APPLICATION OF SUPPORT VECTOR MACHINE WITH OPTIMIZATION PARTICLE SWARM OPTIMIZATION FOR DETECTING MALARIA DISEASE

Arya Putra Sanjaya[1], Nelly Indriani W.,S.Si., M.T.[2]

[1,2] Universitas Komputer Indonesia
Dipatiukur Street No. 112-116, Coblong, Bandung, Jawa Barat 40132
E-mail: putraarya14@gmail.com, nelly.indriani@email.unikom.ac.id[2]

ABSTRACT

Classification can be defined as work that performs the training / learning of the target function f which maps each set of attributes (features) x to one of a number of available y class labels. The training work will produce a model which is then stored as memory. One example is the classification of malaria disease because it has typical clinical symptoms, in this research using the method of support vector machine in detecting malaria disease, and using particle swarm optimization in determining SVM parameters, because SVM has shortcomings in the selection of parameters that affect the process Classification, PSO was chosen because many have been shown to have good performance in that case because PSO have simple concept and efficient in reaching global solution both minimize and maximize.
This research used 164 training data and 70 test data. yielding accuracy of 91.4286% in testing optimized with PSO and without optimization to get an accuracy of 87.1429%. These results indicate that PSO is effective in improving classification accuracy.

Keywords: Malaria Detecting, Support Vector Machine, Particle Swarm Optimization

.

1. Introduction

Support Vector Machine (SVM) is a learning system that uses space hypothesis in the form of linear functions in a feature space (feature space) high-dimensional, trained with learning algorithms based on optimization theory by implementing learning bias derived from the theory of learning statistics [1]. SVM has the advantage of being able to identify a separate hyperplane that maximizes margins between two different classes however, SVM has a deficiency to the problem of selecting parameters especially C parameters to determine the penalty due to misclassified data and the value specified by the user so that different parameter values result in different accuracy [2].

Particle Swarm Optimization (PSO) is one of the optimization methods that can be used to improve SVM accuracy by optimizing parameters on SVM [3]. SVM-PSO produces higher accuracy than SVM, PSO performs noise feature selection. It is evident that with the selection of parameters on SVM [4]. In the research of Shin Wei Lin tested several optimization algorithms to determine parameters on SVM with breast cancer dataset, SVM-PSO with RBF kernel obtained 94% better accuracy than SVM-GA algorithm with dimension range C 0.01 - 35.000 and gamma range 0, 0001 - 32, for the PSO parameter the acceleration values 1 and 2 are 2, the number of particles 8 and the iteration of 250 [5]. In Bo Tsu Chen's research applying PSO for SVM optimization on RBF kernel with C 0-100 san gamma (γ) 0-100 dimension and PSO parameter with 20 particle jumps, 100 iterations and 2 and 2 acceleration values by predicting the company's financial crisis 97.76% accuracy and without optimization get an accuracy of 81.15% [6]. In the study of Abd. Samad Hasan implements SVM-PSO using the RBF kernel by reviewing opinions on the movie to negative review class and positive review getting accuracy for SVM without optimization of 71.97% and with optimization with PSO getting accuracy of 77% [7]. In the Xiaoyung Liu study applied SVM-PSO using RBF kernel on breast cancer data by searching C dimension range 0.01 - 100 and gamma range (γ) 0.01 - 100 and for PSO parameter of particle number 20, iteration 100 and acceleration value 1 , 5 and 1.7 get an accuracy of 90%[8]. In the Fatima Ardjina study applying SVM-PSO using RBF kernel to classify multiclass benchmarks for the 0.01-100 and 0.00-1.000 0.010 gamma ranges and for the PSO parameter of particle number 30, 100 iterations and acceleration values of 1.5 get an accuracy of 86 , 50% and without optimization 76.62% [9]. In Ali Khazaee's study applying SVM-PSO with Electrocardiograms (ECG) heart case using RBF kernel for the range of 1 to 0 million descended, gamma dimension (γ) 0-1 and for PSO parameter of particle number 20, iteration 200 and acceleration value 2 get accuracy 95.1970% and without optimization PSOO get an accuracy of 93.847%[10]. In ResearchYan Hui Feng applied SVM-PSO using the RBF kernel to classify hospital customers for the C 1-1000 search range and for PSO parameters the number of iterations of 100 particles of 20 and acceleration values of 0.4 and 0.8 obtained an accuracy of 98.32% [11]. In the Ermiono Marami

study applying SVM-PSO using RBF kernel in case of face detection with PSO optimization obtains an accuracy of 93.95% [12]. In Tang Faming research applying SVM-PSO using RBF kernel on UCI dataset with search range C 0-1, gamma dimension range (γ) 0-1 for PSO parameter of particle number 24, iteration 100 and acceleration value 0,4 and 0,9 Get 100% accuracy[13In the S P Yazhini study on the case of detecting the disorder get SVM accuracy with optimization of PSO equal to 94,09% and without optimation 92,065%[14]. In Vithal Manekar's research applying SVM-PSO to the instrumentation detection system by optimizing for C 0-1 dimension range, gamma 0-1 and for parameter PSO particle number 30 and iteration 100 get 81,8% accuracy and without optimization 47,99%[15]. In the T-study the kumaresan apply SVM-PSO to Image Spam Filtering (ISF) for the search range of dimensions C 0-100, gamma 0-10 and for PSO parameters the number of particles 10 and iteration 100 get an accuracy of 82.5% [16]. In the Onuma Kosanan study applied SVM-PSO in the prediction of Thai agricultural statistics with dimension range C 1-10, gamma 0-10 and PSO parameter of particle number 20, 100 iterations and acceleration values 0.4 and 0.9 get an error value of 0 , 0040% and without PSO optimization get error 0,0388%[17].

Malaria is still one of the most important public health problems in the world. In Indonesia, especially outside Java and Bali, up to now the high malaria morbidity rate is around 400,000 positive cases each year, 4,000 cases of complications or late death due to late treatment. Clinical diagnosis is the most widely used approach in peripheral or remote areas where laboratory testing to support a clinical diagnosis of malaria is not available. The introduction of typical clinical symptoms in malaria endemic areas is helpful in the rapid, precise and rational treatment of malaria, in order to reduce morbidity and mortality from malaria. Health workers in endemic areas are expected to recognize typical clinical symptoms and signs in their area, as a preliminary step before being confirmed in blood smear and Rapid Diagnostic Test (RDT) [18].

Based on this case, in this study apply SVM with PSO optimization to detect malaria disease based on clinical signs and symptoms that aims to get good accuracy.

2. Discussion Method

Process analysis is defined as the decomposition of the main process into subprocesses in order to optimize the Support Vector Machine (SVM) with particle swarm optimization (PSO) by using a prediction case of a malaria patient based on symptoms of the patient experienced and analyzing SVM without PSO and SVM with PSO. The general overview of process analysis of the algorithm needs can be seen in Figure 1 overview of the process as follows:

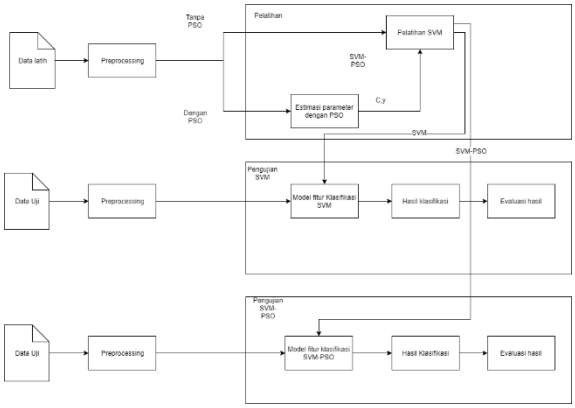


Figure 1 Process Analysis

Information

1. Practice data and test data is patient data collection at Jaifuri Arso III Community Health Center of Skanto District of Keerom Regency. The data contain the results of anamnesis in Patients. Then the data is divided into training data and test data with the file format .xls..
2. The input data processing is done to be modeled in the classification model. The pre-processing step on this system is
   a. Feature selection is the selection of input data used as a feature that matches the symptoms and clinical signs in malaria patients.
   b. Data transformation is the process of changing data that is in a certain scale, changed into the range between (0-1).
3. Training

   Is a process for getting feature models based on training data.
   a. PSO Parameter Estimation

      Generating parameters for SVM training so optimizing parameters on the training process up to the optimal parameters.
   b. Training SVM

      Is a learning stage of trainer data, at this stage to do the training process of training data and parameter values used both without PSO and with PSO.
4. Testing

   Testing is the process of classifying test data from feature models that have been obtained from the training process.

a. Model Classification

   It is the result of SVM and SVM-PSO training that has obtained an optimal feature model for classifying data testing.

b. Classification Result

   Is a model validation process with testing based on data testing.

c. Evaluation of Results

   Evaluation of model results by using confusison matrix method.

2.1 Feature Selection

Feature selection is the feature selection process, the data selected for processing. Data input from records of puskesmas Arso III will be selected first according to symptoms and clinical signs of malaria to be used as features because not all input data is not an important feature in the symptoms and clinical signs of malaria disease In this study the symptoms associated with malaria is fever , Chills, sweating, headache, pain, nausea, vomiting, lymph enlargement, body temperature, anemia and diarrhea based on interviews with health workers and papers [19][20].

2.2 Transformation Data

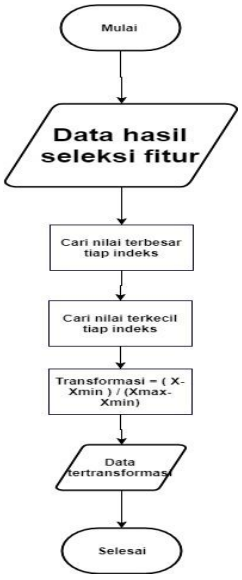
The transformation process is the process of structuring the input data to be entered into the classification model because the data model is needed between 0 to 1 to be processed into the classification model both test data and training data. In this research, data transformation method using scaling method. Scaling method is a method that can change the data in a certain scale, to scale in the range (0-1) with equation 1 [22].

$$\hat{x} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Information :

Xmin is a minimal value and Xmax is the maximum value.

Flowchart Transformation Data can be seen in figure 2.



Gambar 2 Flowchart Transformasi Data

2.3 Analisis Metode SVM

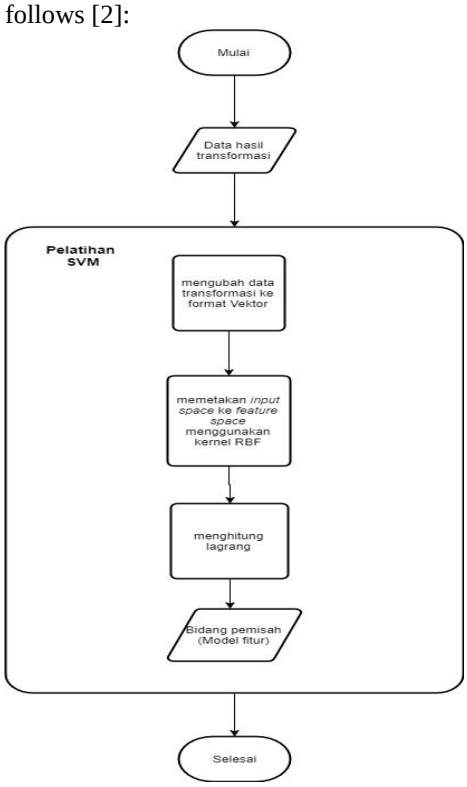
Support Vector Machine (SVM) is a technique to make predictions, both in case of classification and regression. SVM has a linear classifier basic principle that is a classification case that can be separately separated, but SVM has been developed in order to work on a non-linear type by incorporating kernel concepts in high-dimensional workspaces. In the high-dimensional space, hyperplane (hyperplane) will be able to maximize the distance (margin) between data classes [2].

2.3.1 Training SVM

According to the tutorial on the recommended kernel function to be tested for the first time is the function of the RBF kernel because it has the same performance with the linear kernel on certain parameters, has behavior like the sigmoid kenel function with the parameter of course and the small value range. The kernel function used in this research is the RBF kernel function with equation 2 [2].

$$K(x_i, x_j) = \exp(-\gamma||x_i - x_j||^2) \quad (2)$$

Flowchart SVM training can be seen in Figure 3 as follows [2]:



Gambar 3 Flowchart Pelatihan SVM

1. Converting data to vector format
   [+1 1:1 2:1 3:0 4:1 5:1 6:1 7:0 8:1 9:0,4285 10:0 11:0]
   +1 represents the class and 1 represents the index and after the index declares the contents of the index value..
2. Mapping the input space to feature space
   Mapping the input space to feature space using the RBF kernel.
3. search *Lagrange multiplier*
   Look for lagrange multiplier using equation 3[2]

$$L_D = \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i,j} a_i a_j y_i y_j x_i x_j \quad (3)$$

with $constraint$ 1

$$0 \leq a_i \leq C \quad (4)$$

constraint 2

$$a_1, a_2, a_3, \ldots . a_n \geq 0 \quad (5)$$

In this process to get the number of support vector, ai and bias.

4. Division field or feature model
   From the process that has been done then subtitusikan to separator function with the equation 6 [2].

$$f(x_d) = \sum_{=1}^{4} a_i y_i \phi(x_i)\phi(x_d) + b \quad (6)$$

2.3.2 Testing SVM

From the training process, it can then test the test data into the +1 or -1 classes with the feature model or the delimited field already obtained. With the operation sign value greater than 0 will go into the +1 class and if it is smaller then go to the class -1.

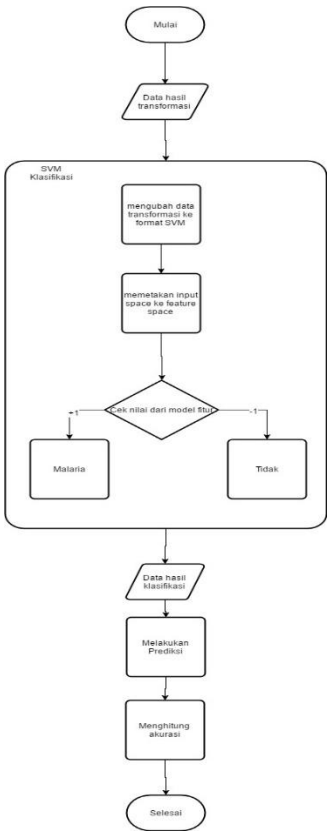


Figure 4 Flowchart Testing SVM

1. The transformed data is converted to vector format.
2. Mapping the input space to feature space

Mapping test data with equations 7

$$K(x_{latih}, x_{uji}) = \exp(-\gamma||x_{latih} - x_{uji}||^2) \quad (7)$$

3. Classification
   From the model features that have been obtained and do kernelisasi on the test data so as to get the feature model.

$$f(x_d) = \sum_{=1}^{n} a_i y_i K(x_{latih} - x_{uji}) + b \quad (8)$$

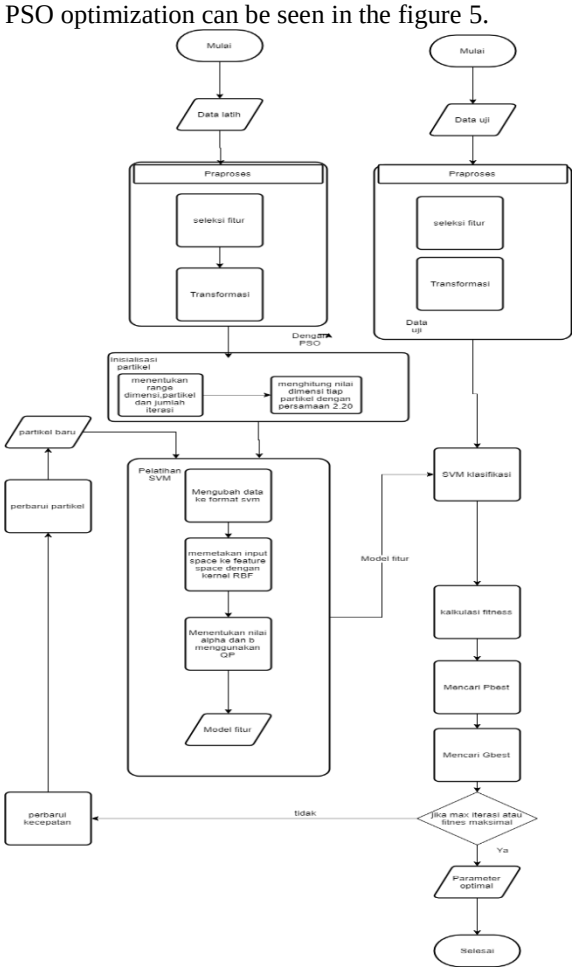
If f (xd)> 1 then it is detected malaria and if <1 then it is detected not.

4. Callculate accuracy.

Calculates this accuracy from the correct classification results.

2.4 Analysis of SVM Method with PSO Optimization

The flow in analyzing the SVM method with PSO optimization can be seen in the figure 5.



Figur 5 Flowchart SVM berbasis PSO

Train data and test data have gone through the pre-process stage.

1. Initialization of particles

The process of initialization of particles is the process of representing each particle with values C and γ based on the upper and lower limits of the search space of each dimension determined by equation 8 [21].

$$\overline{X}_{1,}c = x_{min,c} + r_1(x_{max,c} - x_{min,c}) \quad (8)$$

2. Evaluate fungsi *fitness*.

After obtaining the value of the particle dimension in the previous stage, then the evaluation of fitness function by doing SVM training Here is the flow done in this stage.

a. Do SVM Train.

At this stage it aims to get the dividing field or feature model.

b . Test Model.

From the training process test the feature model with test data.

c. Calculate the fitness function of each particle.

Calculate the fitness of each particle with the following equation 9

$$F=\frac{jumlah\ klasifikasi\ benar}{jumlah\ data} \quad (9)$$

3. Looking for pBest

Phase search for pBest is the stage of seeking the best individual positions. This stage begins by comparing the pBest fitness values and the current iteration particles. If the pBest fitness value is greater than the current iterative particle fitness value, the pBest position value remains.

4. Looking for Gbest

The gBest looking stage is the stage of finding the best global position. This stage begins by comparing the gBest fitness values and pBest. If the gBest fitness value is greater than the pBest fitness value, the gBest position value remains. If the gBest fitness value is less than the pBest fitness value, the gBest position value is replaced by the pBest position value, if it has not maximum iteration or the fitness value is not optimal then to step 5 and 6.

5. Update Velocity

The stage of updating the particle velocity by the formula is the calculation phase of particle velocity based on pBest, gBest, and inertia weight obtained from the previous stage. With the result of the fitness value then found Gbest because Gbest is the highest fitness value.

Calculate the weight of inertia with equation 10 as follows[21] :

$$w(t) = w_{max} - \frac{t}{maksiterasi} * (w_{max} - w_{min}) \quad (10)$$

After finding the weight of the inertia calculate the velocity for the iteration + 1 with the following equation [21]:

$$V_{1,c}(t+1) = W_{1,c} * v_{1c}(t) + (r_1 * c_1) * \left(p_{best1C}(t) - x(t)\right) + (r_2 * c_2) * (G_{bestc}(t) - x_I(t)) \quad (10)$$

Where r is the random number random 0 to1 and c_1, c_2 is value akselerasi.

6. Update Particle

At this stage update the position of the new particle with the equation for each dimension as follows [21]

$$x_{1c}(t+1) = x_{1c}(t) + v_{1c}(t+1) \quad (11)$$

From new particles do training and testing up to maximum iteration or fitness value = 1.

3. Research Results

Training data used were 164 with malaria class and not as many as 82 and 70 test data. This test tested SVM method without optimization and with optimization. For PSO parameter values can be seen in Table

Table 1 Parameter PSO

Parameter	Nilai
C_1	1
C_2	1
W_{max}	0,8
W_{min}	0,4

3.1 Testing Range Dimension C

For testing on the C dimension, the search space range used is 0.01-10, 0.01-50, 0.01-100 0.1-10 0.1-50 0.1-100, Detail parameters used In the test range C-dimensional search space is as follows:

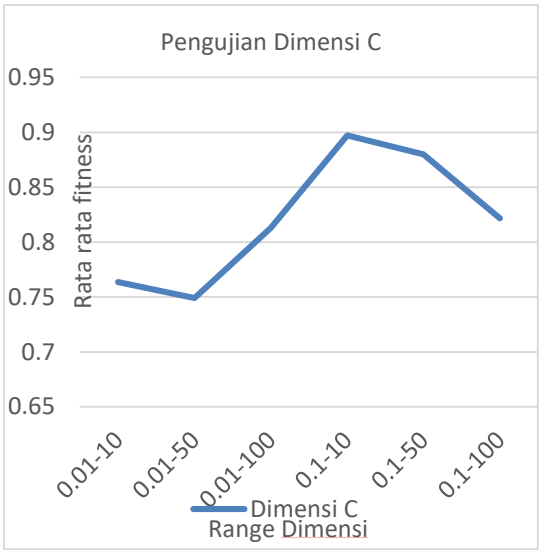


Figure 6 Testing Dimension C

From the test, the value of C dimension range with the highest fitness average is 0,1-10.

3.2 Testing Range Dimension gamma

For testing on gamma dimensions, the search space range used is 0.01-4, 0.01-6, 0.01-10, 0.1-4, 0.1-6, 0.1-10, detail parameters used in Test of gamma dimension search space range is as follows:

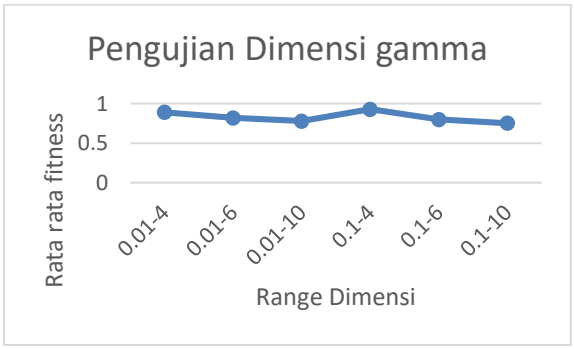


Figure 7 Testing Dimension gamma

From the tests conducted to get the highest average fitness that is between the dimension range of 0.1-4.

3.3 Testing number of particle

For the test Number of particles, the length of the search space used is 5, 10, 15, 20, 25, Detail parameters used in the experiment Testing the number of particles is as follows:

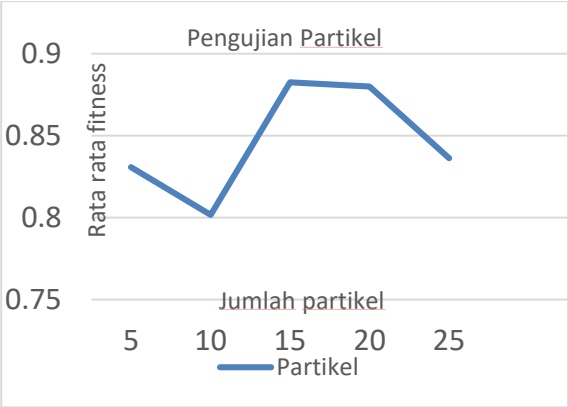


Figure 8 Testing Number of particle

From the test results for the number of particles with the highest average fitness value is 15.

3.4 Testing Number of Iterati

For the test Number of iterations, the length of the search space used is 10, 20, 30, 40, 50, Detail parameters used in the experiment Testing the number of particles is as figure 9:

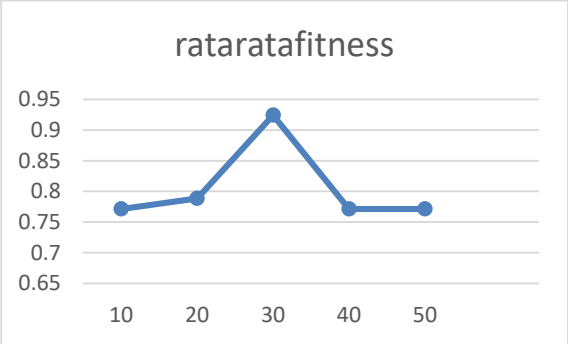


figure 9 Pengujian Jumlah Iterasi

3.5 Testing performance SVM

Performance tests used 184 training data and 70 test data for testing values of C = 1 and gamma (γ) = 0.5.

Table 2 Classification SVM

No	Status	No	Status
1	TP	21	TP
2	TP	22	TP
3	TP	23	TN
4	TN	24	TP
5	TP	25	TP
6	TP	26	TP
7	TN	27	TN
8	TP	28	TP
9	TP	29	TP
10	TP	30	TN
11	TN	31	TP
12	TP	32	TN
13	FP	33	TP
14	TP	34	TP
15	TP	35	TP
16	TN	36	TN
17	TN	37	TP
18	TP	38	FP
19	TP	39	TP
20	TN	40	TP
41	TP	56	TN
42	TP	57	TP
43	TN	58	TP
44	TP	59	TP
45	TP	60	FN
46	TP	61	FN
47	TP	62	FN
48	TN	63	FN
49	TP	64	TN
50	TP	65	TN
51	TP	66	TN
52	TN	67	TN
53	TP	68	FP
54	TP	69	FP
55	TP	70	FP

From the classification result get TN = 20 TP = 41, FP = 4, FN = 5..

$$Accuracy = \frac{61}{70} X\ 100\% = 87{,}1429\%$$

3.6 Testing performance with PSO

Performance tests used 164 training data and 70 test data with the previous optimization process performed with range C 0.1-10 and gamma dimension range 0.1-4.

Table 3 Classification SVM- PSO

No	Status	No	Status
1	TP	21	TP
2	TP	22	TP
3	TP	23	TN
4	TN	24	TP
5	TP	25	TP
6	TP	26	TP
7	TN	27	TN

No	Status	No	Status
8	TP	28	TP
9	TP	29	TP
10	TP	30	TN
11	FP	31	TP
12	TP	32	TN
13	FP	33	TP
14	TP	34	TP
15	TP	35	TP
16	TN	36	TN
17	TN	37	TP
18	TP	38	FP
19	TP	39	TP
20	FP	40	TP
41	TP	56	TN
42	TP	57	TP
43	TN	58	TP
44	TP	59	TP
45	TP	60	TP
46	TP	61	FN
47	TP	62	FN
48	TN	63	FN
49	TP	64	TN
50	TP	65	TN
51	TP	66	TN
52	TN	67	TN
53	TP	68	FP
54	TP	69	FP
55	TP	70	FP

The result is TN = 21, TP = 43 and FP = 2, FN = 4.

$$Accuracy = \frac{64}{70} X\ 100\% = 91{,}42\%$$

4 Closing

4.1 Clonclusion

Based on the results of research that has been done, it can be concluded that the application of Support Vector Machine (SVM) method with train data amounted to 164 and 70 test data, based on the test results using SVM without optimization, the accuracy is 87.1429% accuracy test using Confussion Matrix . Then for the hybrid SVM method with optimization of particle swarm optimization (PSO) with the test results get get 91.4286%.

5.1 Suggestion

This research uses inertia weight and acceleration coefficient which influence particle velocity and position to find the best or optimal parameter, because if choosing the value of both of these factors then PSO can be optimal so it can enable to get better result, hopefully the next research try many acceleration value And inertia weight.

Since the initial particles formed are dependent on the random value, the determination of the number of particles and iterations has an effect, so as not to get caught in the best local position.

Bibliography

[1] T. Zhang, "An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods," *AI Mag.*, vol. 22, no. 2, p. 103, 2001.

[2] K. Sembiring, "penerapan Teknik Support Vector Machine untuk pendeteksian intrusi pada Jaringan," no. September, 2007, pp. 1–28.

[3] H. L. Chen *et al.*, "Towards an optimal support vector machine classifier using a parallel particle swarm optimization strategy," *Appl. Math. Comput.*, vol. 239, pp. 180–197, 2014.

[4] B. Chen and M. Chen, "Applying particles swarm optimization for support vector machines on predicting company financial crisis," *Iternational Conf. Bus. Econ. Res.*, vol. 1, pp. 301–305, 2011.

[5] S.-W. Lin, K.-C. Ying, S.-C. Chen, and Z.-J. Lee, "Particle swarm optimization for parameter determination and feature selection of support vector machines," *Expert Syst. Appl.*, vol. 35, no. 4, pp. 1817–1824, 2008.

[6] B. Chen and M. Chen, "Applying particles swarm optimization for support vector machines on predicting company financial crisis," *Iternational Conf. Bus. Econ. Res.*, vol. 1, pp. 301–305, 2011.

[7] B. Abd Samad Hasan, H. Burairah, I. Ananta, J. Zeniarja, Others, and B. A. S. Hasan, "Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization," *Malaysian Tech. Univ. Conf. Eng. Technol.*, pp. 545–552, 2012.

[8] X. Liu, H. Fu, X. Liu, and H. Fu, "PSO-Based Support Vector Machine with Cuckoo Search Technique for Clinical Disease Diagnoses," *Sci. World J.*, vol. 2014, pp. 1–7, 2014.

[9] F. Ardjani, K. Sadouni, and M. Benyettou, "Optimization of SVM multiclass by particle swarm (PSO-SVM)," *2010 2nd Int. Work. Database Technol. Appl. DBTA2010 - Proc.*, no. December, pp. 32–38, 2010.

[10] A. Khazaee, "Heart Beat Classification Using Particle Swarm Optimization," no. May, pp. 25–33, 2013.

[11] I. Management, "Optimization and Application of Support Vector Machine Based on SVM Algorithm Parameters," vol. 11, no. 2, pp. 165–169, 2013..

[12] E. Marami and A. Tefas, "Optimization and Support Vector Machines," vol. 248434, pp. 369–374, 2010.

[13] E. Marami and A. Tefas, "Optimization and Support Vector Machines," vol. 248434, pp. 369–374, 2010.

[14] S. P. Yazhini and R. Devipriya, "Support Vector Machine with Improved Particle Swarm Optimization Model for Intrusion Detection," vol. 7, no. 2, pp. 37–42, 2016.

[15] V. Manekar and K. Waghmare, "Intrusion Detection System using Support Vector Machine (SVM) and Particle Swarm Optimization (PSO)," no. 3, pp. 2–6, 2014.

[16] T. Kumaresan, "Image Spam Filtering using Support Vector Machine and Particle Swarm Optimization," pp. 17–21, 2015.

[17] O. Kosanan, "A Hybrid Particle Swarm Optimization Algorithm and Support Vector Machine Model for Agricultural Statistic of Thailand Forecasting," pp. 1167–1178, 2016.

[18] L. Siahaan, "Gejala dan Tanda Klinis Malaria di Daerah Endemis," *Maj. Kedokt. Indones.*, vol. 58, pp. 211–215, 2008.

[19] Budi Santosa, *Data Mining,Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu, 2007.

[20] Eko Prasetyo, *Data Mining : Mengolah Data menjadi informasi menggunakan MATLAB*. Yogyakarta: Andi Offset, 2014.

[21] F. Marini and B. Walczak, "Chemometrics and Intelligent Laboratory Systems Particle swarm optimization (PSO). A tutorial," *Chemom. Intell. Lab. Syst.*, 2015.

[22] H. Himawan, "Rancang Bangun Pengembangan Pola Pengetahuan," vol. 2010, no. semnasIF, pp. 263–271, 2010.