
		Breast Cancer	Colon Tumor	Lung Cancer	Ovarian Cancer	Prostate Cancer	
BP	0,05	62,11	64,74	83,42	90,51	78,86	75,93
BP + BPSO	0,05	61	64,23	82,84	82,64	75,87	73,32
BP	0,1	69,16	62,82	84,01	94,48	85,95	79,28
BP + BPSO	0,1	64,95	64,62	82,91	93,67	78,62	76,95
BP	0,5	67,84	82,18	93,33	97,22	74,97	83,11
BP + BPSO	0,5	65,89	71,53	92,79	97,62	76,4	80,85

TABLE VII. BPSO FEATURE SELECTION RESULTS

BPSO Performance Measure	No. of Selected Features				
	Breast Cancer	Colon Tumor	Lung Cancer	Ovarian Cancer	Prostate Cancer
Original Features	24481	2000	12533	15154	12600
Backpropagation	11972	938	6057	7405	6242
CG BP (Powell-Beale)	12068	919	6113	7402	6086
CG BP (Fletcher-Reeves)	11923	910	6092	7338	6225
CG BP (Polak-Ribiere)	11886	947	6096	7385	6062

In the results presented in Table VI, the highest average accuracy is obtained when the learning rate is 0.5 which results in a value of 83.11% and 80.85% when the classification is combined with BPSO. This can also be seen in Colon Tumor and Lung Cancer data which received a significant increase in accuracy. In Colon Tumor data, the learning rate with a value of 0.5 can increase the average accuracy value from 64.1% to 76.85%. The same can also be seen in Lung Cancer data which the average increased from 83.29% to 93.06%. The average accuracy also got an increase in Ovarian Cancer data, although it was not significant.

However, a learning rate of 0.5 does not always provide the best value. In Prostate Cancer data, the best average accuracy is obtained when the learning rate is 0.1, which is 82.28%. Whereas for Breast Cancer data, a learning rate of 0.1 provides the best accuracy results when the testing is not combined with BPSO.

E. The effect of classification method

This testing phase aims to compare classic Backpropagation classification method with Conjugate Gradient Backpropagation. Three methods of updating the Conjugate Gradient, namely Powell-Beale, Fletcher-Reeves, and Polak-Ribiere were each tested to compare the results with each other. Classification with and without BPSO was also tested to compare the classification results of data that had passed the dimensionality reduction process.

Before the classification, the data goes through dimensionality reduction process with BPSO. The parameters used in BPSO are cognitive learning ( $c_1$ ) with a value of 1, cognitive learning ( $c_2$ ) with a value of 2, inertia weight ( $w$ ) with a value of 0.5, and a population size of 50. The performance measure used to assess the selected feature subset is the method which corresponds to the classification method. The results of dimensionality reduction with BPSO can be seen in Table VII.

TABLE VIII. CLASSIFICATION RESULTS WITH CLASSIC BP AND CONJUGATE GRADIENT BP

Testing method	Accuracy (%) on Data					Average (%)
	Breast Cancer	Colon Tumor	Lung Cancer	Ovarian Cancer	Prostate Cancer	
BP	67,84	82,18	93,33	97,22	74,97	83,11
BP + BPSO	65,89	71,53	92,79	97,62	76,4	80,85
CG BP (Powell-Beale)	68,89	83,72	99,44	98,42	90,5	88,19
CG BP (Powell-Beale) + BPSO	62,84	83,97	99,44	98,81	92,65	87,54
CG BP (Fletcher-Reeves)	57,53	78,85	96,14	92,43	88,2	82,63
CG BP (Fletcher-Reeves) + BPSO	59,68	78,85	94,52	98,42	86,75	83,64
CG BP (Polak-Ribiere)	65	80,77	98,89	97,65	91,9	86,84
CG BP (Polak-Ribiere) + BPSO	65,89	80,76	99,44	99,21	90,45	87,15

In the data listed in Table VII, BPSO has succeeded in selecting features with a similar ratio in each data, which is around 48% of the number of original features. The method of performance measure that tends to select the most data is

Conjugate Gradient Backpropagation with Polak-Ribiere method, with an average of selecting 6475.2 features per data. However, the difference is not significant with other methods. For example, Fletcher-Reeves method selects an average of

6497.6 features per data.

The test results in Table VIII show that in most cases, Conjugate Gradient Backpropagation method can produce a high accuracy, but this does not apply to Fletcher-Reeves Conjugate Gradient method. From all experiments, Fletcher-Reeves method always produces the smallest accuracy compared to other Conjugate Gradient methods, with an average of 83.14%. The best accuracy value was obtained by Conjugate Gradient Powell-Beale and Polak-Ribiere methods on Lung Cancer data, both of which produce an accuracy of 99.44%.

#### F. Test results analysis

Based on the conducted tests, the classification performance is very dependent on the type of data being classified. Lung Cancer and Ovarian Cancer data always show quite high results without having to use feature selection, in contrast to Breast Cancer data which shows fairly low accuracy value in all tests. This can happen due to several factors, one of which is the number of features contained in the data. It can be seen in Table 1 that Breast Cancer data has the highest number of features, namely 24481 features with a sample of only 97 samples. Even after going through BPSO feature selection, the number of features in Breast Cancer data still fairly high, with an average of 11962 features selected, as can be seen in Table VII.

The application of BPSO as feature selection also does not guarantee an increase in accuracy. In some cases, the accuracy decreased when the data had been feature-selected by BPSO. For example, in Breast Cancer data, BPSO decreases the accuracy when combined with classic Backpropagation and Conjugate Gradient Powell-Beale. The average value of 68.37% decreased to 64.37% when BPSO was used. The most drastic decrease can be seen in Colon Tumor data, where BPSO combined with classic Backpropagation reduces the accuracy from 82.18% to 71.53%.

The application of Conjugate Gradient Backpropagation is seen to increase the accuracy quite drastically in Prostate Cancer data. The average accuracy which was only 75.69% was successfully increased to 90.08% by using Conjugate Gradient Backpropagation. However, the combination of Conjugate Gradient Backpropagation and BPSO seems to reduce the accuracy on Fletcher-Reeves and Polak-Ribiere.

#### V. CONCLUSION

In this study, a classification system with Modified Backpropagation with Conjugate Gradient combined with Binary Particle Swarm Optimization (BPSO) feature selection has been successfully developed. Several BPSO parameters such as cognitive learning, social learning, inertia weight, and population size as well as Backpropagation parameters such as number of neurons, number of hidden layers, and learning rate have also been tested to determine the best parameter values to be used for the final test.

The classification results show that the Modified

Backpropagation with Conjugate Gradient succeeded in increasing the average accuracy when compared to classic Backpropagation. Conjugate Gradient Backpropagation succeeded in increasing the average accuracy to 85.89%. Combining Conjugate Gradient Backpropagation with BPSO can also increase the average accuracy even higher. The average accuracy obtained from all tests with Conjugate Gradient Backpropagation and BPSO is 86.11%. However, combining BPSO with classic Backpropagation reduces the accuracy, which decreases the average value to 80.85%.

Classification methods and parameter determination are not the only factors that can determine the goodness of the accuracy. Different types of data can also affect classification performance. Data with a very large number of features can also reduce classification performance, such as Breast Cancer data which always produces a low accuracy when compared to other data.

#### REFERENCES

- [1] W. H. Organization, "Cancer fact sheet," September 2019. [Online]. Available: <http://www.who.int/news-room/fact-sheets/detail/cancer>. [Accessed 18 September 2019]
- [2] Purbolaksono, M.D., Widiastuti, K.C., Adiwijaya, Mubarak, M.S., and Ma'ruf, F.A., 2017. Implementation of Mutual Information and Bayes Theorem for Classification Microarray Data. In *Journal of Physics: Conference Series* (Vol.921, No.1, p. 012011). IOP Publishing
- [3] Dashtban, M. and Balafar, M., 2017. Gene selection for microarray cancer classification using a new evolutionary method employing artificial intelligence concepts. *Genomics*, 109(2), pp.91-107.
- [4] Cao, J., Zhang, L., Wang, B., Li, F. and Yang, J., 2015. A fast gene selection method for multi-cancer classification using multiple support vector data description. *Journal of biomedical informatics*, 53, pp.381-389.
- [5] Pradana, A.C., Adiwijaya, A. and Aditsania, A., 2018. Implementasi Algoritma Binary Particle Swarm Optimization (BPSO) dan C4.5 Decision Tree untuk Deteksi Kanker Berdasarkan Klasifikasi Microarray Data. *eProceedings of Engineering*, 5(3).
- [6] Chen, K.H., Wang, K.J., Tsai, M.L., Wang, K.M., Adrian, A.M., Cheng, W.C., Yang, T.S., Teng, N.C., Tan, K.P. and Chang, K.S., 2014. Gene selection for cancer identification: a decision tree model empowered by particle swarm optimization algorithm. *BMC bioinformatics*, 15(1), p.49.
- [7] Nurfalah, A. and Suryani, A.A., 2016. Cancer Detection Based on Microarray Data Classification Using PCA and Modified Back Propagation. *Far East Journal of Electronics and Communications*, 16(2), p.269.
- [8] Guo, Q.M., 2003. DNA microarray and cancer. *Current opinion in oncology*, 15(1), pp.36-43.
- [9] Siang, T.C., Soon, T.W., Kasim, S., Mohamad, M.S., Howe, C.W., Deris, S., Zakaria, Z., Shah, Z.A. and Ibrahim, Z., 2015. A review of cancer classification software for gene expression data. *International Journal of Bio-Science and Bio-Technology*, 7(4), pp.89-108.
- [10] Sarhan, A.M., 2009. Cancer classification based on microarray gene expression data using DCT and ANN. *Journal of Theoretical & Applied Information Technology*, 6(2).
- [11] Mafarja, M., Jarrar, R., Ahmad, S. and Abusnaina, A.A., 2018. Feature selection using binary particle swarm optimization with time varying inertia weight strategies. In *Proceedings of the 2nd International Conference on Future Networks and Distributed Systems* (pp. 1-9).
- [12] Kennedy, J. and Eberhart, R.C., 1997. A discrete binary version of the particle swarm algorithm. In *1997 IEEE International conference on*

- systems, man, and cybernetics. *Computational cybernetics and simulation* (Vol. 5, pp. 4104-4108). IEEE.
- [13] Pashaei, E., Ozen, M. and Aydin, N., 2015. A novel gene selection algorithm for cancer identification based on random forest and particle swarm optimization. In *2015 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)* (pp. 1-6). IEEE.
- [14] Nezamabadi-pour, H., Rostami-Shahrbabaki, M. and Maghfoori-Farsangi, M., 2008. Binary particle swarm optimization: challenges and new solutions. *CSI J Comput Sci Eng*, 6(1), pp.21-32.
- [15] Adiwijaya, A., Wisesty, U.N., Lisnawati, E., Aditsania, A. and Kusumo, D.S., 2018. Dimensionality Reduction using Principal Component Analysis for Cancer Detection based on Microarray Data Classification. *Journal of Computer Science*, 14(10).
- [16] Heravi, A.R. and Hodtani, G.A., 2018. A new correntropy-based conjugate gradient backpropagation algorithm for improving training in neural networks. *IEEE transactions on neural networks and learning systems*, 29(12), pp.6252-6263.
- [17] Adiwijaya, W. and Nhita, F., 2014. Study of line search techniques on the modified backpropagation for forecasting of weather data in Indonesia. *Far East J. Math. Sci*, 86(2).
- [18] "ELVIRA Biomedical Data Set Repository," February 2005. [Online] Available at: <http://leo.ugr.es/elvira/DBCRepository>. [Accessed 16 June 2020]
- [19] Eesa, A.S. and Arabo, W.K., 2017. A normalization methods for backpropagation: a comparative study. *Science Journal of University of Zakho*, 5(4), pp.319-323.
- [20] Patro, S. and Sahu, K.K., 2015. Normalization: A preprocessing stage. *arXiv preprint arXiv:1503.06462*.
- [21] Jain, S., Shukla, S. and Wadhvani, R., 2018. Dynamic selection of normalization techniques using data complexity measures. *Expert Systems with Applications*, 106, pp.252-262.
- [22] Annavarapu, C.S.R., Dara, S. and Banka, H., 2016. Cancer microarray data feature selection using multi-objective binary particle swarm optimization algorithm. *EXCLI journal*, 15, p.460.
- [23] Vieira, S.M., Mendonça, L.F., Farinha, G.J. and Sousa, J.M., 2013. Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients. *Applied Soft Computing*, 13(8), pp.3494-3504.
- [24] "Pyswarms Source Code," May 2020. [Online] Available at: <https://github.com/ljvmiranda921/pyswarms/pull/404#issuecomment-632149623> [Accessed 5 August 2020]
- [25] Dwivedi, A.K., 2018. Artificial neural network model for effective cancer classification using microarray gene expression data. *Neural Computing and Applications*, 29(12), pp.1545-1554.
- [26] Lam, H.K., Ekong, U., Liu, H., Xiao, B., Araujo, H., Ling, S.H. and Chan, K.Y., 2014. A study of neural-network-based classifiers for material classification. *Neurocomputing*, 144, pp.367-377.
- [27] Majid, M.A., Abidin, A.F.Z., Anuar, N.D.K., Kadiran, K.A., Karis, M.S., Yusoff, Z.M., Anuar, N.H.K. and Rizman, Z.I., 2018. A comparative study on the application of binary particle swarm optimization and binary gravitational search algorithm in feature selection for automatic classification of brain tumor MRI. *Journal of Fundamental and Applied Sciences*, 10(2S), pp.486-498.
- [28] Philip, J.T. and George, S.T., 2019. Performance Comparison of Neural Network Backpropagation Algorithms in Detecting P300 Signals from Mind-Speller Data. In *2019 2nd International Conference on Signal Processing and Communication (ICSPC)* (pp. 117-121). IEEE.
- [29] Pratiwi, M.S., Aditsania, A., Adiwijaya, A., 2018. Cancer Detection Based on Microarray Data Classification using Genetic Bee Colony (GBC) and Conjugate Gradient Backpropagation with Modified Polak Ribiere (MBP-CGP). In *2018 International Conference on Computer, Control, Informatics and its Applications (IC3INA)* (pp. 163-168). IEEE.
- [30] Aydadenta, H., 2018. A Clustering Approach for Feature Selection in Microarray Data Classification Using Random Forest. *Journal of Information Processing Systems*, 14(5).
- [31] Astuti, W. and Adiwijaya, A., 2019. Principal Component Analysis Sebagai Ekstraksi Fitur Data Microarray Untuk Deteksi Kanker Berbasis Linear Discriminant Analysis. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 3(2), pp.72-77.
- [32] Priyono, I., Adiwijaya, A., Aditsania, A., 2020. Cancer Detection Based on Microarray Data Classification Using Principal Component Analysis and Functional Link Neural Network. *Journal of Data Science and Its Applications*, 3(2).
- [33] Munzir, A.F.H., Adiwijaya, A., Aditsania, A., 2018. Analisis Reduksi Dimensi pada Klasifikasi Microarray Menggunakan Powell-Beale. *E-Jurnal Matematika*, 7(1), pp.17-24.