BAREKENG: Journal of Mathematics and Its Applications
December 2023 Volume 17 Issue 4 Page 2281-2292
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss4pp2281-2292

DEVELOPMENT OF HEALTH INSURANCE CLAIM PREDICTION METHOD BASED ON SUPPORT VECTOR MACHINE AND BAT ALGORITHM

Syaiful Anam[1*], Abdi Negara Guci[2], Fery Widhiatmoko[3], Mila Kurniawaty[4], Komang Agus Arta Wijaya[5]

[1]*Computer and Data Science Laboratory, Mathematics Department, Brawijaya University*
[1,3,4,5] *Actuarial Science Study Program, Mathematics Department, Brawijaya University*
[1,2] *Mathematics Study Program, Mathematics Department, Brawijaya University*
Malang, 65145, Indonesia

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

ABSTRACT

Article History:

Received: 26th July 2023
Revised: 11th October 2023
Accepted: 10th November 2023

Keywords:

Claim Prediction;
HEALTH INSURANCE;
Support Vector Regression;
Bat Algorithm;
Parameter Selection.

Health insurance industry is very much needed by the community in handling the financial risks in the health sector. The number of claims greatly affects the achievement of profits and the sustainability of the health insurance industry. Therefore, filing claims by insurance users from year to year is important to be predicted in insurance firm. The Machine Learning (ML) method promises to be a good solution for predicting health insurance claims compared to conventional data analytics methods. Support Vector Machine (SVM) is one of the superior ML approaches. Nonetheless, SVM performance is controlled by the suitable selection of SVM parameters. The SVM parameters is typically selected by trial and error, sometimes resulting in not optimal performance and taking a long time to complete. Swarm intelligence-based algorithms can be used to select the best parameters from SVM. This method is capable of locating the global best solution, is simple to implemented, and doesn't involve derivatives. One of the best swarm intelligence algorithms is the Bat Algorithm (BA). BA has a faster convergence rate than other algorithms, for example Particle Swarm Optimization (PSO). Based on this situation, this paper offers the new classification model for predicting health insurance claim based on SVM and BA. The metrics utilized for evaluation are accuracy, recall, precision, f_1-score, and computing time. The experimental outcomes show that the proposed approach is superior to the conventional SVM and the hybrid of SVM and PSO in forecasting health insurance claims. In addition, the proposed method has a substantially shorter computing time than the hybrid of SVM and PSO. The outcomes of the experiments also indicate that the new classification model for predicting health insurance claim based on the SVM and BA can avoid over-fitting condition.

How to cite this article:

S. Anam, A. N. Guci, F. Widhiatmoko, M. Kurniawaty, K. A. A. Wijaya and Z. Fitriah, "DEVELOPMENT OF HEALTH INSURANCE CLAIM PREDICTION METHOD BASED ON SUPPORT VECTOR MACHINE AND BAT ALGORITHM," *BAREKENG: J. Math. & App.*, vol. 17, iss. 4, pp. 2281-2292, December, 2023.

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

Individuals or organizations often experience losses due to risk. One of the causes of risk loss is disaster, for example floods, disease outbreaks, earthquakes, volcanic eruptions and so on. Natural disasters can have an impact on human health risks. Losses due to risks can be reduced or eliminated by taking insurance [1], for example financial risks to individual health. The health insurance industry is critical to the community's ability to overcome the financial hazards in the healthcare area. Health insurance has recently become one of the most popular choices for protection. Health insurance companies assist insurance customers in medical treatment by covering the costs of treatment, either in full or in part through filing insurance claims [2]. Insurance claims are requests from policyholders to insurance companies to cover losses incurred. The sustainability of the health insurance industry depends on the number of claims filed by insurance users. Due to the enormous number of claims filed, several health insurance firms faced losses. The number of claims is a crucial factor in deciding how profitable health insurance businesses [3]. Forecasting the quantity of claims is a critical challenge in the insurance business, particularly for health insurance leaders, financial specialists, and underwriters [4]. Precise and accurate prediction of the quantity of claims is necessary for consideration in preparing the insurance company's annual financial budget [5], and for determining the amount of premiums to be paid by insurance users. As a result, insurance firms must forecast the quantity of claims filed by insurance customers.

The claim prediction problem is an insurance business problem that is difficult to model mathematically, making it difficult to solve using traditional analytic approaches, such as regression and linear programming [6]. The ability of the machine learning method to solve complex problems and difficult to model mathematically holds promise in solving the problem of predicting health insurance customer claims. Various Machine Learning (ML) approaches have been widely used in numerous challenging issues. Several researchers have applied machine learning methods to predict insurance claims. Quan and Valdez proposed a multivariate Decision Tree (DT) predicting study of insurance claims [7]. Bhardwaj in 2020 has also proposed predicting health insurance claims using Artificial Neural Networks (ANN) [5]. Alamir et al. in 2021 have also proposed predicting motor vehicle insurance claims using Support Vector Machine (SVM) and Random Forest (RF), with RF and SVM performance that are not much different [8].

Some research investigations have been performed to evaluate the accuracy of various ML approaches for problem solving. Imaduddin et al. have shown that SVM is superior to DT in breast cancer classification [9]. Kusumawati et al. also have shown that SVM is superior to Naïve Bayes for the classification of Tokopedia services [10]. Zanaty in 2012 also has shown that SVM is superior to ANN Multilayer Perceptron (MLP) in data classification [11]. The results of research on SVM show that SVM solutions are global and unique [12]. SVM is able to work well on incomplete data. SVM can work with unstructured data. The results also show that SVM is capable of solving complex or nonlinear problems. SVM additionally proves effective on data with large dimensions[13]. SVM has also been used satisfactorily in a variety of applications, including covid-19 forecasting [14], identifying leaf diseases [15] [16], forecasting stock prices [17], classifying handwriting[18], identifying fraudulent banking [19], identifying fraudulent credit cards[20], recognition of faces[21], and numerous others. Based on the results of the previous research, this research will develop the SVM method to predict health insurance claims.

However, SVM performance depends on the parameters and features selected [22]. SVM choosing parameters is usually performed by trial and error, resulting in less-than-optimal results. Optimization techniques may be employed when selecting the best parameter values in SVM [23]. The global optimization algorithm offers the best parameter selection of the SVM method. The primary benefit of the global optimization approach is that it requires few or no assumptions about the issue to be improved. Furthermore, the global optimization method has the ability of producing optimum or nearly-optimal results in huge areas with reasonable computing costs [24]. PSO, GA, and BA are a few examples of heuristics optimization algorithms. Particularly, PSO and BA include in the Swarm Intelligence (SI) group. This technique can identify a global optimum from numerous local optimums, it doesn't need derivatives, it is resilient, and it is simple to apply [25].

SI-based algorithms have been proposed by several researchers. This algorithm has been used to overcome the weaknesses of machine learning algorithms. Anam et al. has utilized an algorithm based on swarm intelligence (PSO) to segment disease on tomato leaves [26]. Anam et al. also has proposed PSO for finding the parameters of SVM. They use it for predicting the health insurance claim [27]. This study demonstrates that the method works well. However, it needs the expensive computational time. The BA is another swarm intelligence method. Yang presented the technique in 2010, which is an artificial intelligence algorithm influenced by bat echolocation activity. Echolocation refers to bats' capacity to generate waves in order to navigate, discover prey, and avoid obstacles in totally dark surroundings. BA has the ability to find points automatically to regions where potential solutions exist, which is one of its many benefits. As a consequence, the Bat Algorithm has a faster convergence rate than other algorithms, such as the GA and PSO[28].

Based on the previous paragraphs, this paper offers a new classification model for predicting health insurance claims based on SVM and BA. It is hoped that the method developed will be used for further purposes, for example, determining the amount of premiums charged to health insurance customers. BA in this study is used to select the right parameters of the SVM method so that SVM produces optimal performance.

2. RESEARCH METHODS

This paper offers a new classification model for predicting health insurance claim method by utilizing SVM and BA. The new method consists of several steps, as seen in the flowchart in Figure 1. These involve the input data and configuring the BA parameters, pre-processing the data, separating the data, developing a model classification using SVM and BA, and evaluating the model. The method's goal is to build a new classification model for predicting health claim insurance.

2.1 Inputting Data and Configuring BA Parameters

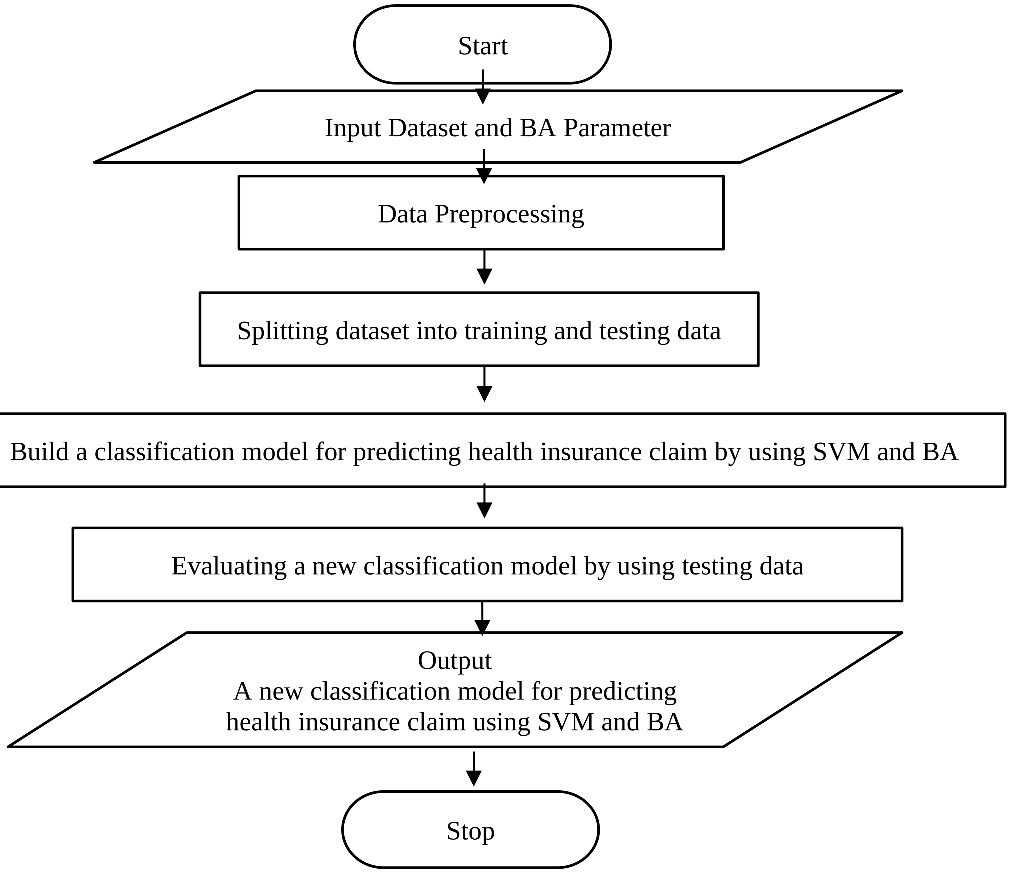


Figure 1. Flowchart of a New Classification Model for Predicting Health Insurance Claim by Combining SVM and BA.

Some of the factors / characteristics used in this study to construct the classifier model to predict health claim insurance include age, gender, Body Mass Index (BMI), the quantity of children, smokers, area costs, and insurance claim. BA is employed to fine-tune the SVM's parameters. Before running the BA algorithm, the BA requires parameters to be set. There some parameters of BA should be set which are an iteration maximum (t_{max}), an initial loudness (A), an initial pulse rate (r_0), a parameter alpha (α), a parameter gamma

(γ), a frequency maximum ($f_{\max}$) and a frequency minimum ($f_{\min}$). The values of BA parameter which is used in this experiment are A=1, r_0=1, $\alpha = 0.97$, $\gamma = 0.1$, $f_{\max}$ =2 and f_{min}=0[28]. The numbers of bat population in the experiments are 5, 10, 20 and 50. The termination condition of the algorithm is the maximum iteration or the stagnation condition. The iterations maximum number is 1000 iteration. If the global best of bat doesn't improve in 20 iterations, the algorithm will be stopped although the maximum iteration hasn't yet reached.

2.2 Data pre-processing

One of the data mining process stages is data pre-processing. The data pre-processing needs to prepare the raw data so that it is ready to be processed. The removal of irrelevant data is typically how data pre-processing is accomplished. The data will also be modified during this process so that the system can understand it better. According to another view, data pre-processing is a procedure made to alleviate various concerns that may develop while sending data. This is due to the inconsistency of numerous data formats. Data normalization is typically used prior to utilizing a data mining approach. In ML and data mining, this approach is employed for scaling down the numerical values of attributes in the collection of data. Typically, small scales like -1 to 1 or 0 to 1 are used for data normalization. For classification algorithms, this is typically useful.

Data normalization techniques are very beneficial in data mining because they offer a number of benefits. The data mining approach has the advantages of being quicker, more effective, and more efficient to use. Data must be normalized before using certain methods of analysis.

2.3 Separating the dataset into training data and testing data

The dataset must be split into training and test sets. In order to develop a data mining model, this is one of the key steps. The majority of the data is used during the training process. A lesser amount of the data is used in the testing process. Different ratios of 80/20 or 70/30 can be used to describe the ratio of training data to testing data. Training data and testing data are not overlapping to ovoid a negative impacts training model. The ML model is built by utilizing the training data. Following the training stage, the testing data is utilized to examine the model's performance. The testing procedure is also used to determine whether a fit is excessively, inadequately, or appropriately good. This study employs 30% testing and 70% training.

2.4 Develop a model classification for health claim insurance prediction using SVM and BA

The following phase is to develop a model of classification to forecast health claim insurance by utilizing SVM and BA. The employment of a linear classifier is fundamental to SVM. The classification instances can be addressed by linearly separating them. SVM, on the other hand, has been used to tackle non-linear issues. The kernel concept can be utilized to address the nonlinear problem. By optimizing the distance between data classes, a hyperplane is formed in the SVM algorithm's space with high dimensions. The SVM is employed to build the new classification model, and the c and γ of the SVM are searched using the BA. The training procedure was used to develop the classification model. The flowchart in **Figure 2** depicts the procedure for constructing a classification model for predicting health claim insurance by utilizing SVM and BA. Bat locations in SVM with BA indicate candidates for the best SVM parameters of c and γ_{svm}. The SVM with BA in the first phase generates randomly the N candidates for the SVM parameters, which are expressed as $\boldsymbol{x}_i^0$, $i = 1,2,...,N$. In addition, SVM and BA investigate and use the area of searching in order to find the global optimal point. The calculation of the fitness function in the SVM with BA is calculated by f_1 score which is defined by equation $1/(0.001 + f_1\ score)$. The classification model's entire method for forecasting health insurance claims utilizing SVM and BA is as following.

Algorithm 1. A new classification model base SVM and BA for predicting health claim insurance

Input:

The training data ($\boldsymbol{X_{train}}$) with size of $n \times m$, n and m indicate the amount of training data and the number of features, respectively.

Some of the BA parameters

Output:

$\boldsymbol{x}_*$ represents the best solution of parameter of SVM which are c and γ_{svm}.

a. Set the bat locations and speeds to their initial values, $\boldsymbol{x}_i^0$ and $\boldsymbol{v}_i^0$,$(i = 1, 2, \ldots, N)$. Each bat location $\boldsymbol{x}_i$ represents a candidate of SVM parameters, which are $\boldsymbol{x}_i^0 = (c_i, \gamma_i)$.

b. Set the minimum frequency f_{min}, the maximum frequency f_{max}, a pulse rate r_i, and a loudness A_i

c. t=0

d. **WHILE** $(t < \ Iteration\ Maximum)$ **DO**

1. **FOR** $i = 1$ to N **DO**

i. The frequency, velocity and location of bats are updating by the formulations in **Equation (1)**, **Equation (2)**, and **Equation (3)**.

$$f_i = f_{min} + (f_{max} - f_{min})\beta, \tag{1}$$

$$\boldsymbol{v}_i^{t+1} = \boldsymbol{v}_i^t + \left(\boldsymbol{x}_i^t - \boldsymbol{x}_*\right) f_i, \tag{2}$$

$$\boldsymbol{x}_i^{t+1} = \boldsymbol{x}_i^t + \boldsymbol{v}_i^{t+1}. \tag{3}$$

ii. **if** $(rand < r_i$ **then**

Note: $rand$ is a random number which has uniform distribution $U(0,1)$.

Equation (4) is used to generate the local solutions randomly.

$$\boldsymbol{x}_{new} = \boldsymbol{x}_{old} + \sigma\epsilon_t A^{(t)} \tag{4}$$

ϵ_t shows random values generated from a normal distribution with a Gaussian shape $N(0,1)$, $A^{(t)}$ represents the average of bat loudness over time t, and σ represents the scale factor. For simplicity, it could be utilized $\sigma = 0.01$.

end if

iii. Assess fitness utilizing the f_1 score which is resulted by a classification model using training data and SVM which corresponds with parameter $\boldsymbol{x}_i^{t+1}$.

iv. **if** $(rand > A_i$ **and** $f(\boldsymbol{x}_i) < \mathcal{F}(\boldsymbol{x}_*))$ **then**

The solutions obtained by steps (i) or (ii) should be used to update the current solution.

end if

v. Using **Equations (5)** and **(6)** to increase r_i and reduce A_i.

$$A_i^{t+1} = \alpha A_i^t, \tag{5}$$

$$r_i^{t+1} = r_i^0[1 - \exp(-\gamma t)], \tag{6}$$

α has range $0 < \alpha < 1$,while and $\gamma > 0$. In accordance with Yang (2014), to ease the search procedure, it has value of α and γ can be equated by value. $\alpha = \gamma = 0.9$. Yang (2014) claims that A_i and r_i can equate the values of and, with value $\alpha = \gamma = 0.9$., to simplify the search procedure.

vi. Sort the bats in order to get the optimum solution $(\boldsymbol{x}_*)$

end for

end while

e. Output *Best* $(\boldsymbol{x}_*)$ to obtain SVM parameters.

Algorithm 2. Assessment of a new classification model by utilizing SVM and BA for predicting the health claim insurance.

Input:

The training data $(\boldsymbol{X}_{train})$ with size of $n \times m$

Parameter c and γ_{svm}

y_{train} (Training data set's class label)

Output:

Accuracy (*Ac*), Recall (*Rc*), Precision (*Pr*), f_1 *Score* (f_1Sc).

1. Train SVM Model using training data
2. Determine the label prediction y_{pred} based on SVM Model with BA.
3. Determine *Ac*, *Rc*, *Pr* and f_1Sc.

Algorithm 3. Evaluation of the new model classification by utilizing SVM and BA for predicting health claim insurance

Input:

The testing data

SVM model

$y_{testing}$ (each testing data's class labels)

Output:

Ac, *Rc*, *Pr* and f_1Sc.

1. Determine the label prediction y_{pred} based on SVM Model with BA.
2. Determine *Ac*, *Rc*, *Pr* and f_1Sc.

There is an assumption. The global optimum point has been discovered when the fitness improvement of the global best does not occur in 20 iterations. BA finds the SVM parameters, then they are used in the classification model.

2.4 Evaluate the Model Classification

The classification model needs to be evaluated by calculating the assessment metrics. The SVM parameters acquired using BA use to build the classification model. Therefore, the model is then assessed using the test data. The *Ac*, *Rc*, *Pr* and f_1Sc of training and testing data are calculated to assess the success of the classification model. These are the formulas of assessment metrics.

a. Classification Rate / Accuracy, that is computed by the approach described in **Equation (7)**.

$$Ac = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

Positive data that is appropriately observed is commonly referred to as a True Positive (*TP*). True Negative (*TN*) is the number of tuples which is classified into the negative category correctly. False Positive (*FP*) the number of tuples which is classified into the positive category incorrectly. False Negative (*FN*) is the number of tuples which is classified into the negative category. The ratio of accurate predictions to all cases considered is used to calculate accuracy.

b. Recall is defined by **Equation (8)**. Recall is a metric for determining the percentage of correctly identified positive patterns.

$$Rc = \frac{TP}{TP+FN} \tag{8}$$

c. Precision is defined by the **Equation (9)**. When comparing the entire anticipated pattern in the positive class to the accurately predicted positive pattern, precision is employed.

$$Pr = \frac{TP}{TP+FP} \tag{9}$$

d. The formula in **Equation (10)** is used to determine the f_1 Score. The f_1 score is obtained by calculating the harmonic mean of recall and precision.

$$f_1\ sc = 2 \cdot \frac{Pr \cdot Rc}{Pr + Rc} \tag{10}$$

The experimental data are examined in order to make conclusions after the evaluation metrics have been calculated.

Table 1. An Example of a Dataset For Insurance Claim Prediction

No	Age	Sex	BMI	Num. of Children	Smoker	Region	Charges	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
…	…	…	…	…	…	…	…	…

3. RESULTS AND DISCUSSION

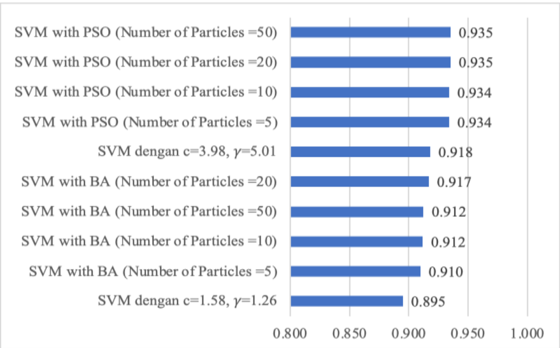
The dataset for the SVM method with BA assessment will be covered in this section. This section will also discuss the experiment's findings. The experiment's findings will also be examined, reported, and discussed. It is vital to repeat trials since the solution generated by the SVM method with BA may vary from experiment to experiment. This study repeats each experiment 25 times. The spread of the results generated by the methods is calculated using the average and deviation standard of the *Ac*, *Rc*, *Pr* and f_1Sc, as well as the processing time.

3.1 Dataset for the evaluation of proposed method

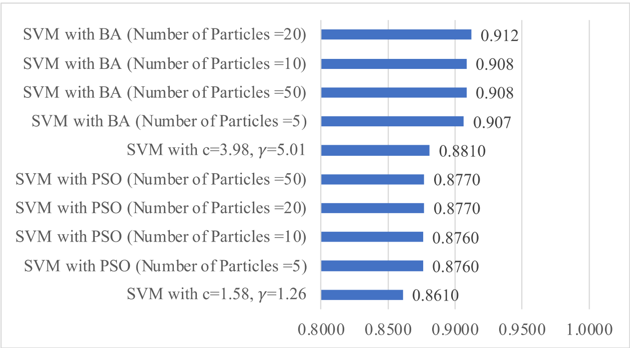
The classification model for predicting health insurance claims was evaluated using 1338 data points from Kaggle.com. Some of the features of dataset are used to predict the insurance claim. The features are age, sex, BMI, the number of children, smokers, region fees. It can be seen that the dataset example in Table 1. Each variable has various of range of values so that the min max normalization technique should be used, in order for machine learning techniques to function well.

Table 2. An Example of Dataset After Applying The Min Max Normalization

No	Age	Sex	BMI	Num of Children	Smoker	Region	Charges	Insurance Claim
1	0.02	0	0.32	0	1	1	0.252	1
2	0	1	0.48	0.2	0	0.67	0.010	1
3	0.22	1	0.47	0.6	0	0.67	0.053	0
…	…	…	…	…	…	…	…	…

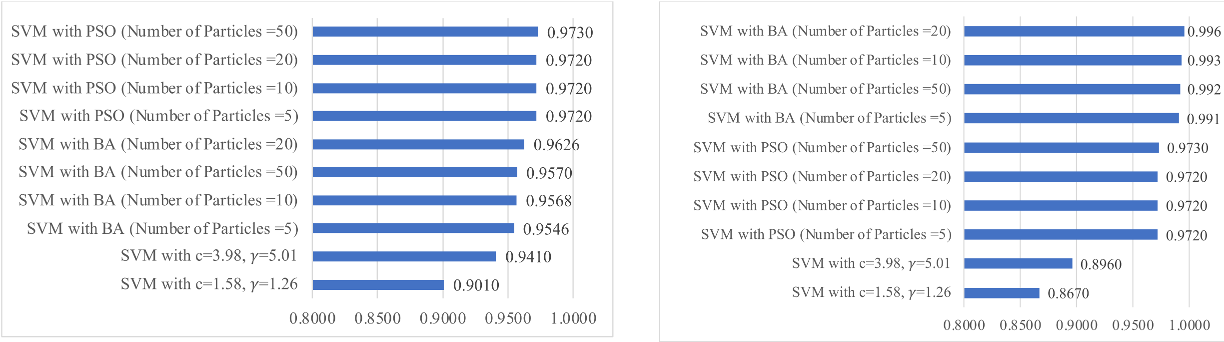


(a)



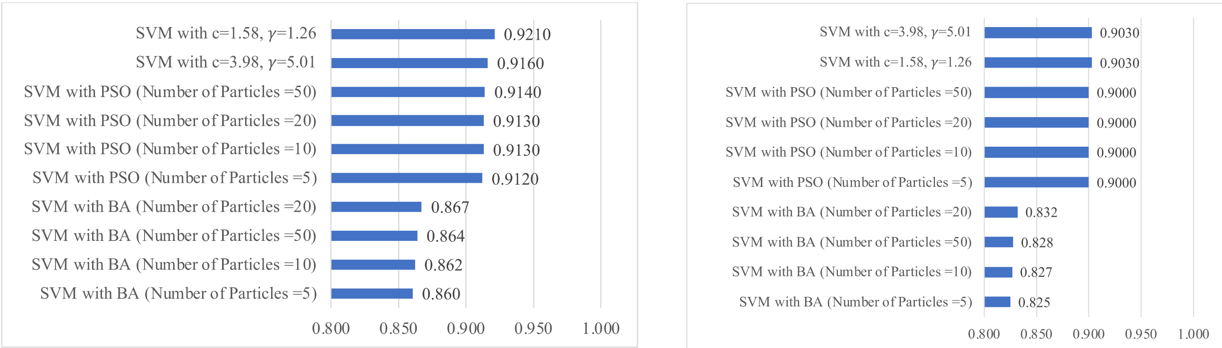
(b)

Figure 2. The Average of The Accuracy of SVM, SVM- PSO and SVM-BA for Predicting The Health Insurance Claim. (a) Training data. (b) Testing data



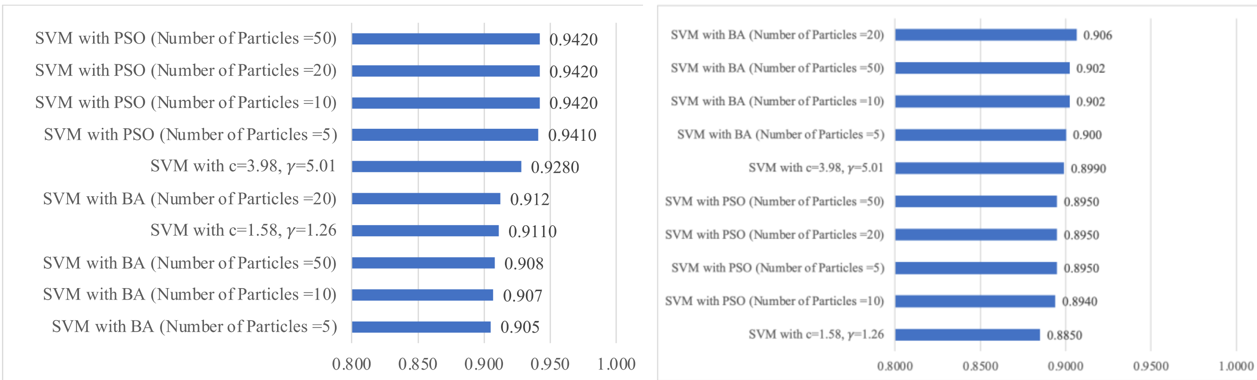
(a) **(b)**

Figure 3. The Average of the Precision of SVM, SVM-PSO and SVM-BA for Predicting The Health Insurance Claim. (a) Training data. (b) Testing data



(a) **(b)**

Figure 4. The Average of the Recall of SVM, SVM-PSO and SVM-BA for Predicting The Health Insurance Claim. (a) Training data. (b) Testing data



(a) **(b)**

Figure 5. The Average of the f_1 Score of SVM, SVM-PSO and SVM-BA for Predicting The Health Insurance Claim. (a) Training data. (b) Testing data

3.2 The Experimental Results and Discussions

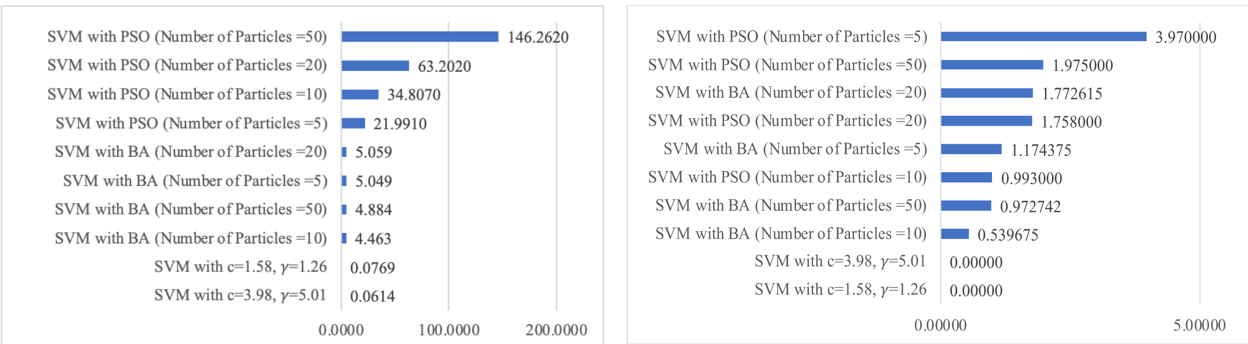
The findings of the experiment will be shown in this part. The experimental findings will also be examined. The dataset in **Table 2** is results of the min-max normalization process. It is demonstrable that all data have the same [0, 1] range. This procedure seeks to improve the SVM method's effectiveness. The dataset is divided into two subsets. The ratio of the number of training data and testing data are 70% and 30%, respectively.

The performance of the classification model for predicting health insurance by utilizing SVM, SVM with PSO, and SVM with BA is compared in **Figure 2** - **Figure 5**. According to the experimental findings, the SVM approach with BA performs significantly better than the normal SVM for the majority of evaluation criteria. Figures 2 - Figure 5 show that the Ac, Rc, Pr and f_1Sc values resulted by the SVM technique BA are greater than the Ac, Rc, Pr and f_1Sc values provided by the standard SVM and the SVM with PSO. The outcomes of the experiment also demonstrate that there is no difference in performance between 5, 10, 20, and 50 the number of bats. The suggested quantity of particles is therefore 5.

The standard deviations of the SVM with BA, the SVM with PSO dan the conventional SVM for predicting health insurance claims are shown in Table 3. For all possible the number combinations of bats and particle, the resulting standard deviation is extremely minimal. The both methods may give an optimal value with little variance, which makes it effective. The computational times for the standard SVM, the SVM-PSO and the SVM-BA could be shown in Figure 6. The SVM - BA for predicting health insurance claims requires less computation time than the SVM-PSO. While a graph of the association between iterations and the fitness value is shown in Figure 7. The graphic in Figure 7 shows a convergent BA of less than 90 iterations for all the number of bats 5, 10, 20 and 50. The bigger number of bats is recommended since generally the fitness values are more optimal than small number of bats. From Figure 7, it can also be seen that the optimality can be found faster for the bigger number of bats.

Table 3. The Deviation Standard of the Metric Performance of SVM, SVM-PSO and SVM-BA for Predicting the Health Insurance Claim

Method	Accuracy		Recall		Precision		f_1 score	
	Training Data	Testing Data	Training Data	Testing Data	Training Data	Testing Data	Training Data	Testing Data
SVM-PSO (Num. of Swarm =5)	0.0005	0.0021	0.0008	0.0021	0.0007	0.0021	0.0004	0.0018
SVM-PSO (Num. of Swarm =10)	0.0005	0.0030	0.0007	0.0032	0.0009	0.0024	0.0005	0.0026
SVM-PSO (Num. of Swarm =20)	0.0005	0.0022	0.0005	0.0021	0.0007	0.0020	0.0004	0.0018
SVM-PSO (Num. of Swarm =50)	0.0003	0.0022	0.0000	0.0023	0.0005	0.0020	0.0002	0.0019
SVM with $c=10^{0.2}$, $\gamma=10^{0.1}$	0	0	0	0	0	0	0	0
SVM with $c=10^{0.6}$, $\gamma=10^{0.7}$	0	0	0	0	0	0	0	0
SVM-BA (Number of Particles =5)	0.0104	0.0083	0.0129	0.0108	0.0104	0.0084	0.0111	0.0091
SVM-BA (Num. of Bat=10)	0.0100	0.0079	0.0129	0.0109	0.0101	0.0077	0.0108	0.0087
SVM-BA (Num. of Bat =20)	0.0058	0.0062	0.0054	0.0073	0.0078	0.0073	0.0060	0.0067
SVM-BA (Num. of Bat =50)	0.0108	0.0079	0.0124	0.0101	0.0124	0.0089	0.0114	0.0086



(a) **(b)**

Figure 6. The Computing Time of SVM, SVM-PSO and SVM-BA for Predicting Health Insurance Claim. (a) Average. (b) Deviation Standard

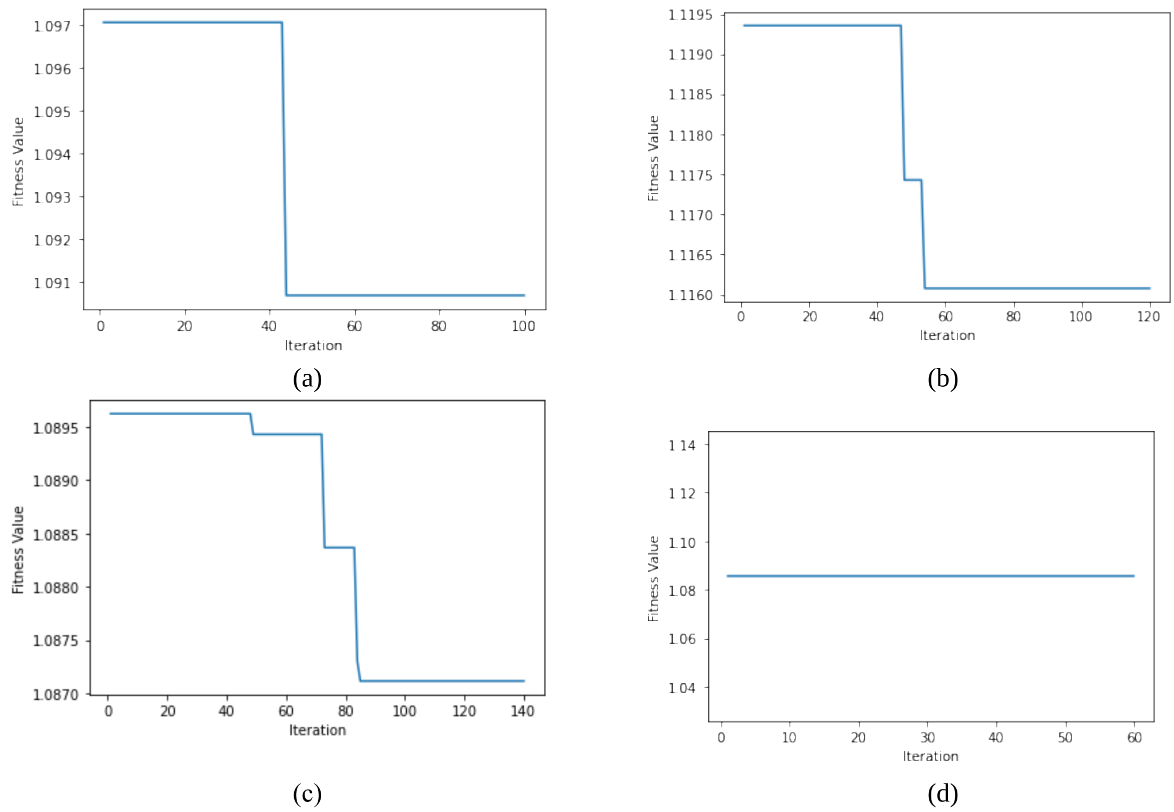


Figure 7. The Relationship Between Iterations and Fitness Value Is Depicted Graphically. (a) number of bats =5, (b) number of bats =10, (c) number of bats =20, and (d) number of bats =50.

4. CONCLUSIONS

From the experimental findings, it can be concluded that the proposed method is superior to the conventional SVM and the SVM- PSO for predicting health insurance claims. The proposed method also has much better computational time than the SVM-PSO. The experimental results also show that the health insurance claim prediction method based on the SVM-BA can avoid over-fitting condition.

ACKNOWLEDGMENT

The authors are pleased to express their gratitude to the University of Brawijaya for funding this research under the **Doktor Non Lektor Kepala Research Grant 2023**.

REFERENCES

[1] M. hanafy and O. M. A. Mahmoud, "Predict Health Insurance Cost by using Machine Learning and DNN Regression Models," *International Journal of Innovative Technology and Exploring Engineering*, vol. 10, no. 3, pp. 137–143, Jan. 2021, doi: 10.35940/ijitee.C8364.0110321.

[2] R. R. Bovbjerg and J. Hadley, "Why Health Insurance Is Important," *Health Policy Briefs*, vol. 11, pp. 1–3, 2007, [Online]. Available: http://www.iom.edu/CMS/3809/

[3] T. Hwang and B. Greenford, "Across-section Analysis of the Determinants of Life Insurance Consumption in Mainland China, Hong Kong, and Taiwan," *Risk Management and Insurance Review*, vol. 8, no. 1, pp. 103–125, 2005.

[4] M. Bärtl and S. Krummaker, "Prediction of claims in export credit finance: a comparison of four machine learning techniques," *Risks*, vol. 8, no. 1, Mar. 2020, doi: 10.3390/risks8010022.

[5] S. Goundar, S. Prakash, P. Sadal, and A. Bhardwaj, "Health Insurance Claim Prediction Using Artificial Neural Networks," *International Journal of System Dynamics Applications*, vol. 9, no. 3, pp. 40–57, Jun. 2020, doi: 10.4018/ijsda.2020070103.

[6] K. A. Smith, R. J. Willis, and M. Brooks, "An Analysis of Customer Retention and Insurance Claim Patterns Using Data Mining: A Case Study," *Source: The Journal of the Operational Research Society*, vol. 51, no. 5, pp. 532–541, 2000.

[7] Z. Quan and E. A. Valdez, "Predictive analytics of insurance claims using multivariate decision trees," *Dependence Modeling*, vol. 6, no. 1, pp. 377–407, Dec. 2018, doi: 10.1515/demo-2018-0022.

[8] E. Alamir, T. Urgessa, A. Hunegnaw, and T. Gopikrishna, "Motor Insurance Claim Status Prediction using Machine Learning Techniques," *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 12, no. 3, p. 2021, [Online]. Available: www.ijacsa.thesai.org

[9] H. Imaduddin, B. Aditya Hermansyah, and F. B. Aura Salsabilla, "Comparison of Support Vector Machine and Decision Tree Methods in the Classification of Breast Cancer," *Jurnal Pendidikan Teknologi Informasi*, vol. 5, pp. 22–30, 2021.

[10] R. Kusumawati, A. D'Arofah, and P. A. Pramana, "Comparison Performance of Naive Bayes Classifier and Support Vector Machine Algorithm for Twitter's Classification of Tokopedia Services," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Nov. 2019. doi: 10.1088/1742-6596/1320/1/012016.

[11] E. A. Zanaty, "Support Vector Machines (SVMs) versus Multilayer Perception (MLP) in data classification," *Egyptian Informatics Journal*, vol. 13, no. 3, pp. 177–183, 2012, doi: 10.1016/j.eij.2012.08.002.

[12] D. L. Olson and D. Delen, *Advanced Data Mining Techniques*. Berlin Heidelberg: Springer-Verlag , 2008.

[13] N. Çolakoğlu and B. Akkaya, "Comparison of Multi-class Classification Algorithms on Early Diagnosis of Heart Diseases Recommendation System for Spotify View project," in *Proceeding of Conference: Recent Advances in Data Science and Business Analytics*, 2019, pp. 162–171. [Online]. Available: https://www.researchgate.net/publication/338950098

[14] D. N. Avianty, Prof. I. G. P. S. Wijaya, and F. Bimantoro, "The Comparison of SVM and ANN Classifier for COVID-19 Prediction," *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, vol. 13, no. 2, p. 128, Aug. 2022, doi: 10.24843/lkjiti.2022.v13.i02.p06.

[15] B. K. R, B. K, hreyas M. S, onu S, A. H, and A. B. J, "SVM Based Plant Diseases Detection using Image Processing," *International Journal of Computer Sciences and Engineering*, vol. 7, no. 5, pp. 1263–1266, May 2019, doi: 10.26438/ijcse/v7i5.12631266.

[16] M. Hussein and A. H. Abbas, "Plant Leaf Disease Detection Using Support Vector Machine," *Al-Mustansiriyah Journal of Science*, vol. 30, no. 1, pp. 105–110, Aug. 2019, doi: 10.23851/mjs.v30i1.487.

[17] N. Tripathy, "Stock Price Prediction Using Support Vector Machine Approach," in *International Academic Conference on Management & Economics*, 2019, pp. 44–59.

[18] 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.

[19] 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*, Institute of Electrical and Electronics Engineers Inc., Jan. 2019, pp. 37–41. doi: 10.1109/IEMCON.2018.8614994.

[20] D. Zhang, B. Bhandari, and D. Black, "Credit Card Fraud Detection Using Weighted Support Vector Machine," *Appl Math (Irvine)*, vol. 11, no. 12, pp. 1275–1291, 2020, doi: 10.4236/am.2020.1112087.

[21] Z. Rustam and A. A. Ruvita, "Application Support Vector Machine on Face Recognition for Gender Classification," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Dec. 2018. doi: 10.1088/1742-6596/1108/1/012067.

[22] S. Sulistiana and M. A. Muslim, "Support Vector Machine (SVM) Optimization Using Grid Search and Unigram to Improve E-Commerce Review Accuracy," *Journal of Soft Computing Exploration*, vol. 1, no. 1, pp. 8–15, 2020.

[23] W. M. Czarnecki, S. Podlewska, and A. J. Bojarski, "Robust optimization of SVM hyperparameters in the classification of bioactive compounds," *J Cheminform*, vol. 7, no. 1, Aug. 2015, doi: 10.1186/s13321-015-0088-0.

[24] W. Dubitzky, O. Wolkenhauer, K.-H. Cho, and H. Yokota, *Encyclopedia of Systems Biology*. London: Springer International Publishing, 2013. doi: 10.1007/978-1-4419-9863-7.

[25] A. K. Kordon, *Applying computational intelligence: How to create value*. Springer Berlin Heidelberg, 2010. doi: 10.1007/978-3-540-69913-2.

[26] S. Anam and Z. Fitriah, "Early Blight Disease Segmentation on Tomato Plant Using K-means Algorithm with Swarm Intelligence-based Algorithm," *Int J Math Comput Sci*, vol. 16, no. 4, pp. 1217–1228, 2021, [Online]. Available: http://ijmcs.future-in-tech.net

[27] 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: Jurnal Ilmu Matematika dan Terapan*, vol. 17, no. 2, pp. 0797–0806, Jun. 2023, doi: 10.30598/barekengvol17iss2pp0797-0806.

[28] X.-S. Yang, "A New Metaheuristic Bat-Inspired Algorithm," Apr. 2010. [Online]. Available: http://arxiv.org/abs/1004.4170