

Available online at [www.sciencerepository.org](http://www.sciencerepository.org)

Science Repository



## Research Article

# Classification of Diseases Using a Hybrid Binary Bat Algorithm with Mutual Information Technique

Firas Ahmed Yonis AL\_Taie\* and Omar Saber Qasim

Department of Mathematics, University of Mosul, Mosul, Iraq

### ARTICLE INFO

#### Article history:

Received: 25 November, 2019

Accepted: 18 December, 2019

Published: 28 December, 2019

#### Keywords:

Binary bat algorithm (BBA)

classification

mutual information

bigdata

### ABSTRACT

Many features change the results of calculations and change this may be a negative impact on the accuracy of the results, especially if the data used is large. Evolutionary algorithms are used to find the fastest and best way to perform these calculations, such as the bat algorithm (BA) by reducing the dimensions of the search area after changing it from continuous to discrete. In this paper, we will propose a method of linking and conclude the selection of the best and most influential features on the results by neglecting the negative impact features through three stages: the first will be the arrangement of the features columns according to their importance Starting of the most important using mutual information technology and the second stage the process of cutting these features into A certain limit and content with the most important and the calculations using the workbook NAVI\_BAIS and then the final stage using the bat algorithm (BBA). The proposed algorithm describes speed, efficiency, and accuracy so that it produces high-precision results based on fewer features

© 2019 Firas Ahmed Yonis AL\_Taie. Hosting by Science Repository.

### Introduction

Often preferable to find the best and simplest features and extract them to describe the problem of searching in different patterns. The process of selecting the best features can be considered a very important step and also depends on the nature of the problem [1]. Several algorithms for the behavior of physical or biological systems in nature have been proposed as powerful ways to improve global problems. Kennedy and Eberhard proposed a binary version of the known Optimization of a Swarm Particle (PSO) called BPSO, where the conventional PSO algorithm was modified in order to address binary optimization problems, Rashedi et al. proposed a binary version of the Gravitational Search Algorithm (GSA) called (BGSA) for feature selection, and Ramos et al. presented their version of the Harmony Search (HS) for the same purpose in the context of theft detection in power distribution systems [2]. This research contains three stages in the work start from the information technology Mutual information MI, which reduces several features and is very useful if the data is large by arranging and then cut at a certain extent and then start the next stage using the binary bat algorithm BBA that performs a random selection of partial totals Of the feature by the

value vector (0,1) Then in the final stage the internal classification process is done using the NAVI\_BAIS workbook on partial features [1, 3].

### Bat Algorithm

Considered as Bats are one of the fascinating creatures and have been of interest to many scientists and researchers for their ability to echolocation, this property can be considered a natural sonar type [4, 5]. All bats especially the small bats produce high sound pulses but it is very fast and short then waits the returned of these pulses to it after colliding in the body or a creature, the bat then calculates the distance between it and the body and also has a magnificent guiding technic to be able to recognize the Barriers and the creatures to be able to hunt at night and dark [1]. The completeness of this feature makes it able to distinguish between fixed and moving targets, and this operation is done in a small fraction of time [6, 7].

The author of this algorithm YANG considers the behavior of the bat to be an ideal method of selection and path improvement. He then

\*Correspondence to: Firas Ahmed Yonis AL\_Taie, Department of Mathematics, University of Mosul, Mosul, Iraq; E-mail: [firmas\\_a.al\\_taie@uomosul.edu.iq](mailto:firmas_a.al_taie@uomosul.edu.iq)

developed them to resemble a group of bats looking for food/prey using echo-positioning[8]. Some improvement rules have been developed as follows:

1. All bats have echo position property to estimate distance and also know the difference between (obstacles and prey) in an amazing way.
2. Bats fly in randomly velocity ( $V_i$ ) and position( $X_i$ ) with constant frequency ( $f_{min}$ ) changing in wavelength ( $\lambda$ ) and sound loudness ( $A_0$ ), bats have the ability to adjust wavelength or pulses frequency produced from it and is able to adjust the rate of pulse  $r \in [0,1]$  According to the prey dimension.
3. The loudness of bats can vary in many ways, but YANG assumes that the loudness can change within a given curve from a large positive value  $A_0$  to a minimum value ( $A_{min}$ ) and be constant.

The velocity  $V_i$  and the initial position  $X_i$  and the frequency  $f_i$  for each bat  $b_i$ . And each time the T move is determined as the maximum number of iterations, then the default movement of the bat is given by updating the position and updating the velocity using the following mathematical model:

$$V_i^k(t) = V_i^k(t - 1) + (\bar{X}^k - X_i^k(t - 1))f_i \quad \dots \dots (1)$$

$$X_i^k(t) = X_i^k(t - 1) + V_i^k(t) \quad \dots \dots (2)$$

$$f_i = f_{min} + (f_{max} - f_{min}) \beta \quad \dots \dots (3)$$

Where  $V_i^k(t)$  represents the velocity of the new bat, and  $X_i^k(t)$  represents the new position,  $\bar{X}^k$  represents the best current global solution for a site that can be considered the best for the decision variable and  $\beta$  is a random number whose value is between [0, 1].

YANG proposed the idea of using the random paths feature to improve the variance of possible solutions by selecting one of the best current solutions and then applying it to create a new solution for all bats in the group as:

$$X_{new} = X_{old} + \varepsilon \bar{A}(t) \quad \dots \dots (4)$$

Where  $\varepsilon$  is a random number and has a value of (-1, 1) and  $\bar{A}(t)$  is the average of all sounds coming from all bats in this step. The technique is balanced by adjusting the loudness and pulse emission rate as follows:

$$A_i(t + 1) = \alpha A_i(t) \quad \dots \dots (5)$$

$$r_i(t + 1) = r_i(0)[1 - \exp(-\gamma t)] \quad \dots \dots (6)$$

( $\alpha$ ) and ( $\gamma$ ) are constants, of the algorithm and in the first steps the emission rate  $r_i(0)$  and loudness  $A_i(0)$  are chosen randomly and within the period  $r_i(0) \in [0, 1]$  and  $A_i(0) \in (0, 1)$ .from the previous equations, it will be noted that bats follow the frequency change, which works to accelerate their access to the best solution.

**Binary Bat Algorithm**

A binary search space is a separate space where molecules can move to all corners of a hypercube using only 0 or 1 values[9, 10]. This means that the bat algorithm (BBA) is somewhat similar to the basic algorithm (BA), but depending on the search space where the BA operates at continuous distances while BBA search is based on the area restricted by 0 and 1 this process requires a function that changes values between 0 and 1 and is called the transfer function[11] .The process of choosing this function depends on certain conditions in which it must be met:

- 1- The scope of this function must be related to the range (0,1), as it represents the probability of a particle changing its position.
- 2- The transfer function should provide the greatest possible change of position to change the absolute velocity value. Some particles with a great absolute value of their velocity can be moved away from the optimal solution and therefore must be switch repositioned in the following iterations
- 3- The previous phrase applies for the smallest probability to change position even if the absolute velocity was too small.

The efficiency of the transfer function depends on its ability to increase the value of return for each increase in the speed of the molecules [12, 13]. in other words, whenever the probability of moving away from the best solution must increase its ability to change its position to the right direction and reduce it by reducing speed [14]. In this paper, we rely on the (sigmoid function) as a transport function by which the locations of all molecules are converted to binary values:

$$S(V_i) = \frac{1}{1 + e^{-v_i^k}} \quad \dots \dots (7)$$

According to this function can be replaced Eq. (2) with Eq. (8):

$$X_i^k = \begin{cases} 1 & \text{if } S(V_i) > \sigma \\ 0 & \text{otherwise} \end{cases} \quad \dots \dots (8)$$

Where  $\sigma \sim U(0, 1)$ .

Now equation Eq.8 can provide binary values for the position of any individual from the bat group in the network.

**Mutual Information (MI)**

There is a lot of information for two interrelated random variables and we can learn about this information and its importance through Mutual information:

$$I(X, Y) = \frac{H(X)}{H\left(\frac{X}{Y}\right)} \quad \dots \dots (9)$$

Where H be the entropy of the discrete random variable X conditioned on the discrete random variable Y taking a certain value [15].

That means the information about y that we got and is meant to be Mutual information  $I(X; Y)$ .

Otherwise,  $I(X; Y) = 0$  if X and Y are completely unrelated. The information exchanged was mainly applied to filter feature settings to measure the correlation between specific features and some class classifications[16]. There is a classical implementation of the mutual information method in several of the feature classification metrics[17]. One of the most important applications of this theory is to choose the feature as it will evaluate the importance of the features and arrange them according to their importance, in other words this theory will arrange the features in ascending order according to the impact of each feature on the final results of the data and thus will reduce the number of features included in the calculations, which accelerates the work of the program [18]. In the following relationship, it will be indicated by a wide range of features and classifications:

$$I(F, G) = \iint P(F, G) \log \frac{P(F, G)}{P(F) * P(G)} dfdc \quad \dots \dots (10)$$

Some methods depend on the evaluation of mutual information (MI) between a particular feature and the class label, this procedure is not a problem [14]. The difficulty is if the method evaluates the entire "feature set" [14]. The reason for the need to evaluate these totals of the whole

feature in a multivariate way is the communication or correlation between the features. Two individual features may not give enough information about a particular part, but it is possible to obtain sufficient information by combining them. Between N variables (X1, X2 ...XN) and the variable (Y), the chain rule is:

$$I(X1, X2, \dots, Xn; Y) = \sum_{i=1}^N I(Xi; Y/Xi - 1, Xi - 2, \dots, x1) \dots \dots (11)$$

$$fitness(Xi) = I(Xi, G) \dots \dots (12)$$

It is possible to specify the exchanged information as a type of measure to reduce uncertainty about classification labels, where the "fitness function" maximizes the values of information exchanged.

**Naïve Bayes Classifier N\_B**

Naïve Bayes is one of the great models used in the classifications, which calculates the probability that a particular feature belongs to a particular category. It is assumed that the constituent features of the research space are restricted and dependent on certain conditions [19]. Naïve Bayes performs mostly well in terms of simplicity of construction and ease of implementation. From the example that describes the "characteristic vector" which is (x\_1, x\_2, ..., x\_n), we will look for the class G which increases the probability of:

$$P\left(\frac{X}{G}\right) = P(x_1, x_2, \dots, x_n | g) \dots \dots (13)$$

Naïve Bayes will allow "conditional independence" between features by expressing that conditional probability  $P\left(\frac{X}{G}\right)$  as a result of those possibilities:

$$P(X|G) = \pi_{i=1}^n P(X_i|G) \dots \dots (14)$$

**Algorithm: The proposed, MI\_BBA-NB, algorithm**

- 1: Set the initial parameters: r, n, cc1, dd1, Max\_iteration,
- 2: Apply MI filter technique to select a subset of features depending on their importance
- 3: Set the initial velocities and positions using, Eq. (1) and Eq. (2)
- 4: Evaluate data by fitness function and set  $X_i^k$ .
- 5: Set iteration **1** from 1 to Max\_iteration.
- 6: Update data velocity and position according to Eq. (1) and Eq. (2).
- 7: When  $i \leq$  Max\_iteration stop satisfied and return get the best global solution.
- 8: Choosing the subset from the features specific by the BBA algorithm.
- 9: Return the good of features (selected features).

**Propose Algorithm**

Researchers tend to always have dependence on the experience of the fastest methods taking into account the accuracy in the results in this research consists of proposed method MI\_BBA\_NB consists of three parts, the first part it is done the use of MI technique, which that are working on reduces the size of bigdata by arranging the features according to their importance in progressive order, which starts with the most influential features on the output and finish with the least impact ,at this stage the process of selecting the most important features is done

by cutting the series of features after a certain number, which thus neglecting a large part of the features that negatively effect on the accuracy of the results. in the second stage, a binary bat algorithm is applied, in which a partial set of features is randomly selected by a vector of values (0,1). vector includes a string of binary values of 1 and 0 represents a subset of features whereas features that correspond to the number 0 are ignored and a Features that corresponds to the 1 number is chosen. The last stage is where the classification operations are performed within the features selected via BBA by the classifier NB.

**Experimental Results**

The suggested algorithm MI\_BBA\_NB is evaluated and compared with another evolutionary algorithm.

**I Datasets**

The application of the proposed algorithm to a group of bigdata is one way to verify its efficiency in solving classification problems. Table 1 is illustrated applying the algorithm to some data obtained from the UCI repository [20].

The target variable is a binary variable representing a negative or sick state = 0 and a positive or healthy state = 1.

**Table 1:** Description of the data sets used.

Dataset	# Samples	# Features	Target class
Breast	168	2906	2
Leukemia	76	7130	2
wisconsin	569	31	2

**II Evaluation Criteria**

Classification efficiency is measured by quality (SP), matthew correlation coefficient (MCC), classification accuracy (AC) and sensitivity (SE) these metrics are defined by the following:

$$SE = \frac{TP}{TP + FN} \times 100\% \dots \dots (15)$$

$$SP = \frac{TN}{FP + TN} \times 100\% \dots \dots (16)$$

$$AC = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \dots \dots (17)$$

Where TP, TN, FP, and FN be the numbers of true positive, true negative, a false positive and false negative of the confusion matrix, respectively, where the values of these criteria represent the strength of the classification process and the proportionality between them is direct.

**III Discussion and Analysis**

The algorithm proposed in this paper is compared with both the binary genetic algorithm (BGA) and the BBA original algorithm. The data set is divided into 30% as a test group in our experiments and the rest is used as training data. a 20-fold is set to obtain the best reliable rating. in (Table 2), the evaluation criteria are compared with some other evolutionary algorithms. it shows that fewer than the total number of features have

been selected and the accuracy of the proposed algorithm classification compared to other classification algorithms.

achieved the best results classification. for example, in the leukemia dataset, the accuracy (AC) of the data for the test group represented 95% in the proposed algorithm MI\_BBA and 90% in BBA. And 92% in BGA.

Table 2 shows that the data for both the training group and the test group

**Table 2:** Classification of the performance of algorithms on average of more than 20 sections (the number in parentheses represents the standard error).

Datasets	Methods	Training dataset					Testing dataset
		# selected descriptors	AC	SE	SP	MCC	AC
Breast	MI-BBA_NB	<b>13.65 (0.2497)</b>	<b>0.9653 (0.0031)</b>	<b>0.9878 (0.0029)</b>	<b>0.9852 (0.0048)</b>	<b>0.9562 (0.0056)</b>	<b>0.9563 (0.0252)</b>
	BBA	494.95 (0.2528)	0.8317 (0.0307)	0.8507 (0.0285)	0.7856 (0.0549)	0.6156 (0.0729)	0.9143 (0.0383)
	BGA	401.85 (0.3108)	0.9131 (0.0870)	0.9753 (0.0872)	0.8847 (0.1437)	0.8826 (0.1870)	0.9175 (0.0285)
Leukemia	MI-BBA_NA	<b>14.3 (0.2693)</b>	<b>0.9563 (0.0222)</b>	<b>0.9856 (0.0861)</b>	<b>0.9745 (0.0259)</b>	<b>0.9632 (0.0041)</b>	<b>0.9489 (0.0096)</b>
	BBA	465.95 (0.3086)	0.9000 (0.0789)	0.9222 (0.0750)	0.8682 (0.1496)	0.8830 (0.1749)	0.9116 (0.0506)
	BGA	397.05 (0.3621)	0.9289 (0.0821)	0.9289 (0.0735)	0.8940 (0.1862)	0.8527 (0.2086)	0.8949 (0.0471)
wisconsin	MI-BBA_NB	<b>14.2 (0.2608)</b>	<b>0.9995 (0.0017)</b>	<b>0.9925 (0.0154)</b>	<b>0.9992 (0.0058)</b>	<b>0.9989 (0.0036)</b>	<b>0.9995 (0.0024)</b>
	BBA	474 (0.2721)	0.9324 (0.0180)	0.9306 (0.0299)	0.9338 (0.0199)	0.8544 (0.0402)	0.9397 (0.0190)
	BGA	383.55 (0.3021)	0.9548 (0.248)	0.9658 (0.0258)	0.9320 (0.0403)	0.8992 (0.0541)	0.9366 (0.0096)

By comparing the previous implementation between BBA and BGA with the proposed algorithm MI-BBA\_NB, it shows us that it has a great ability in terms of accuracy of classification and also efficiency by applying them to four groups of bigdata.

**Conclusion**

The MI\_BBA\_NB method is proposed in this paper in order to improve the performance of the classification of large data, which is based on the selection of important features via the MI method and then apply BBA to the remaining features randomly and then sent to the workbook NB. The results of the proposed MI\_BBA method were compared with the results of both BBA and BGA through the (Table2). Experimental results from the dataset in the (Table2) indicate that the proposed MI\_BBA method has fewer features and has a classification performance that is higher than both the BBA and BGA.

**REFERENCES**

1. X-S Yang (2010) A new metaheuristic bat-inspired algorithm. *Nat inspired Coop Strateg Optim (NICSO 2010)* 65-74.
2. CA Moraes, EJ De Oliveira, M Khosravy, LW Oliveira, LM Honório et al. (2019) A Hybrid Bat-Inspired Algorithm for Power Transmission Expansion Planning on a Practical Brazilian. *Appl Nature-Inspired Comput Algorithms Case Stud* 71.
3. M Mafarja, I Aljarah, H Faris, AI Hammouri, A-Z Ala'M et al. (2019) Binary grasshopper optimisation algorithm approaches for feature selection problems. *Expert Syst Appl* 117: 267-286.
4. N Kwak, C-H Choi (2002) Input feature selection for classification problems. *IEEE Trans. neural networks* 13: 143-159.
5. X Deng, Y Li, J Weng, J Zhang (2019) Feature selection for text classification: A review. *Multimed. Tools Appl* 78: 3797-3816.
6. L Chaib, A Choucha, S Arif (2017) Optimal design and tuning of novel fractional order PID power system stabilizer using a new metaheuristic Bat algorithm. *Ain Shams Eng J* 8: 13-125.
7. A Entesar, O Saber, W Al-Hayani Hybridization of Genetic Algorithm with Homotopy Analysis Method for Solving Fractional Partial Differential Equations.
8. D Pebrianti, NQ Ann, L Bayuaji, NR H Abdullah, ZM Zain et al. (2019) Extended Bat Algorithm (EBA) as an Improved Searching Optimization Algorithm. *Proceedings of the 10th National Technical Seminar on Underwater System Technology 2018* 229-237.
9. NA Al-Thanoon, OS Qasim, ZY Algamal (2018) Tuning parameter estimation in SCAD-support vector machine using firefly algorithm with application in gene selection and cancer classification. *Comput Biol Med* 103: 262-268.
10. M Kang, J Kim, J-M Kim (2015) Reliable fault diagnosis for incipient low-speed bearings using fault feature analysis based on a binary bat algorithm. *Inf Sci (Ny)* 294: 423-438.
11. I Koyuncu (2018) Implementation of high-speed tangent sigmoid transfer function approximations for artificial neural network applications on FPGA. *Adv Electr Comput Eng* 18: 79-87.

12. S Ngah, RA Bakar (2017) Sigmoid Function Implementation Using the Unequal Segmentation of Differential Lookup Table and Second Order Nonlinear Function. *J Telecommun Electron Comput Eng* 9: 103-108.
13. L Petrescu, E Cazacu, C Petrescu (2015) Sigmoid functions used in hysteresis phenomenon modeling. *2015 9th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*, 521-524.
14. OS Qasim, ZY Algamal (2018) Feature selection using particle swarm optimization-based logistic regression model *Chemom Intell Lab Syst* 182: 41-46.
15. R Futrell, P Qian, E Gibson, E Fedorenko, IA Blank (2019) Syntactic dependencies correspond to word pairs with high mutual information. *Proceedings of the Fifth International Conference on Dependency Linguistics (Depling, SyntaxFest 2019)* 3-13.
16. H Patel, M Shah (2016) Rough Set Feature Selection Using Bat Algorithm.
17. M Bennasar, Y Hicks, R Setchi (2015) Feature selection using joint mutual information maximization. *Expert Syst Appl* 42: 8520-8532.
18. H Yang, J Moody (1999) Feature selection based on joint mutual information. *Proceedings of international ICSC symposium on advances in intelligent data analysis* 22-25.
19. NA Al-Thanoon, OS Qasim, ZY Algamal (2019) A new hybrid firefly algorithm and particle swarm optimization for tuning parameter estimation in penalized support vector machine with application in chemometrics. *Chemom Intell Lab Syst* 184: 142-152.
20. CL Blake, CJ Merz (2003) UCI repository of machine learning databases. University of California, Department of Information and Computer Science, Irvine, CA (1998).