Iraqi Journal of Intelligent Computing Informatics (IJICI)
Vol. 1, 1. 1, June 2022, pp. 11~22
ISSN: 2791-2868, DOI: 10.52940/ijici.v1i1.3

Bat algorithm optimisation technique for feature selection on different dimension of datasets

Hussein Fouad Almazini[1], Farzana Kabir Ahmad[2]
[1]Department of Computer Science, Shatt Al-Arab University College, Basra, Iraq
[2] School of Computing, Universiti Utara Malaysia 6010 Sintok, Kedah, Malaysia

Article Info

Article history:

Received February 10, 2022
Revised March 10, 2022
Accepted June 1, 2022

Keywords:

Metaheuristic
Feature Selection
Particle Swarm Optimisation
Genetic Algorithm
Classification

ABSTRACT (10 PT)

The advanced of Information Technology has resulting in the generation of numerous datasets with different dimensions. However, dealing with multi-dimensional datasets which typically contain large number of attributes, p has cause problems to classification process. Classifying different dimensional numerical data is a difficult problem as dealing with various feature spaces, could cause the performance of supervised learning method to suffer from the curse of dimensionality. This condition eventually degrades both classification accuracy and efficiency. In a nutshell, not all attributes in the dataset can be used in the classification process since some features may lead to low performance of classifier. Feature selection (FS) is a good mechanism that minimises the dimensions of high-dimensional datasets and solve classification problems. This paper proposed Bat Algorithm (BA) for FS that were trained using a Support Vector Machine (SVM) classifier. The proposed algorithm was tested on six public datasets with different sizes and compared with other benchmark algorithms, such as Particle Swarm Optimisation (PSO) and Genetic Algorithm (GA). The experimental results indicated that the BA has outperformed the other two algorithms. In addition, the comparison details showed that binary BA is more competitive in terms of accuracy and the number of features when assessed on datasets with different sizes.

Corresponding Author:

Hussein Fouad Almazini
Department of Computer Science, Shatt Al-Arab University College, Basra, Iraq
Email: hussein.f.abbas@sa-uc.edu.iq

1. INTRODUCTION

Machine learning methods have been used by researchers for various data analytics tasks in the past several years. Machine learning methods are used to get insight of data by analyzing and examining sample data or experience. Regardless the potential of machine learning in extracting information from datasets, the usage of a machine learning method alone guarantees no ideal solutions. For example, a neural network learning algorithm necessitates the use of an optimization strategy in order to get the best results. Therefore, optimisation is a significant component to produce an ideal machine learning outcomes.

Machine learning methods face a big problem when it comes to choosing relevant attributes from data and getting rid of irrelevant ones [1]. Normally, massive data contains a high number of irrelevant features. The inclusion of such non-relevant features might make analysis time-consuming as well as computationally costly [2], [3]. Irrelevant features from raw data can be eliminated by FS technique [4]. The FS algorithm is used in the training dataset before to the analysis phase of machine learning to pick a subset from the primary set of characteristics that will be utilized to build patterns in the analysis dataset. Since the 1970s, FS has been regarded as an important topic in research and development.

FS has been effectively implemented in a variety of issue situations during the last several years, including data mining applications, genomic analysis, text categorization, pattern classification and information retrieval processing [5], [6]. The advantages of FS in data mining classification are as follows: (1) increases comprehensibility, (2) reduces the complexity of the induced model, (3) develops inductive learning,

(4) improves predictive accuracy and (5) it speeds up algorithm for data mining [7], [8]. Wrapper, filter, and embedding techniques are the three primary ways to look at FS in general [4]. Compared to the wrapper approach, the filter approach is quicker but less precise and complicated [9], [10]. Due to its sufficient results and efficiency in dealing with bigger and more complicated datasets than that using the filter approach, the wrapper approach is commonly utilized [11].

Recently, metaheuristic methods have been used by researchers to determine near-optimal solutions because these methods can search in the full search space. Many studies have been presented to use metaheuristic algorithms to tackle FS problems [12]–[17]. For example, Zhang et al., 2015 extended a bare-bone particle swarm optimisation (BPSO) algorithm for FS with binary variables, which is called the binary BPSO [18]. In addition, a new FS algorithm based on a new variant of ant colony optimisation (ACO), that is enriched ACO with a new function that could perform better in searching mechanism [19].

Bioinspired optimization approaches are utilized to address complicated optimization issues because of their simplicity, efficiency, and robustness [20]–[25]. Bat algorithm (BA) can be a possible solution to solve problems in extracting data, such as classification and attribute selection [26]–[31]. Bats can distinguish between several insects and determine preys even in the dark. Thus, proposed a new metaheuristic algorithm and termed it BA [32]. Motivated by this algorithm, in this study we proposed the binary BA (BBA) for FS. This algorithm mainly aims to establish a set of binary coordinates for each bat and examine whether the coordinate will belong to the last subset of the attributes. This task can be maximised by using supervised classifier fineness.

Given that the solution fineness is linked with each bat, we classify an evaluating group and evaluate each bat by training a classifier with the attributes, which are selected and encoded in accordance with the bat position. Hence, we need a fast and powerful classifier. As such, we opted to use the Support Vector Machine (SVM) classifier [33]–[35] which is as effective as the optimum-path forest (OPF) classifier yet faster for training. In this study, performance evaluation was performed on six general datasets with different sizes: (1) small, (2) medium and (3) large datasets. Experiment on the proposed algorithm was evaluated by comparing it with other relevant approaches, such as PSO and GA.

2.Bat Algorithm for Feature Selection

Feature selection generally aims to discover the most discriminative information to understand and get insight into several application domains. Finding features/attributes that are easy to discern, invariant to geometric and affine transformations, noise-insensitive, and useful for describing patterns in several categories is often suitable. The choice of such features is a crucial step and mainly depends on the nature of the problem. Accordingly, many researchers has used BA in the domain of FS to maximize the supervised classifier's accuracy and it relative performances [30]. Section below explained in detail the BA and BBA mechanisms.

Bat Algorithm

The last few years have seen a lot of researchers use BA because microbats can find their prey and tell the difference between different types of insects even in complete darkness. The BA method has been shown to outperform certain well-known nature-inspired methods. In terms of echolocation and recognizing prey and impediments, bats are very skilled and sophisticated. This exceptional efficiency has piqued the interest of researchers interested in its use in a variety of fields. Bats send short- and high-pulse sounds and wait for them to hit a certain target. Echoes will return to their ears after a short period of time [36]. In this manner, bats are capable of calculating the distance between themselves and an object is [37]. Bats also have a unique treadmill system that allows them to discern between prey and obstacles when pursuing in full darkness [36], [37]. On the basis of the bats' behaviour and their ability to track preys in darkness, [32] developed a notable idea called the BA. The technique has been improved further by other researchers who developed a new version by enhancing its capacity to locate food, prey, and obstacles via echolocation. The BA generally deals with the following rules [32]:

1. Bats sense their environment by echolocation and are capable of discriminating between danger and food;
2. Bats (bi) flies randomly with velocity (vi) at position (xi) with a fixed frequency (fmin), with varying wavelength '(λ)' and loudness '(A0)' to search for preys. Bats (bi) can automatically set

the wavelength (or frequency) of their emitted pulses and adjust the rate of pulse emission r ∈ [0, 1] depending on the proximity of their target.; and

3. Although loudness might vary greatly, assumed that it ranges from a large (positive), $A0$ to a minimum constant value $Amin$.

Algorithm 1 shows below is the BA, which was adopted from [32]:

```
1.  While t < T
2.      For each bat b_i, do
3.          Generate new solutions through Equations (1),
4.          (2) and (3).
5.          If rand > r_i, then
6.              Select a solution among the best solutions.
7.              Generate a local solution around the
8.              best solution.
9.          If rand < A_i and f(x_i) < f(x̂), then
10.             Accept the new solutions.
11.             Increase r_i and reduce A_i.
12. Rank the bats and find the current best x̂.
```

Figure 1. Algorithm 1 (adopted from [32])

At first, the initial position of xi, frequency "fi, velocity "vi" are initiated for every bat bi"."For every time phase "t, being T is the maximum number of iterations, , the movement of the virtual bats is set by updating the position and velocity using Equations 1 to 3 as follows:

$$fi = f\min + (f\min - f\max)\ \beta, \quad (1)$$

$$\text{“}vj\ i\ (t) = vj\ i\ (t-1) + [\hat{}xj - xj\ i\ (t-1)]\ fi,\text{”} \quad (2)$$

$$\text{“}xj\ i\ (t) = xj\ i\ (t-1) + vj\ i\ (t)\text{”} \quad (3)$$

where "β" represents a randomly created number for during interval period of [0,1]. "$xji\ (t)$" on the other hand, is a value of decision variable "j" for bat "i" at time phase "t". The result of "fi" in Equation 1 is used to control the range of the motion of the bats and its pace. A variable "xj" performs the existing the global best solution (position) for decision variable "j" which is done by comparing all the solutions offered by m" bats.

As a way to increase the number of possible solutions, Random walks were suggested in [32] by Yang. Typically, one of the best solutions is chosen from the rest. Subsequently, the casual walk is then utilized to generate a new solution for each bat that takes the case in algorithm 1 Line 5:

$$\text{“}xnew = xold + \epsilon A(t)\text{”} \quad (4)$$

At every "$A(t)$, where A(t) is the average loudness of all the bat at time t, and $\epsilon \in [-1, 1]$ is the strength of the random walk and attempts the direction, the emission pulse rate "ri" and the loudness "Ai are updated", as follows:

$$\text{“}Ai(t+1) = \alpha Ai(t)\text{”} \quad (5)$$

and

$$\text{“}ri(t+1) = ri(0)[1 - exp(-\gamma t)\text{”} \quad (6)$$

where "γ" and "α" are ad-hoc constants. At first step of this method, the loudness "$Ai(0)$" and the emission rate "$ri(0)$" are often randomly selected for any event, "$Ai(0)$" ∈ [1, 2] and "$ri(0)$" ∈[0, 1] [12].

BBA: Binary Bat Algorithm

Each bat movement in the search space is across continuous valued locations. However, there is an issue of attribute selection as the search space is modelled as a n- dimensional Boolean trellis. Since the problem is mainly due to the selecting of a given attribute, bat's location is performed by binary vectors. This study adopts a binary part of BA by limiting the new bat's location to binary values using a sigmoid service:

$$S(vj\ i) = \frac{1}{11+e-Vji} \quad (7)$$

Thus, in Equation 3 can be replaced to:

$$xj\ i = \begin{cases} -x, x < 0 \\ 1 \text{ if } S(Vji) > \sigma \end{cases} \quad (8)$$

in which every" $\sigma \sim$ (0, 1)". Thus, Equation 8 can supply the binary values only for every bat coordinates in the Boolean trellis, which stand for the presence of absence of the features.

3.Proposed Method for Feature Selection

Figure 2 depicts the six-step experimental framework that has been presented. In the first step, different datasets sizes are selected that are used to evaluate the suggested algorithm's performance. Table 1 specifies the number of characteristics and the class. The use of test data, which are subsets of the original dataset of varied sizes that are trained using an SVM classifier to achieve accuracy prior to the feature selection process, is the second step in this recommended methodology. The bat algorithm is then used to choose features in the third step. In the fourth step, the experiment is set up to pick the best subsets of features using BBA. The best subsets are determined through the last step, in which the experimental results are evaluated in term of the accuracy and number of features by comparing the results with other previous methods.

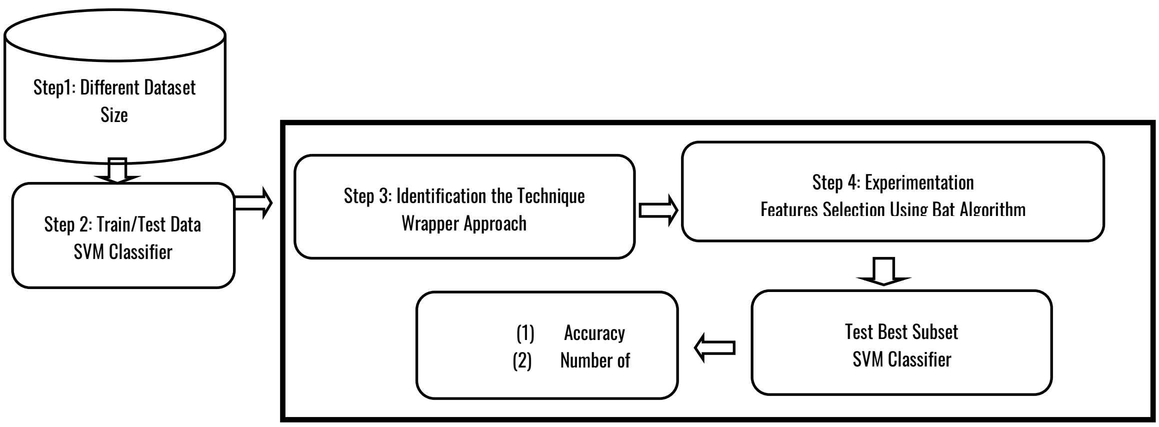

Figure 2. Experimental Framework

This research contributes to the literature as follows: firstly, BBA is used with classification technique; secondly, the heuristic method of BBA is utilized to increase the efficacy of SVM. The location of each bat in the search distance encodes and executes a set of qualities. As a result, for each subset, an SVM classifier is trained in one portion and assessed in another section that was not observed during the training to determine each bat's fitness score. Training and assessing subsets may be repeated across bats since each bat has a large number of distinct attribute subsets. Firstly, the bat population is established. Each location is subsequently prefaced by random picking of a binary amount, which matches if the attribute will select or not to create the incoming dataset. Secondly, a new training and testing subset is created with the selected attributes. Thirdly, each attribute is evaluated by updating its fitness rate. The rate of pulse emission (ri) and the loudness (Ai) are updated if the new solution is acceptable. By contrast, the loudness generally reduces one time, bat finds its prey and the average pulse emission increases (Equations 5–6). Fourthly, the maximum task yields the bat fitness value that maximises the fitness function and the index. When the global best position is updated, i.e., $\hat{}x$, the bat position reaches the highest fitness function. Fifthly, the maximum task results in the bat fitness function that maximises the index. Afterwards, the global best location'x' is updated with the bat location, which reaches the highest fitness value. Seventhly, bat location and velocity are updated. In the same stage, a qualified local search is implemented. Finally, each frequency is updated according to Equation (1). The '$f$max' values and '$f$min are 1 and '0, respectively. Finally, in the last step, the procedure obtains and returns the output vector (F) with the selected attributes.

4.Expenrimental Results and Discussion

This section highlights the assessment of the BBA performance for feature selection. Six public datasets were used in this work. These datasets have different sizes and were obtained from the UCI Machine Learning Repository [38]. The datasets include, WBCD, Australian, German, WPBC, lung cancer and sonar datasets. We classified the datasets according to sizes: large scale (over 50 features) features, medium (20-49 features) and small (1-19 features) [39]–[42]. Table 1 also shows the main characteristics of each dataset.

Table 1.Description of the datasets from UCI repositories

Dataset	Size	#Sample	Features	Classes
WBCD	small	683	9	2
Australian	small	690	14	2
German	medium	1000	24	2
WPBC	medium	198	32	2
Lung Canser	large	32	56	3
Sonar	large	208	60	2

In this study, two experiments were performed. At first, the performance of SVM classifier is tested on different datasets prior to any feature selection algorithm. The second experiment is the comparative analysis of BBA against the PSO and the Genetic Algorithm (GA). To compute the classification of the mean and feature number, the experiment was performed through 10-folds cross validations to obtain robust results. The most ideal subset of features was used to maximise the area of curve (AUC) of featured values, in which could maximising accuracy of results. Table 2 presents the results of SVM accuracy when tested on different datasets without any feature selection algorithm.

Table 2.Classification accuracy without feature selection on the original datasets

Data Name	SVM Accuracy
WBCD	***66.6%***
Australian	***60.9%***
German	***75.4%***
WPBC	***77.2%***
Lung Canser	***43.8%***
Sonar	***80.3%***

Error! Reference source not found. to Figure 8 illustrate the experimental results for BA, PSO and GA in 10 runs, with 70% of training datasets and 30% of testing datasets. The experimental results show that SVM achieved low classification accuracy when all the features are used (see Table 2). For instance, in WBCD and Australian datasets that denotes to small size of features, the classification accuracies are observed at a rate of 66.6% and 60.9% respectively when no algorithm was applied to reduce the attributes. This phenomenon is generally due the non-informative features that degrade the classification accuracy. Figure 3 and Figure 4 show the performance of BBA, PSO and GA on WBCD and Australian datasets, which were categorized under small datasets. BBA selected four features from the nine features for WBCD dataset and five features from the fourteen features for Australian dataset. The WBCD dataset achieved a high accuracy of 96.90%, as shown in Fig. 3. Meanwhile, PSO and GA selected only two features with an improved classification accuracy of 96.30% and 96.40% for PSO and GA, respectively. The identical performance also been seen in Australian dataset with BBA outperform compare to PSO and GA.

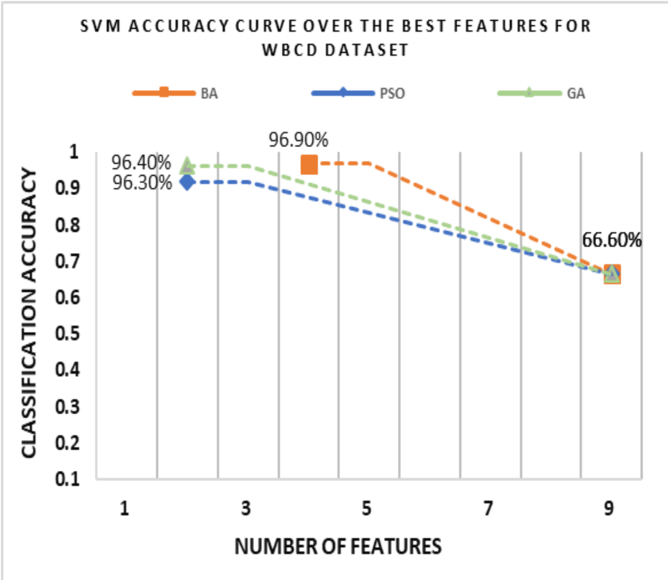

Figure 3. WBCD dataset

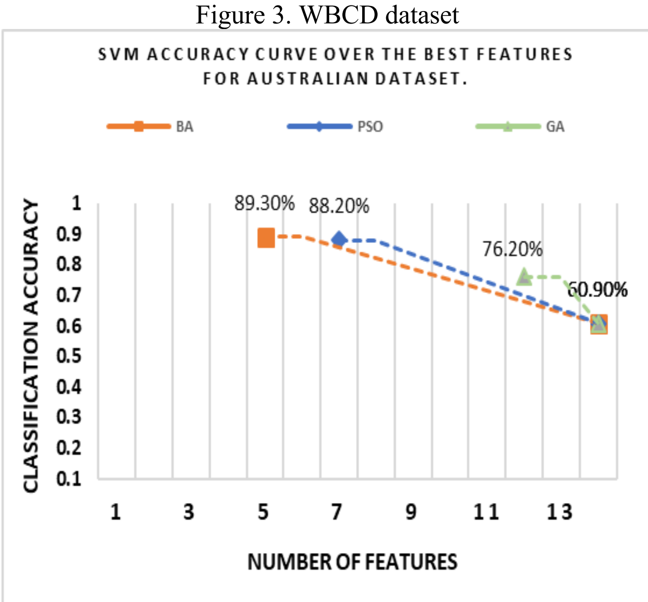

Figure 4. Australian dataset

The classification accuracy results for BBA, PSO and GA algorithms using SVM classifier on the datasets with medium size were 75.4% and 77.2% for German and WPBC, respectively, with all the features, as represented in Table 3 and Table 4. The performance of SVM classifier improved and displayed a high accuracy for the datasets that have medium size when BBA selected six features from the 24 attributes in the German database and seven features from the 32 attributes in the WPBC database. The classification accuracy (Fig. 5) obtained in the German database was 86.40%, and PSO achieved approximately 76.00% with three features and 76.70% for GA with four features. In the WPBC database, the accuracy rate (Fig. 6) increased to

84.10% and 78.90% for BBA and PSO, respectively, with three features. GA obtained 77.90% with four features.

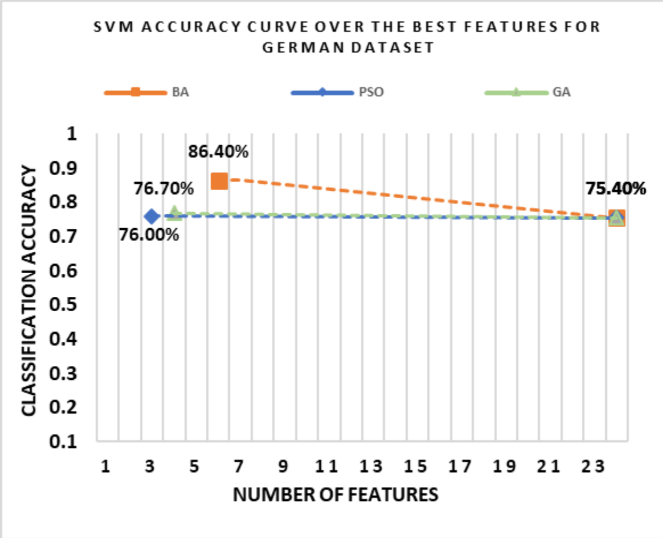

Figure 5.German dataset

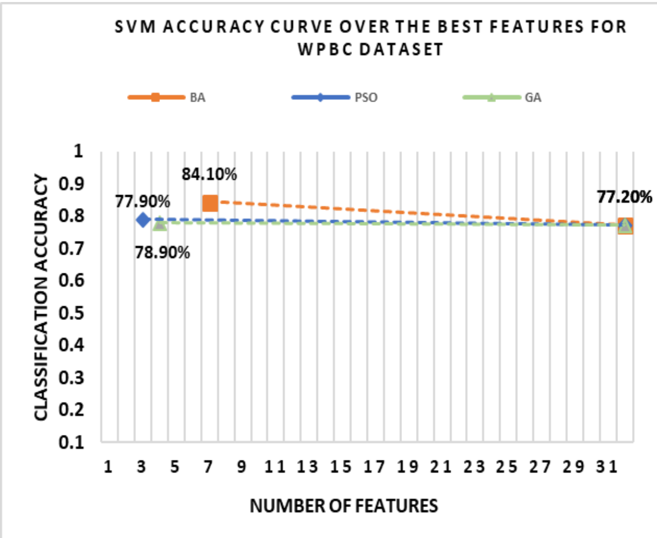

Figure 6. WPBC dataset

In larger datasets, it discovered that BBA selects less number of features compare to PSO and GA while producing much higher accuracy, which is quite opposite from the other two dataset sizes. Figure 7 and Figure 8 presented the SVM accuracy when run on lung and sonar datasets. BBA selected 12 features from the 56 features of lung cancer database and 9 features from the 60 features of sonar database. Additionally, PSO and GA selected 23 and 27 features with improved classification accuracy of 88.30% and 89.40%, respectively. These findings show that BBA works better when tested on larger number of features.

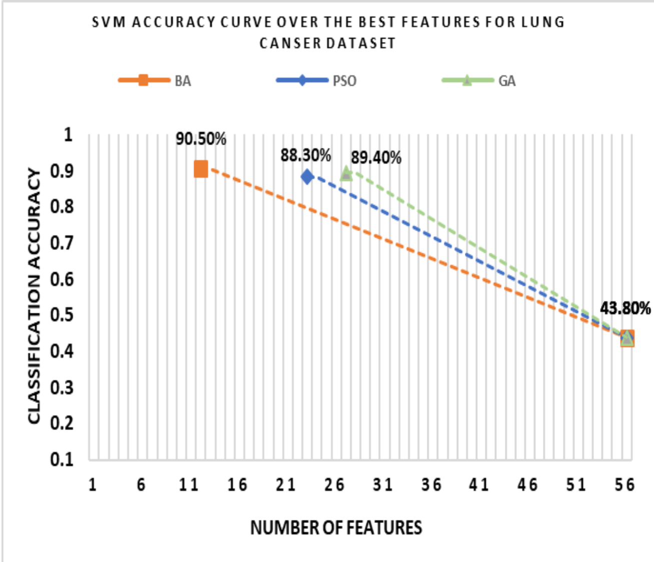

Figure 7.Lung dataset

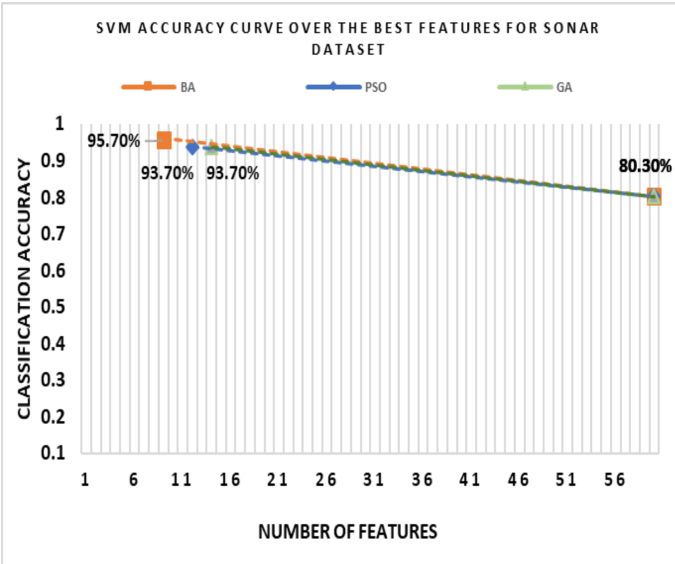

Figure 8.Sonar dataset

Table 3 displays the SVM accuracy over the most ideal subset, which maximises the presented curves. BA outperformed other algorithms for all the datasets with different sizes.

Table 3.Classification accuracy for the best subset of attributes

Dataset	BA	PSO	GA
WBCD	96.9%	96.3%	96.4%
Australian	89.3%	88.2%	76.2%
German	86.4%	76.0%	76.7%
WPBC	84.1%	78.9%	77.9%
Lung Canser	90.5%	88.3%	89.4%
Sonar	95.7%	93.7%	93.7%

Finally, Table 4 shows the number of selected attributes for each optimisation method. In particular, BBA derived lesser features than those of other methods, with the exception of WBCD, German and WPBC datasets.

Table 4 Number of the features selected for each optimization technique

Dataset	BA	PSO	GA
WBCD	4	2	2
Australian	5	7	12
German	6	3	4
WPBC	7	3	4
Lung Cancer	12	23	27
Sonar	9	12	14

According to the overall measurement results, our proposed method outperforms the other methods in term of the classification accuracy as well as it's the ability to select the most ideal subset of attributes on datasets with different sizes. Consequently, improved results can be achieved overall. The proposed algorithm also selected a larger number of attributes than those of other algorithms in datasets with small data size (WBCD) and medium data size (German and WPBC), and produced classification rate higher than those of PSO and GA. From existing literatures the experimental result revealed that the proposed algorithm it's used in reducing dimension features and it can efficiently improve classification accuracy. Clearly, the number of dataset features has an effect on each individual's fitness. According to popular belief, if the number of features is too limited, the number of chosen features will have a greater impact on fitness than classification accuracy. In fact, feature selection aims to reduce the dimension of feature and achieve high percentage of accuracy. Recall that feature selection prefers a highly percentage of accuracy when datasets have a relatively small number of features. Nevertheless, it seems that in high-dimension datasets with huge sizes, which include selection, the better subset with relatively high classification accuracy is preferred. Furthermore, in case that the number of dataset features exceeds 20 (medium and large size), the algorithm produced better classification accuracy and less features.

5.Conclusions and Future works

We proposed the usage of BBA together with SVM for feature selection. The general concept of this algorithm is the establishment of a set of binary coordinates for each bat and assessing whether this coordinate will belong to the last subset of the attributes. Furthermore, we evaluate the efficiency of BBA in conjunction with an SVM classifier in order to extract the most relevant subset of features from a variety of dataset sizes. Moreover, we used the wrapper technique for feature selection during the learning process. The following methods typical procedures for selecting wrapper features were used: Forward selection, backward elimination, and stochastic search. Nevertheless, datasets with different sizes were examined to evaluate the effectiveness of the proposed algorithm. However, The classifier's performance may be improved by picking the most suitable features from a given feature set for training and assessment. Result was determined by comparing with those of other methods that used the same datasets for FS (e.g. PSO and GA). BBA was shown to be competitive on datasets of various sizes in terms of accuracy and number of features. Finally, as future research directions, we should focus on the hybridisation of BBA for FS with local search in stochastic methods to determine many improvements in each candidate solution. Parameter control mechanisms can also be used to adjust the values for the evaluated BBA parameters. These parameter mechanisms can be either pre-deterministic rules or based on the feedback from the search itself (e.g. solution quality).

Acknowledgement

The authors would like to acknowledge the Ministry of Education (FRGS Grant) and Universiti Utara Malaysia (UUM) for the financial support of this research project.

REFERENCES

[1] K. I. Rufai, R. C. Muniyandi, and Z. A. Othman, “Improving bee algorithm based feature selection in intrusion detection system using membrane computing,” *J. Networks*, vol. 9, no. 3, pp. 523–529, 2014.

[2] I. Cheng, “Hybrid Methods for Feature Selection,” Western Kentucky University, 2013.

[3] Qinbao Song, Jingjie Ni, and Guangtao Wang, “A Fast Clustering-Based Feature Subset Selection Algorithm for High-Dimensional Data,” *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 1, pp. 1–14, 2013.

[4] M. S. Packianather and B. Kapoor, “A wrapper-based feature selection approach using Bees Algorithm for a wood defect classification system,” *10th Syst. Syst. Eng. Conf. SoSE 2015*, pp. 498–503, 2015.

[5] M. Mahmuddin and Y. Yusof, “A Hybrid Simplex Search and Bio-Inspired Algorithm for Faster Convergence,” *2009 Int. Conf. Mach. Learn. Comput.*, vol. 3, no. February 2015, pp. 203–207, 2009.

[6] I. Mandli and P. Mahesh, “Selection of Most Relevant Features from High Dimensional Data using IG-GA Hybrid Approach,” *Int. J. Comput. Sci. Mob. Comput.*, vol. 3, no. 2, pp. 827–830, 2014.

[7] Z. Matta and W. Lee, “The digital universe of opportunities: rich data and the increasing value of the internet of things,” vol. 2, no. 3, pp. 1–9, 2014.

[8] S. Alelyani, *On Feature Selection Stability: A Data Perspective*, no. May. Arizona State University, 2013.

[9] D. T. Pham, M. Mahmuddin, and S. Otri, “Application of the Bees Algorithm to the Selection Features for Manufacturing Data,” *Manuf. Eng.*, no. January, 2007.

[10] P. Gupta, S. Jain, and A. Jain, “A Review of Fast Clustering-Based Feature Subset Selection Algorithm,” *Int. J. Sci. Technol. Res.*, vol. 3, no. 11, pp. 86–91, 2014.

[11] Y. Sun, S. Todorovic, and S. Goodison, “Local-learning-based feature selection for high-dimensional data analysis,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1–18, 2010.

[12] H. Almazini and K. R. Ku-Mahamud, “Grey Wolf Optimization Parameter Control for Feature Selection in Anomaly Detection,” *Int. J. Intell. Eng. Syst.*, vol. 14, no. 2, pp. 474–483, 2021.

[13] V. Agrawal and S. Chandra, “Feature selection using Artificial Bee Colony algorithm for medical image classification,” in *Proc. of International Conf.on Contemporary Computing*, 2015, pp. 171–176.

[14] L. Y. Chuang, C. H. Yang, and J. C. Li, “Chaotic maps based on binary particle swarm optimization for feature selection,” *Appl. Soft Comput.*, vol. 11, no. 1, pp. 239–248, 2011.

[15] H. M. Zawbaa, E. Emary, A. E. Hassanien, and B. Parv, “A wrapper approach for feature selection based on swarm optimization algorithm inspired from the behavior of social-spiders,” in *Proc. of International Conf.on Soft Computing and Pattern Recognition*, 2015, pp. 25–30.

[16] M. Sudana, R. Nalluri, Saisujana, H. Reddy, and V. Swaminathan, “An Efficient Feature Selection using Artificial Fish Swarm Optimization and SVM Classifier,” in *Proc. of International Conf. on networks & advances in computational technologies*, 2017, no. 7, pp. 407–411.

[17] J. André and G. Sargo, “Binary Fish School Search applied to Feature Selection,” Tecnico Lisboa, 2013.

[18] Y. Zhang, D. Gong, Y. Hu, and W. Zhang, “Feature selection algorithm based on bare bones particle swarm optimization,” *Neurocomputing*, vol. 148, pp. 150–157, 2015.

[19] R. Forsati, A. Moayedikia, R. Jensen, M. Shamsfard, and M. R. Meybodi, “Enriched ant colony optimization and its application in feature selection,” *Neurocomputing*, vol. 142, pp. 354–371, 2014.

[20] A. M. Jabbar, K. R. Ku-Mahamud, and R. Sagban, “An improved ACS algorithm for data clustering,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 17, no. 3, pp. 1506–1515, 2020.

[21] S. A. Yasear and K. R. Ku-Mahamud, “Taxonomy of Memory Usage in Swarm Intelligence-Based Metaheuristics,” *Baghdad Sci. J.*, vol. 16, no. 2(SI), p. 0445, 2019.

[22] A. M. Jabbar, K. R. Ku-Mahamud, and R. Sagban, “Modified ACS centroid memory for data clustering,” *J. Comput. Sci.*, vol. 10, no. 10, pp. 1439–1449, 2019.

[23] H. F. Almazini, S. Mortada, H. F. A. Al-Mazini, H. N. K. Al-Behadili, and J. Alkenani, “Improved discrete plant propagation algorithm for solving the traveling salesman problem,” *IAES Int. J. Artif. Intell.*, vol. 11, no. 1, pp. 13–22, 2022.

[24] R. Tiwari, “Correlation-based Attribute Selection using Genetic Algorithm,” *Int. J. Comput. Appl.*, vol. 4, no. 8, pp. 28–34, 2010.

[25] V. Kumar and S. Minz, “Feature Selection : A literature Review,” *SmartCR*, vol. 4, no. 3, 2014.

[26] H. N. K. Al-behadili, K. R. Ku-Mahamud, and R. Sagban, “HYBRID ANT COLONY OPTIMIZATION AND ITERATED LOCAL SEARCH FOR RULES-BASED CLASSIFICATION,” *J. Theor. Appl. Inf. Technol.*, vol. 98, no. 04, pp. 657–671, 2020.

[27] K. Hussain, M. N. Mohd Salleh, S. Cheng, and Y. Shi, “Metaheuristic research: a comprehensive survey,” *Artif. Intell. Rev.*, vol. 52, no. 4, pp. 2191–2233, 2019.

[28] H. N. K. Al-behadili, K. R. Ku-Mahamud, and R. Sagban, “Hybrid Ant Colony Optimization and Genetic Algorithm for Rule Induction,” *J. Comput. Sci.*, vol. 16, no. 7, pp. 1019–1028, 2020.

[29] R. Mohamed *et al.*, “The Effectiveness of Bat Algorithm for Data Handling in Various Applications,” no. November, pp. 25–27, 2016.

[30] B. Yang, Y. Lu, K. Zhu, G. Yang, and J. Liu, “Feature Selection Based on Modified Bat Algorithm,” no. 8, pp. 1860–1869, 2017.

[31] R. Y. M. Nakamura, L. A. M. Pereira, K. A. Costa, D. Rodrigues, J. P. Papa, and X. S. Yang, “BBA: A binary bat algorithm for feature selection,” *Brazilian Symp. Comput. Graph. Image Process.*, pp. 291–297, 2012.

[32] X. S. Yang, “A new metaheuristic Bat-inspired Algorithm,” *Stud. Comput. Intell.*, vol. 284, pp. 65–74, 2010.

[33] R. G. Brereton and G. R. Lloyd, “Support Vector Machines for classification and regression,” *Analyst*, vol. 135, no. 2, pp. 230–267, 2010.

[34] A. Ng, “Margins : Intuition Support Vector Machines,” *Intell. Syst. their Appl. IEEE*, vol. pt.1, no. x, pp. 1–25, 2000.

[35] F. Markowetz, “Classification by Support Vector Machines,” *Discret. Methods Epidemiol.*, vol. 0000, no. 2, pp. 1–9, 2000.

[36] D. R. Griffin, F. A. Webster, and C. R. Michael, “The echolocation of flying insects by bats,” *Anim. Behav.*, vol. 8, no. 3–4, pp. 141–154, 1960.

[37] W. Metzner, “Echolocation behaviour in bats,” *Sci. Prog.*, pp. 453–465, 1991.

[38] K. Bache and M. Lichman, “UCI Machine Learning Repository,” *Univ. Calif. Irvine Sch. Inf.*, vol. 2008, no. 14/8, p. 0, 2013.

[39] M. A. Hall and G. Holmes, “Benchmarking Attribute Selection Techniques for Discrete Class Data Mining,” *IEEE Trans. Knowl. Data Eng.*, vol. 15, no. 6, pp. 1437–1447, 2003.

[40] S. J. Raudys and A. K. Jain, “Small sample size effects in statistical pattern recognition: recommendations for practitioners,” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1991, vol. 13, no. 3, pp. 252–264.

[41] A. S. Ogale and C. Fermu, “Short Papers On the Selection and Classification of Independent Features,” *Analysis*, vol. 27, no. 6, pp. 988–992, 2005.

[42] M. Kudo and J. Sklansky, “Comparison of algorithms that select features for pattern classifiers,” *Pattern Recognit.*, vol. 33, no. 1, pp. 25–41, 2000.