Implementation of cuckoo search algorithm for support vector machine parameters optimization in pre collision warning

A Puspaningrum*, A Suheryadi and A Sumarudin
Department of Informatics, Politeknik Negeri Indramayu, Jl. Lohbener Lama No. 08 Lohbener, Indramayu 45252, Indonesia.

*alifia.puspaningrum@polindra.ac.id

Abstract. Nowadays, pre-collision warning is one of the substantial aspects of the transportation sector. One of the steps to detect the collisions is by classifying and predicting the collisions. There are many supervised machine learning algorithms used, such as Least Square – Support Vector Machine (LS-SVM). Radial Based Function (RBF) is one of the LS-SVM kernels, which is a well-known method to support reliable performance. However, C and Gamma of its parameters are chosen randomly. This makes the performance of the classifier less optimal. To overcome that problem, this paper proposed a cuckoo search algorithm for optimizing two parameters to get optimal accuracy. The proposed approach is applied to 8437 transportation records and evaluated by using Accuracy. In addition, the performance of the proposed method is compared to other well-known meta-heuristic optimization algorithms, namely: Bat Algorithm (BA-SVM) and Firefly Algorithm (FA-SVM). Experimental results show that the Cuckoo Search Algorithm (CSA-SVM) yields the best performance for each of the 10-folds cross-validation by reaching 84.817% for accuracy, compared to the Bat Algorithm and Firefly Algorithm.

1. Introduction
Nowadays, pre-collision warning is one of the substantial aspects of the transportation sector. It provides many advantages to reduce and mitigate collisions by using various kinds of approaches, such as visual and audio recommendations [1,2]. Since the 1990s, many automotive companies have proposed many algorithms to assist the driver and ensure safety in driving [3].

One of the steps to detect the collisions is by classifying and predicting the collisions. There are several machine learning methods have been proposed to predict collisions, such as Least Square Support Vector Machine (LS-SVM), Random Forests (RF), and Artificial Neural Networks (ANN) as supervised machine learning algorithms [4-6]. Among those methods, LS-SVM provides faster computation compared to other machine learning methods. LS-SVM has several well-known kernels, namely: linear, polynomial, radial basis function (RBF), etc. Usually, linear and polynomial kernels consume less time and get less accuracy compared to RBF. In addition, RBF is one of the LS-SVM kernels which support reliable performance. There are two parameters of LS-SVM, which are chosen randomly, such as C and Gamma. So, the performance of the classifier becomes less optimal.

Wherefore, many researchers have conducted various nature-inspired metaheuristic methods to optimize the value of LS-SVM parameters. Finding the optimal solution is the aim of estimating the
value of LS-SVM parameters. Some of the metaheuristic methods include Particle Swarm Optimization, Genetic Algorithm, and Bat Algorithm [7-9]. On the other hand, in 2009, the Cuckoo Search Algorithm (CSA) was developed by Yang and Deb to overcome optimization problems. CSA was inspired by Cuckoo Bird's life behavior. Levy Flight of CSA combines local and global explorative random walk. It, therefore, helps the system not to be trapped in local optima while searching the best solution [10]. Moreover, CSA has been implemented in many sectors, for instance, water distribution networks, software effort estimation, wireless sensor networks, contrast enhancement of medical images [11-14].

Thus, to optimize the value of C and Gamma parameters, this paper proposed a cuckoo search algorithm to get the optimal performance. The performance is evaluated by Accuracy then. After implementing the method to 8437 transportation records and finding the optimal performance, the parameters can be implemented widely in the classification process in order to mitigate the collisions.

The structure of research paper is organized as follows: Section 1 contains the introduction, Section 2 contains a literature review, Section 3 explains the proposed method which has been done to improve the accuracy, result of the experiment is analyzed in section 4, and Section 5 presents the conclusion and future work of this research.

2. Related works
LS-SVM is one of robust intelligent technique which has faster computation time compared to other supervised techniques [15]. However, a random value of the C and Gamma parameter in LS-SVM can influence the classification performance. So, some researches have been conducted to optimize the parameters. Huang C and Wang C proposed parameter optimization by using the Genetic Algorithm in 2006 [8]. The experiment was evaluated in 11 real-world datasets of the UCI repository. The experiment shows that the best performance of accuracy achieved in fewer features of classification.

Therefore, it has widely implemented by using Particle Swarm Optimization (PSO) Algorithm in 2008. This paper has evaluated the same dataset as the research [8] used. But the result shows that the proposed method cannot yield better performance compared to Genetic Algorithm performance. These two types of research were also tested by Tharwat A in 2018 [9]. Tharwat A used the Bat Algorithm to optimize the parameters of LS-SVM in nine UCI Machine Learning dataset [9]. The results demonstrated that Bat Algorithm can achieve optimal performance and has a lower classification error compared to the Genetic Algorithm and PSO.

In 2009, CSA was developed by Yang and Deb as one of the modern metaheuristic methods to answer optimization problems [12]. This method also has been implemented widely in many kinds of research, for example, water distribution networks, software effort estimation, wireless sensor networks, contrast enhancement of medical images [11-14].

From the literature above, most of those researches attempt to achieve the best optimization performance. To the best of knowledge, the proposed method of this paper is the only research that attempts to optimize the parameter values of LS-SVM by using a real transportation dataset. The dataset implemented consists of six attributes, as Huang C and Wang C recommended to use fewer features in the previous research [8].

3. Method
In optimizing LS-SVM parameters, CSA is implemented to find the optimal performance. CSA-SVM was implemented in a PC with the following features: Intel(R) Core (TM) i7-8550U CPU @ 1.80GHz, 16G RAM, a Windows 10 Operating System, and MATLAB 2014a.

| Dataset Information       |       |
|---------------------------|-------|
| 0 (Not Danger)            | 4907  |
| 1 (Danger)                | 3530  |

8437
As shown in Table 1, the label of the classification consists of two labels, danger and no danger. The experimental data used in this research consists of six attributes, as seen in Table 2.

**Table 2. Parameters of CSA.**

| Input                  | Description                                      |
|------------------------|--------------------------------------------------|
| Transportation Dataset | (id, speed, distance, period, warning, pair, label) |

Output: Accuracy, Parameter C and Gamma

K-fold cross-validation has also been used in all experiments. The value of k was set to 10. The dataset then divided into training and testing data. The input and output of the dataset are described in Table 2.

![Figure 1. CSA-SVM flowchart.](image)

Actually, the proposed method generally consists of two main processes, namely optimization, and classification process, as shown in Figure 1. The description of implemented steps as described in this following steps:

- Preparing the dataset
- Splitting dataset into training and testing data
- The training data is used to construct the classification model. The value of C and Gamma parameters are taken from CSA optimization include CSA parameters and Generate Population.
- By using the classification model constructed, the testing data is classified using SVM then
- The accuracy value is being the fitness value
- If the stopping criterion is satisfied, the optimized parameters are achieved
- If it is not, CSA will generate new solutions and generate population then
- Step 3-5 will be repeated until the stopping criterion is achieved
Table 3. Parameters of CSA-SVM.

| Parameters     | Values                      |
|----------------|-----------------------------|
| Number of Cuckoos | 5                           |
| Number of Eggs    | 15                          |
| Iteration        | 2                           |
| Position         | Minimal = 50                |
|                 | Maximal = 125               |
| Pa             | 0.25                        |

Table 3 shows the parameters used in CSA-SVM. Then, the performance is evaluated by using Accuracy. It’s also used to calculate the fitness value of CSA-SVM. The equation of the accuracy is shown in Equation 1

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\] (1)

According to Equation 1, some cases with True Positive (TP) define the danger class, which correctly classified as dangerous. But, when the danger class is classified as not dangerous, it will be classified as False Negative (FN). In addition, when 'no danger' class is correctly classified as 'no danger,' the record will be classified as True Negative (TN). However, when 'no danger' is classified as 'danger,' it will be classified as False Positive (FP).

4. Experimental result

In this section, the proposed method has been implemented for LS-SVM parameters optimization. In addition, the Cuckoo Search Algorithm (CSA-SVM) is compared with other recent metaheuristic methods such as Bat Algorithm (BA-SVM) and Firefly Algorithm (FA-SVM). We considered 8437 records to be tested in five experiments. Each experiment tests the performance of each method.

According to Figure 2, the average value of CSA-SVM yields 83.847% as the best performance compared to BA – SVM and FA – SVM. The average performance is achieved based on five scenarios that have been done. The detail performance for each experiment is described in Table 4.

Based on experimental results show in Table 4, those three methods give accuracy that is not much different for each experiment. CSA-SVM becomes the best method except in the first experiment. FA-SVM gives accuracy that is not much different. On the contrary, BA-SVM gives the worst performance compared to other benchmarking methods. The aim of the proposed method is to know whether CSA-
SVM can optimize the value of each parameter. So, the value of CSA-SVM accuracy in Table 4 has been analyzed deeper, as shown in Table 5.

**Table 4.** Accuracy of experimental results.

| Experiment | BA – SVM | FA – SVM | CSA – SVM |
|------------|----------|----------|-----------|
| 1          | 77.657   | 83.217   | 82.659    |
| 2          | 78.726   | 83.205   | 83.655    |
| 3          | 76.034   | 83.229   | 83.561    |
| 4          | 81.368   | 83.146   | 84.544    |
| 5          | 79.447   | 83.146   | 84.817    |
| Average    | 78.646   | 83.189   | 83.847    |

**Table 5.** Parameters performance of CSA-SVM.

| Experiment | Gamma | C   | CSA – SVM |
|------------|-------|-----|-----------|
| 1          | 11.354| 68.248 | 82.659    |
| 2          | 55.492| 120.809 | 83.655    |
| 3          | 34.464| 97.491 | 83.561    |
| 4          | 38.889| 93.936 | 84.544    |
| 5          | 62.25 | 108.307 | 84.817    |

C and Gamma are two parameters optimized in this proposed method. When the optimal value of these two parameters is found, the classification performance of LS-SVM is going to be achieved, and the accuracy will be increased. However, not all performance of CSA-SVM experiments shows the best result. According to Table 5, the higher value of C can reduce the misclassified sample. On the other hand, the higher value of Gamma tends to increase the misclassified data, as shown in the value of accuracy of the performance. In LS-SVM, parameter C influence the margin width of the classification boundary for each class. Thus, the classifier will help to get correct classification data easier. However, the larger value of gamma can maximize the kernel value. So, incorrect data will be difficult to differentiate more. However, discussing these two parameters, under-fitting, is an issue for this topic.

**Table 6.** Time execution.

| BA – SVM | FA – SVM | CSA – SVM |
|----------|----------|-----------|
| 11.148   | 65.878   | 61.457    |

Besides comparing the accuracy value of each method, this paper also observed the time execution, which is described in Table 6. It shows that BA-SVM reached the fastest time execution compared to, and FA-SVM gets the slowest time execution. Indeed, the performance of time execution for each method is determined by how the method literally works. Moreover, Table 7 shows the best optimized CSA-SVM parameters in the traffic records dataset.

**Table 7.** Optimized parameters of CSA.

| Coefficient | Value  |
|-------------|--------|
| C           | 62.25  |
| Gamma       | 108.31 |

To sum up, the parameters optimized in this research are C and Gamma. The value of these parameters is balancing parameters to change the random number in calculating the accuracy. So, when the optimal parameters are found, the optimal performance of classification is going to be achieved. CSA-SVM shows the best performance for accuracy, but the time execution is in contrast. However, there is no guarantee that this method also gives the same performance when implemented in another sector.
There're also several factors which influence the performance, such as dimension, iteration, number of data, and the quality of the data itself.

5. Conclusion and further work
A Cuckoo Search Algorithm is proposed for optimizing LS-SVM coefficients. The experimental result has done by using 8437 transportation records to verify the proposed method’s reliability. Cuckoo Search Algorithm yields the best performance for each of the 10-folds compared to the other methods by reaching 84.817% for accuracy, compared to the Bat Algorithm and Firefly Algorithm. Regardless of the success of the proposed method, the exploitation of parameter quality still can be explored more. So that, hybrid optimization method can be one of option to get better quality for future research.

Acknowledgments
This research was fully supported by Pusat Penelitian dan Pengabdian Masyarakat, Politeknik Negeri Indramayu, as Penelitian Terapan Dana Lokal Polindra 2019.

References
[1] Reagan I J, Cicchino J B, Kerfoot L B and Weast R A 2018 Crash avoidance and driver assistance technologies – Are they used? Transp. Res. Part F Psychol. Behav. 52 176–90
[2] Eichelberger A H and Mccart T 2015 Toyota drivers’ experiences with Dynamic Radar Cruise Control, Pre-Collision System, and Lane-Keeping Assist
[3] Wang J, Wu J, Zheng X, Ni D and Li K 2016 Driving safety field theory modeling and its application in pre-collision warning system Transp. Res. Part C 72 306–24
[4] Dogru N and Subasi A 2018 Traffic Accident Detection Using Random Forest Classifier 2018 15Th Learn. Technol. Conf. 40–5
[5] Gomes S L, Rebouças E de S, Neto E C, Papa J P, Albuquerque V H C d., Rebouças Filho P P and Tavares J M R S 2017 Embedded real-time speed limit sign recognition using image processing and machine learning techniques Neural Comput. Appl. 28 573–84
[6] Park H S, Kim D J, Kang C M, Kee S C and Chung C C 2018 Object detection in adaptive cruise control using multi-class support vector machine IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC 2018-March 1–6
[7] Lin S, Ying K, Chen S and Lee Z 2008 Particle swarm optimization for parameter determination and feature selection of support vector machines 35 1817–24
[8] Huang C and Wang C 2006 A GA-based feature selection and parameters optimization for support vector machines 31 231–40
[9] Tharwat A, Hassanien A E and Elnaghi B El 2016 A BA-based algorithm for parameter optimization of Support Vector Machine A BA-based algorithm for parameter optimization of support vector machine
[10] Yang X, Cb C and Deb S 2009 Cuckoo Search via L ‘ evy Flights
[11] Naidu M N, Boindana P S, Vasan A and Varma M R R Optimization of Water Distribution Networks Using Cuckoo Search Algorithm (Springer Singapore)
[12] Parwita I M M, Sarno R and Puspaningrum A 2017 Optimization of COCOMO II Coefficients using Cuckoo Optimization Algorithm to Improve The Accuracy of Effort Estimation pp 99–104
[13] Razzaque M A and Hussein M 2015 A Novel Cuckoo Search Based Clustering Algorithm for Wireless Sensor Networks
[14] Daniel E and Anitha J 2016 Optimum wavelet based masking for the contrast enhancement of medical images using enhanced cuckoo search algorithm Comput. Biol. Med. 1–7
[15] Siuly Siuly and Yan Li 2012 Improving the Separability of Motor Imagery EEG Signals Using a Cross Correlation-Based Least Square Support Vector Machine for Brain–Computer Interface IEEE Trans. Neural Syst. Rehabil. Eng. 20 526–38