

HEALTH CLAIM INSURANCE PREDICTION USING SUPPORT VECTOR MACHINE WITH PARTICLE SWARM OPTIMIZATION

**Syaiful Anam^{1*}, M. Rafael Andika Putra², Zuraidah Fitriah³, Indah Yanti⁴,
Noor Hidayat⁵, Dwi Mifta Mahanani⁶**

^{1,2,4} Computer and Data Science Laboratory, Mathematics Department, Brawijaya University

^{1,3,4,5,6} Mathematical Optimization and Computing Research Group, Brawijaya University

² Undergraduate Student of Actuarial Science, Brawijaya University,
Malang, 65145, Indonesia

Corresponding author's e-mail: * syaiful@ub.ac.id

ABSTRACT

Article History:

Received: 20th December 2022

Revised: 12th April 2023

Accepted: 15th April 2023

Keywords:

Claim prediction;

Global optimum;

Health insurance;

Parameter's selection;

Particle Swarm Optimization;

Support Vector Machine.

The number of claims plays an important role in the profit achievement of health insurance companies. Prediction of the number of claims could give significant implications for the profit margins generated by the health insurance company. Therefore, the prediction of claim submission by insurance users in that year needs to be done by insurance companies. Machine learning methods promise a great solution for claim prediction of health insurance users. There are several machine learning methods that can be used for claim prediction, such as the Naïve Bayes method, Decision Tree (DT), Artificial Neural Networks (ANN), and Support Vector Machine (SVM). The previous studies show that the SVM has some advantages over the other methods. However, the performance of the SVM is determined by some parameters. Parameter selection of SVM is normally done by trial and error so that the performance is less than optimal. Some optimization algorithms based heuristic optimization can be used to determine the best parameter values of SVM, for example, Particle Swarm Optimization (PSO) and Genetic Algorithm (GA). They are able to search the global optimum, easy to be implemented. The derivatives aren't needed in its computation. Several researches show that PSO give better solutions if it is compared with GA. All particles in the PSO are able to find the solution near global optimal. For these reasons, this article proposes the health claim insurance prediction using SVM with PSO. The experimental results show that the SVM with PSO gives great performance in the health claim insurance prediction and it has been proven that the SVM with PSO gives better performance than the SVM standard.



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

How to cite this article:

S. Anam, M. R. A. Putra, Z. Fitriah, I. Yanti, N. Hidayat and D. M. Mahanani, "HEALTH CLAIM INSURANCE PREDICTION USING SUPPORT VECTOR MACHINE WITH PARTICLE SWARM OPTIMIZATION," *BAREKENG: J. Math. & App.*, vol. 17, iss. 2, pp. 0797-0806, June, 2023.

Copyright © 2023 Author(s)

Journal homepage: <https://ojs3.unpatti.ac.id/index.php/barekeng/>

Journal e-mail: barekeng.math@yahoo.com; barekeng_journal@mail.unpatti.ac.id

Research Article • **Open Access**

1. INTRODUCTION

Insurance is able to reduce or eliminate the losses cost resulted by some risks. Insurance is needed to protect individuals and organizations from risk [1]. Currently, there are many types of insurance on the market, such as health insurance, vehicle insurance, educational insurance, etc. Health insurance is one of the most favorite products in the insurance industry recently. Health insurance is needed to protect insurance customers from various risks of illness to accidents. Health insurance is necessary because coverage helps insurance customers for getting medical care and improving their lives and health [2]. Customers of health insurance can propose an insurance claim for service in medical care. Insurance claim is a request from a policyholder to an insurance company for coverage for a covered loss. Several health insurance companies suffered a loss due to the large number of claims submitted. Predicting the number of claims is an important task in insurance industry. The number of claims is an important factor in determining the profit achievement of health insurance companies [3]. The implication of the claim number prediction is very important for managers, financial experts, and underwriting in the health insurance industry [4]. An increase in the number of claims will directly increase the company's total expenses, thus affecting the profit margins generated by the insurance company. Prediction of the number of claims appropriately and accurately can be used considered to prepare the annual financial budget of the insurance company [5] and to determine the amount of premium that must be paid by insurance users. Therefore, the prediction of claim submission by insurance customers in that year needs to be done by insurance companies.

Claim prediction of health insurance customers can be done by using machine learning methods. Claim prediction can be solved by the classification methods, such as the Naïve Bayes method, DT, ANN, and SVM. The results of previous studies show that the solution of SVM is global and unique [6]. SVM produces good results even if the data information is not complete. It is able to work with unstructured data and solve complex or nonlinear problems by using a kernel function. SVM also works well in high dimensional data [7]. SVM has been implemented successfully in many applications, such as covid-19 prediction [8], leaf diseases detection [9], [10], stock price prediction [11], handwriting classification [12], bank fraud detection [13], credit card fraud detection [14], face recognition [15], and many other applications. However, the SVM performance is determined by the parameters and the features [16]. Parameter selection from SVM is usually done by trial and error so that the performance is less than optimal. For this reason, parameter selection of SVM becomes an important task to obtain the best performance of SVM.

To overcome these problems in SVM, an optimization algorithm is utilized to find the best parameter values [17]. Some heuristic optimization algorithms can be used for optimizing the SVM parameters. Heuristics optimization methods have few or no assumptions about problem that be optimized. Besides that, they have the ability to get the optimal solution or near optimal solution in the large spaces with an acceptable computational cost [18]. Examples of heuristics optimization methods are PSO and GA. Specifically, PSO belongs to the Swarm Intelligence (SI) group. SI has some advantages. They have the ability to obtain the global optimum and don't need the derivatives calculation. Besides that, they are robust, easy to be implemented [19]. One of the SI methods is PSO. All the particles in the PSO are convergent quickly to the global optimal and near-optimal position [20]. Some researches show that PSO is better than GA. For these reasons, this article proposes the health claim insurance prediction using SVM with PSO.

2. RESEARCH METHODS

This article proposes the health claim insurance prediction using SVM with PSO. The proposed method is divided into several steps, which are shown in the flowchart in **Figure 1**. These steps are the input of data and parameters of PSO, data pre-processing, split the data, create a model classification for health claim insurance prediction using SVM with PSO, and evaluate the model. The output of the proposed method is a classification model for health claim insurance prediction.

2.1 Input data and PSO parameters

This research uses several variables/features for developing the model classification for health claim insurance prediction, which are age, sex, Body Mass Index (BMI), the number of children, smokers, region charges, and insurance claim. For tuning the parameters of SVM, the PSO algorithm is used. The PSO algorithm needs some parameters before the PSO algorithm is run. The PSO has several parameters, which

are cognitive and social parameters c_1 and c_2 , and an inertia weight w [21]. This experiment uses several the number of particles in the experiments, which are 5, 10, 20 and 50. This research uses 1000 iterations for the maximum iterations.

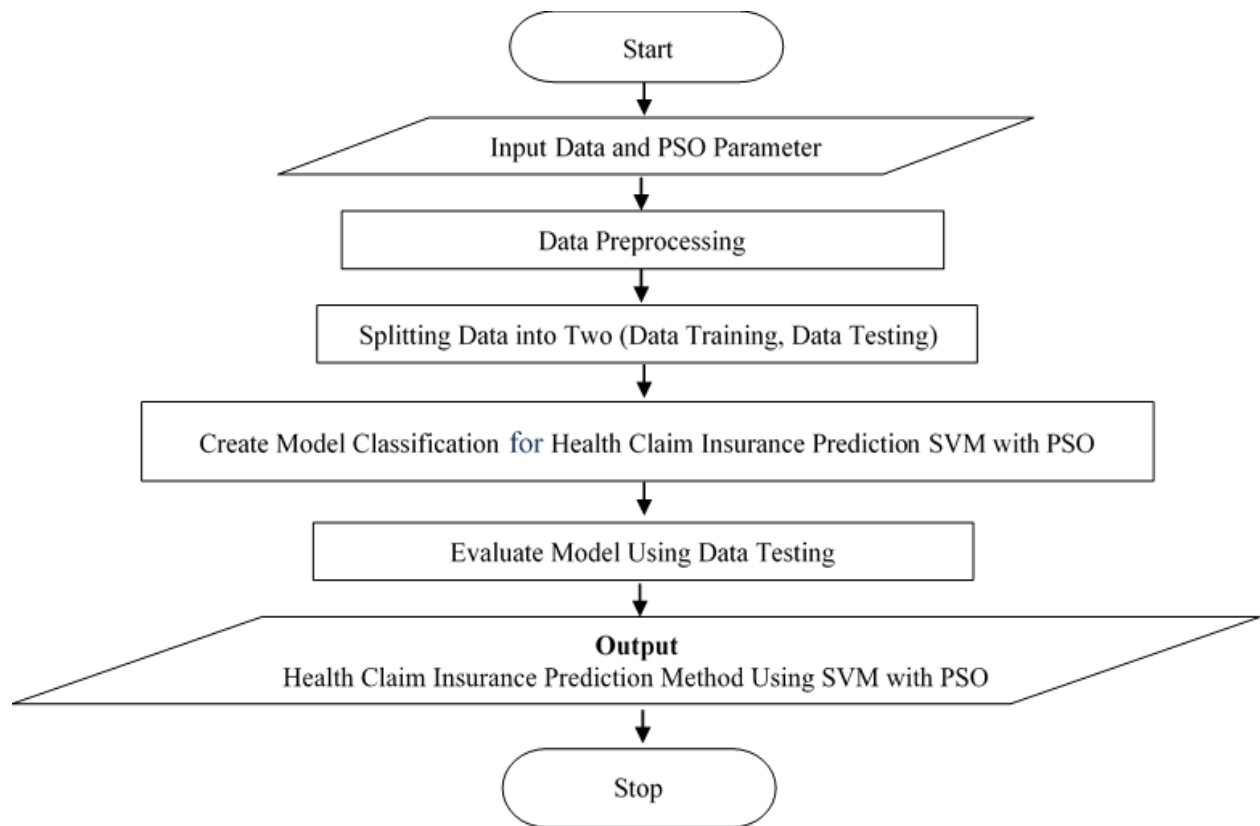


Figure 1. Flowchart of the health claim insurance prediction using support vector machine with particle swarm optimization

2.2 Data pre-processing

Data pre-processing is one of the stages in data mining. Before moving on to the next stage of processing, the raw data will be processed first. Data pre-processing is usually done through the elimination of inappropriate data. In addition, in this process, the data will be changed in a form that will be better understood by the system. Another understanding states that data pre-processing is a step to eliminate some of the problems that can interfere when submitting data. This is because a lot of data are in inconsistent format. Data normalization is commonly used for preparing the data before the data mining technique is applied. In machine learning and data mining, this process is used to convert the numerical values in the dataset to be the same scale. Data normalization usually uses a small scale, such as -1 to 1 or 0 to 1. This is generally useful for classification algorithms.

Data normalization techniques in data mining are very helpful because they provide many advantages. The advantages are that the data mining method becomes easier to be applied, faster, more effective, and efficient. The normalized data needed to be done for analyzing by certain methods.

2.3 Splitting Data

The dataset is necessary to be divided into training and test sets. This is one of important steps in developing a data mining model. The training process uses most of the data. While the testing process uses a smaller part of the data. The composition of the training data and testing data can vary, generally 80% and 20%, or 70% and 30%. Between subsets of data should not overlap because it will damage the training model process. The machine learning model is obtained by learning the training data. After learning, the performance of model is evaluated by using testing data. The testing process is also used to know the overfitting, underfitting, or appropriate-fitting. This research uses 70% training and 30% testing.

2.4 Create Model Classification for Health Claim Insurance Prediction SVM with PSO

The next step is to create a classification model for health claim insurance prediction by using SVM with PSO. The basic principle of SVM is the usage of a linear classifier. The classification cases are able to be solved by separating them linearly. However, SVM has been applied for solving the nonlinear problems. The nonlinear problem can be solved by using the kernel concept. A hyperplane in the high-dimensional space of the SVM method is obtained by maximizing the distance between data classes. The classification model is formed using SVM, where the parameters of SVM, namely c and γ are searched using the PSO algorithm. The classification model was created by carrying out the learning process in training. The steps to obtain a classification model from health claim insurance prediction by using SVM with PSO are shown in the flowchart of **Figure 2**. Particles in SVM with PSO declare candidates from the optimum parameters of SVM, which are c and γ . In the first step of the SVM method with PSO, a number of candidates s of SVM parameters is generated randomly, which are represented as $x_i(0), i = 1, 2, \dots, s$. Furthermore, SVM with PSO explores and exploits the search space to obtain the global optimum point. The SVM method with PSO uses the fitness function, which is obtained by calculating the f_1 score of training SVM on data training which has formula $1/(0.001 + f_1 \text{ Score})$. For exploring and exploiting the search space, every particle tries to find the best position (y_i). The updating of best position is calculated by **Equation (1)**.

$$y_i(t+1) = \begin{cases} y_i(t), & f(x_i(t+1)) \geq f(y_i(t)) \\ x_i(t+1), & f(x_i(t+1)) < f(y_i(t)) \end{cases} \quad (1)$$

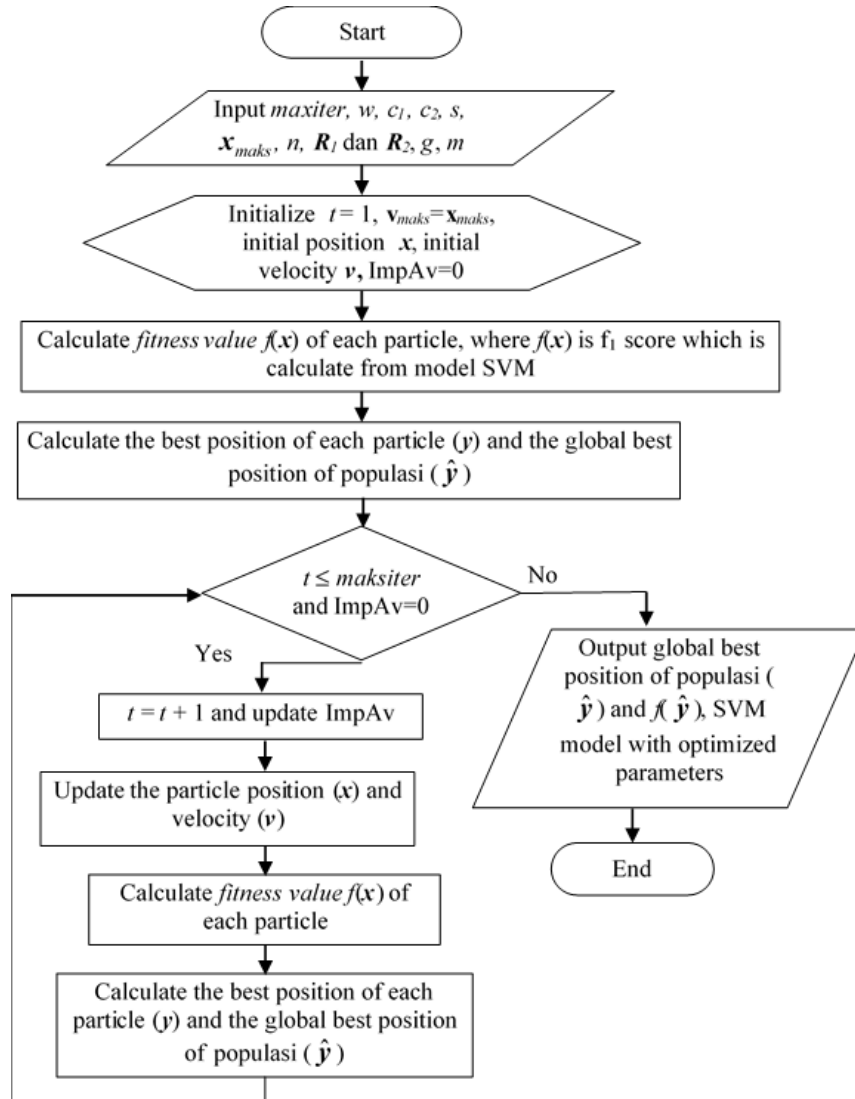


Figure 2. Flowchart of the SVM with PSO

For updating the new position of particles, the global best population is necessary to be calculated, the formulation of global best position represents in **Equation (2)**.

$$\hat{\mathbf{y}} = \operatorname{argmin}_{\mathbf{y} \in G} f(\mathbf{y}), G = \{\mathbf{y}_1(t), \dots, \mathbf{y}_s(t)\} \quad (2)$$

The particle velocity in PSO is calculated by **Equation (3)**. Therefore, it is used for updating the position of particle by calculating **Equation (4)**.

$$\mathbf{v}_i(t+1) = w \cdot \mathbf{v}(t) + c_1 R_1(t)(\mathbf{y}_i(t) - \mathbf{x}_i(t)) + c_2 R_2(t)(\hat{\mathbf{y}}_i(t) - \mathbf{x}_i(t)) \quad (3)$$

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \mathbf{v}_i(t+1) \quad (4)$$

w represents the inertia weight, and c_1 and c_2 represent the social parameters 1 and 2, respectively. $R_1(t)$ and $R_2(t)$ are defined by generating the random numbers $[0,1]$. When a global optimum solution is reached, the SVM with PSO will be stopped. The global optimum point is assumed to be found when the fitness improvement of global best is not happened in 100 iterations. After training the SVM method with PSO, the parameters obtained by PSO are used for classification model.

2.4 Evaluate the model classification

The next step is to evaluate the model that has been obtained by conducting training on the training data and using the parameters obtained with PSO. Next, the model is tested on the test data. The f_1 score, accuracy, recall, and precision in the training and testing data are evaluation metrics which is used to decide the performance classification model.

3. RESULTS AND DISCUSSION

This part will discuss the dataset for evaluating the SVM method with PSO. The results of the experiment also will be discussed in this section. Furthermore, the results of the experiment will be described, investigated, and discussed. Since the solution produced by the SVM method with PSO may be different in every experiment, the repetition of experiments is necessary. Each experiment in this research is repeated 25 times. The average and deviation standard of f_1 score, accuracy, recall, precision, and computational time produced by the methods are used to determine the spread of the values generated by the methods.

Table 1. The example of dataset

No	Age	Sex	BMI	Number of Children	Smoker	Region	Charges	Class of Insurance Claim
1	19	0	27.9	0	1	3	16884.92	1
2	18	1	33.77	1	0	2	1725.552	1
3	28	1	33	3	0	2	4449.462	0
4	33	1	22.705	0	0	1	21984.47	0
5	32	1	28.88	0	0	1	3866.855	1
...

3.1 Dataset for the Evaluation of Proposed Method

To evaluate the classification model for health insurance claim prediction, it is used 1338 data which were taken from Kaggle.com. The dataset has several attributes, which are age, sex, BMI, the number of children, smokers, region charges, and insurance claim. **Table 1** shows the example of dataset. It can be shown that the range of value of each variable is very big different. For this reason, the min max normalization method needs to be applied so that the machine learning methods can work well.

3.2 Analysis of the experimental results

This section discusses the experimental results. Furthermore, the experimental results will be analyzed in this section. **Table 2** shows the dataset after applying the min max normalization. It can be shown that all data is in the same range [0,1]. This process aims to make the SVM method able to work efficiently. Then, the dataset is separated into two subsets which are training and testing data. The percentage of the number of data used for training and testing is 70% and 30%.

Table 3 shows the evaluation of the classification model to predict claims on health insurance users using SVM with PSO. From the experimental results, it can be shown that the performance of the SVM method with PSO is a lot better than the performance of the standard SVM for most of the evaluation metrics. It can be seen from **Table 3** that the values of accuracy, precision, and f_1 score produced by the SVM method with PSO is higher than the values of accuracy, precision, and f_1 score produced by the standard SVM. Although the recall generated by the proposed method is smaller, it is not too different from the standard SVM method. The recall difference is only about 1%. The experimental results also show that the number of particles 5, 10, 20, and 50 produces no different performance. For this reason, the recommended number of particles is 5.

Table 2. The example of dataset after min max normalization

No	Age	Sex	BMI	Number of Children	Smoker	Region	Charges	Class of Insurance Claim
1	0.02	0	0.3212	0	1	1	0.251611	1
2	0	1	0.4792	0.2	0	0.666667	0.009636	1
3	0.22	1	0.4584	0.6	0	0.666667	0.053115	0
4	0.33	1	0.1815	0	0	0.333333	0.33301	0
5	0.3	1	0.3476	0	0	0.333333	0.043816	1
...

Table 3. The average of the metric evaluation of SVM and SVM with PSO for health insurance claim prediction

Method	Accuracy		Recall		Precision		f_1 score	
	Training Data	Testing Data	Training Data	Testing Data	Training Data	Testing Data	Training Data	Testing Data
SVM with PSO (Number of Particles =5)	0.934	0.876	0.912	0.890	0.972	0.900	0.941	0.895
SVM with PSO (Number of Particles =10)	0.934	0.876	0.913	0.890	0.972	0.900	0.942	0.894
SVM with PSO (Number of Particles =20)	0.935	0.877	0.913	0.890	0.972	0.900	0.942	0.895
SVM with PSO (Number of Particles =50)	0.935	0.877	0.914	0.890	0.973	0.900	0.942	0.895
SVM with $c=10^{0.2}$, $\gamma=10^{0.1}$	0.895	0.861	0.921	0.903	0.901	0.867	0.911	0.885
SVM with $c=10^{0.6}$, $\gamma=10^{0.7}$	0.918	0.881	0.916	0.903	0.941	0.896	0.928	0.899

Table 4. The deviation standard of the metric evaluation of SVM and SVM with PSO for health insurance prediction

Method	Accuracy		Recall		Precision		f_1 score	
	Training Data	Testing Data	Training Data	Testing Data	Training Data	Testing Data	Training Data	Testing Data
SVM with PSO (Number of Particles =5)	0.0005	0.0021	0.0008	0.0021	0.0007	0.0021	0.0004	0.0018
SVM with PSO (Number of Particles =10)	0.0005	0.0030	0.0007	0.0032	0.0009	0.0024	0.0005	0.0026
SVM with PSO (Number of Particles =20)	0.0005	0.0022	0.0005	0.0021	0.0007	0.0020	0.0004	0.0018
SVM with PSO (Number of Particles =50)	0.0003	0.0022	0.0000	0.0023	0.0005	0.0020	0.0002	0.0019
SVM with $c=10^{0.2}$, $\gamma=10^{0.1}$	-	-	-	-	-	-	-	-
SVM with $c=10^{0.6}$, $\gamma=10^{0.7}$	-	-	-	-	-	-	-	-

Table 5. The average and deviation standard of the computational time of SVM and SVM With PSO for health insurance claim prediction (testing data)

Method	Avg. of Computational Time	Std. of Computational Time
SVM with PSO (Number of Particles =5)	21.991	3.970
SVM with PSO (Number of Particles =10)	34.807	0.993
SVM with PSO (Number of Particles =20)	63.202	1.758
SVM with PSO (Number of Particles =50)	146.262	1.975
SVM with $c=10^{0.2}$, $\gamma=10^{0.1}$	0.0769	-
SVM with $c=10^{0.6}$, $\gamma=10^{0.7}$	0.0614	-

Table 4 shows the standard deviation of the SVM method with PSO in predicting health insurance claims. The resulting standard deviation is very small for all combinations of particle numbers. This means that the SVM method with PSO is able to work well because it produces an optimum value with a small variation. **Table 5** shows a comparison of the computational time of the SVM method with PSO and the standard SVM. The computation time of the SVM method with PSO for health insurance claim prediction is greater than the standard SVM method. While **Figure 3** shows a graph of the relationship between iterations and the fitness value. PSO converges on a few iterations. The graph shows a convergent PSO of about 4 iterations for the number of particles 5, 10, 20, and 50, so it is recommended to use a small number of iterations so that the computation time can be reduced.

4. CONCLUSIONS

The experimental results show that the SVM method with PSO gives the great performance in the health claim insurance prediction. The SVM method with PSO for predicting claims on health insurance is superior to standard SVM on 3 evaluation metrics used from 4 evaluation tools, but the computation time required is longer.

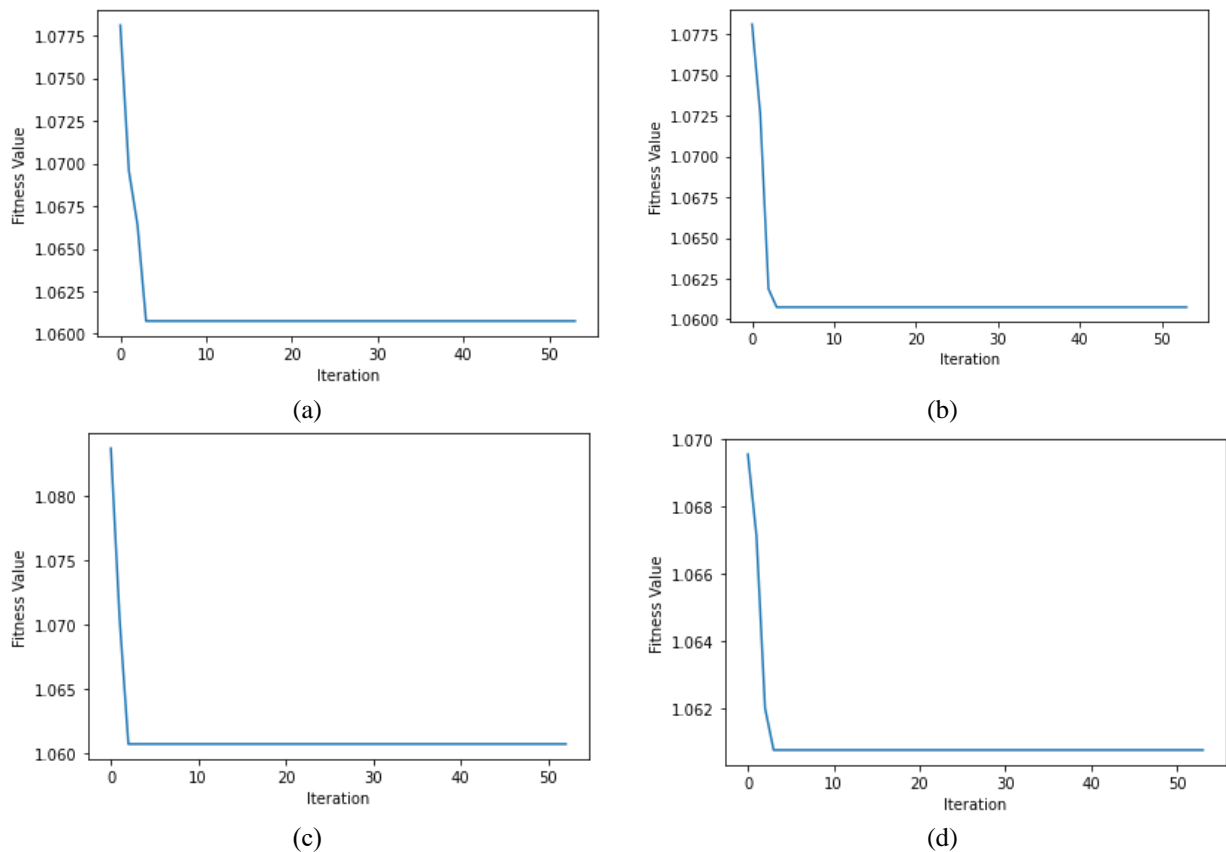


Figure 3. Graph of the relationship between iterations and fitness value. (a) 5 particles, (b) 10 particles, (c) 20 particles, and (d) 50 particles.

ACKNOWLEDGMENT

The authors would like to thank Mathematical Optimization and Computing Research Group, University Brawijaya to support this research.

REFERENCES

- [1] M. Hanafy and O. M. A. Mahmoud, "Predict Health Insurance Cost by using Machine Learning and DNN Regression Models," *Int. J. Innov. Technol. Explor. Eng.*, vol. 10, no. 3, pp. 137–143, 2021, doi: 10.35940/ijitee.c8364.0110321.
- [2] R. R. Bovbjerg and J. Hadley, "Why Health Insurance is Important.," *Heal. Policy Briefs*, no. 11, pp. 1–3, 2007, [Online]. Available: http://www.urban.org/UploadedPDF/411569_importance_of_insurance.pdf
- [3] T. Hwang and B. Greenford, "Across-Section Analysis of the Determinants of Life Insurance Consumption in Mainland China, Hong Kong, and Taiwan," *Risk Manag. Insur. Rev.*, vol. 8, no. 1, pp. 103–125, 2005.
- [4] M. Bärthel and S. Krummacker, "Prediction of Claims in Export Credit Finance: a Comparison of Four Machine Learning Techniques," *Risks*, vol. 8, no. 1, pp. 1–27, 2020, doi: 10.3390/risks8010022.
- [5] S. Goundar, S. Prakash, P. Sadal, and A. Bhardwaj, "Health Insurance Claim Prediction Using Artificial Neural Networks," *Int. J. Syst. Dyn. Appl.*, vol. 9, no. 3, pp. 40–57, 2020, doi: 10.4018/ijdsda.2020070103.
- [6] D. L. Olson and D. Delen, *Advanced Data Mining Techniques*, 1st ed. Berlin: Springer International Publishing, 2008.
- [7] N. Çolakoğlu and B. Akkaya, "Comparison of Multi-class Classification Algorithms on Early Diagnosis of Heart Diseases," 2019, no. March, pp. 162–171.
- [8] D. N. Avianty, Wijaya, I. G. P. S. Wijaya, and F. Bimantoro, "The Comparison of SVM and ANN Classifier for COVID-19 Prediction," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 13, no. 2, p. 128, 2022, doi: 10.24843/lkjiti.2022.v13.i02.p06.
- [9] B. Kumar, R. Bharath Kempegowda, M. S. Shreyas, S. Sonu, H. Anirudh, and B. J. Abhishek, "SVM Based Plant Diseases Detection using Image Processing," *Int. J. Comput. Sci. Eng.*, vol. 7, no. 5, pp. 1263–1266, 2019, doi: 10.26438/ijcse/v7i5.12631266.
- [10] M. Hussein and A. H. Abbas, "Plant Leaf Disease Detection Using Support Vector Machine," *Al-Mustansiriyah J. Sci.*, vol. 30, no. 1, pp. 105–110, 2019, doi: 10.23851/mjs.v30i1.487.
- [11] N. Tripathy, "Stock Price Prediction Using Support Vector Machine Approach," in *International Academic Conference on Management & Economics*, 2019, pp. 44–59. doi: 10.33422/conferenceme.2019.11.641.
- [12] A. A. Hasseim, R. Sudirman, and P. I. Khalid, "Handwriting Classification Based on Support Vector Machine with Cross Validation," *Engineering*, vol. 05, no. 05, pp. 84–87, 2013, doi: 10.4236/eng.2013.55b017.

- [13] N. K. Gyamfi and J. D. Abdulai, "Bank Fraud Detection Using Support Vector Machine," in *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2018*, 2019, no. May 2019, pp. 37–41. doi: 10.1109/IEMCON.2018.8614994.
- [14] D. Zhang, B. Bhandari, and D. Black, "Credit Card Fraud Detection Using Weighted Support Vector Machine," *Appl. Math.*, vol. 11, no. 12, pp. 1275–1291, 2020, doi: 10.4236/am.2020.1112087.
- [15] Z. Rustom and A. A. Ruvita, "Application Support Vector Machine on Face Recognition for Gender Classification," *J. Phys. Conf. Ser.*, vol. 1108, no. 1, 2018, doi: 10.1088/1742-6596/1108/1/012067.
- [16] S. Sulistiana and M. A. Muslim, "Support Vector Machine (SVM) Optimization Using Grid Search and Unigram to Improve E-Commerce Review Accuracy," *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 8–15, 2020, doi: 10.52465/jossex.v1i1.3.
- [17] W. M. Czarniecki, S. Podlewska, and A. J. Bojarski, "Robust Optimization of SVM Hyperparameters in the Classification of Bioactive Compounds," *J. Cheminform.*, vol. 7, no. 1, pp. 1–15, 2015, doi: 10.1186/s13321-015-0088-0.
- [18] W. Dubitzky, O. Wolkenhauer, K.-H. Cho, and H. Yokota, Eds., *Encyclopedia of Systems Biology*. London: Springer International Publishing, 2013. doi: 10.1007/978-1-4419-9863-7.
- [19] A. K. Kordon, "Swarm Intelligence: The Benefits of Swarms," in *Applying Computational Intelligence*, 1st ed., no. October 2001, Berlin: Springer International Publishing, 2010, pp. 145–174. doi: 10.1007/978-3-540-69913-2.
- [20] J. A. Flores-Livas, L. Boeri, A. Sanna, G. Profeta, R. Arita, and M. Eremets, "A Perspective on Conventional High-temperature Superconductors at High Pressure: Methods and Materials," *Phys. Rep.*, vol. 856, pp. 1–78, 2020, doi: 10.1016/j.physrep.2020.02.003.
- [21] Y. He, W. J. Ma, and J. P. Zhang, "The Parameters Selection of PSO Algorithm influencing on Performance of Fault Diagnosis," in *2016 International Conference on Mechatronics, Manufacturing and Materials Engineering*, 2016, vol. 63, no. 2016, pp. 1–5. doi: 10.1051/mateconf/20166302019.

