



ISSN: 0067-2904

## An Integrated Information Gain with A Black Hole Algorithm for Feature Selection: A Case Study of E-mail Spam Filtering

Amaal Mahmood<sup>1\*</sup>, Adnan Hadi Mahdi Al-Helali<sup>2</sup>

<sup>1</sup>Department of fundamentals of Religion, Tarmiya, college Alemam\_Alaadm University, Baghdad, Iraq

<sup>2</sup>Cyber Security Department, Faculty of Science and Information Technology, Irbid National University, Irbid, Jordan

Received: 30/1/2022

Accepted: 28/11/2022

Published: 30/9/2023

### Abstract

The current issues in spam email detection systems are directly related to spam email classification's low accuracy and feature selection's high dimensionality. However, in machine learning (ML), feature selection (FS) as a global optimization strategy reduces data redundancy and produces a collection of precise and acceptable outcomes. A black hole algorithm-based FS algorithm is suggested in this paper for reducing the dimensionality of features and improving the accuracy of spam email classification. Each star's features are represented in binary form, with the features being transformed to binary using a sigmoid function. The proposed Binary Black Hole Algorithm (BBH) searches the feature space for the best feature subsets, and feature selection is based on a fitness function that is proportional to the accuracy achieved using a Naive Bayesian Classifier (NBC). When measuring the performance of the BBH with the SpamBase dataset, the performance of the classifier and the dimension of the selected feature vector used as a classifier input are considered. The experiments revealed that the BBH can produce good FS results even with a small set of selected features. This shows that when utilizing the NBC-based BBH, good spam email categorization accuracy is possible.

**Keywords:** E-mail Spam Filtering, Black Hole Algorithm, Feature Selection, Naive Bayesian Classifier

### 1. Introduction

Email is widely regarded as the most dependable and effective mode of communication, but it has recently become a major target for cyberattacks. Spam or junk emails account for a significant portion of this attack, as they are distributed through various protocols such as the simple mail transfer protocol (SMTP)[1][2]. As a consequence, spam emails may pose a risk to government institutions [3] [4]. In general, spam email detection relies on correctly classifying emails into spam and non-spam categories.

The majority of contemporary spam detection frameworks use machine learning approaches to classify spam emails [5-7]. However, selecting the classifiers' ideal input feature subsets, which is done through an FS process, is a serious issue that concerns email categorization. Meanwhile, most classifiers, such as the Artificial Neural Network (ANN), Support Vector Machine (SVM), and NBC, suffer from the problem of excessive data dimensionality, which is related to the FS process [7-13] High data dimensionality is thought to be prevented by restricting feature space and lowering the message's huge number of features. Irrelevant characteristics can have an effect on categorization accuracy. It can also have an impact on the amount of time it

\*Email: amaal.mohmood@gmail.com

takes to train a classifier, the cost of features, and the number of learning instances needed [14], [15].

Nature was the main source of inspiration for the researchers in developing different types of optimization algorithms [16]. Nature-inspired algorithms have been used to improve the performance of various machine learning models [17-21]. Swarm-based and evolutionary methods such as Ant Colony Optimization (ACO) [22-24], Genetic Algorithm (GA)[25-27], Artificial Bee Colony (ABC) [28], [29] Particle Swarm Optimization (PSO) [30], [31], Bat Algorithm (BA) [32-34], and Harmony Search Algorithm (HSA) have recently been used to solve FS [35] problems. The Black Hole algorithm (BH) has been developed recently for solving different optimization problems. It simulates the natural phenomenon of a black hole in the universe [36].

The problem of feature selection is addressed using a novel hybrid model that combines whale optimization methods and flower pollination algorithms. This model is founded on the idea of opposition-based learning. Experiments are conducted in two stages to assess the performance of the suggested algorithm. Ten sets of data selection features from the UCI data repository were used in the trials that made up the first phase. The proposed algorithm was put to the test in the second step by looking for spam emails. According to the first step's results, the suggested method outperformed other fundamental meta-heuristic algorithms in terms of average selection size and classification accuracy when applied to 10 UCI data sets. The results of the second step also demonstrate that the suggested algorithm has been able to detect spam emails with accuracy [37]

Investigate the potential effects of adversarial scenarios on the safety of machine learning-based systems like email spam filters. They created and tested three intrusive strategies, namely synonym replacement, ham-word injection, and spam word spacing, using natural language processing (NLP) and the Bayesian model as examples. The adversarial examples and results suggest that these techniques are effective in fooling the machine learning models [38].

Three effective binary solutions to the FS problem were described, all of which were based on the Symbiotic Organisms Search (SOS) algorithm. For the binarization of the SOS in the first and second techniques, several S-shaped and V-shaped transfer functions were employed. They were referred to as BSOSS and BSOSV. For binarization of the SOS, two new operators known as the binary mutualism phase (BMP) and the binary commensalism phase (BCP) were presented, resulting in the Efficient Binary SOS (EBSOS). 18 standard UCI datasets were used to test the suggested methods, which were then contrasted with the fundamental and significant meta-heuristic algorithms[39].

The email spam detection concept can be applied to text and image datasets. The invention of adaptive capsule networks and multi-objective feature selection for email spam detection is considered to be the primary contribution in this case. Two feature extraction methods, Term Variance (TV) and Term Frequency-Inverse Document Frequency (TF-IDF), are used when working with text datasets, while Fisher Discriminate Analysis (FDA), Walsh-Hadamard Transform (WHT), and color correlograms are utilized when working with image datasets. The hybrid meta-heuristic method Grey-Sail Fish Optimization (G-SFO) performs the multi-objective feature selection due to the length of the features appearing to be long and to reduce the training complexity [40].

The primary goal of this study is to use a hybrid soft computing model to improve the classification performance of an e-mail spam filtering system. The proposed model consists of a machine learning model, or classifier, called Naïve Bayesian Classifier (NBC), for classifying

the e-mails, and a hybrid filter called the wrapper feature selection algorithm, which is based on the Information Gain (IG) and Black Hole (BH) algorithms.

The remainder of the paper is organized as follows: The usual BH and NBC are described in Section 2, whereas the proposed algorithm is explained in Section 3. The experimental results are depicted in Section 4. Finally, the study's conclusion is presented in the final part.

## 2. Black hole algorithm

A black hole (BH) is a region in space and time (x, y, and t) that has a very strong gravitational field from which nothing can escape. Based on the general relativity concept, a sufficiently compact mass is required to deform space-time to generate a BH. Surrounding the BH is a mathematically defined surface known as the event horizon, which marks the point of no return, as anything that gets closer to it or crosses the Schwarzschild radius gets drawn into the BH and will disappear forever. The BH concept is just a region of space with so much mass that nothing can escape its gravitational pull. Whatever falls into the BH is forever lost, including light. The three main components of the BH algorithm are as follows: First, there's the black hole, which represents the best candidate (or solution) among all the options at any given iteration. Second, there are the "stars," which represent the other standard solutions or candidates. The BH was not created at random and is one of the real candidates in the population. Finally, based on their present location and a random number, all candidates are transported towards the black hole in the movement component. The BH algorithm has been successfully applied to a variety of optimization problems, including training a CNN [41], solving the traveling salesman problem (TSP) [42], and exploring the stars via Levy Flight for global optimization problems and data clustering [43]

## 3. The proposed algorithm

The proposed algorithm consists of two main phases. In the first phase, the IG method is executed. For each feature in the dataset, a specific weight is calculated using the following equation:

$$IG(F) = - \sum_{k=1}^C \frac{N_{C_k}}{N} \log_2 \frac{N_{C_k}}{N} + \frac{N_F}{N} \sum_{k=1}^C \frac{N_{F,C_k}}{N_F} \log_2 \frac{N_{F,C_k}}{N_F} + \frac{N_{\bar{F}}}{N} \sum_{k=1}^C \frac{N_{\bar{F},C_k}}{N_{\bar{F}}} \log_2 \frac{N_{\bar{F},C_k}}{N_{\bar{F}}} \quad (1)$$

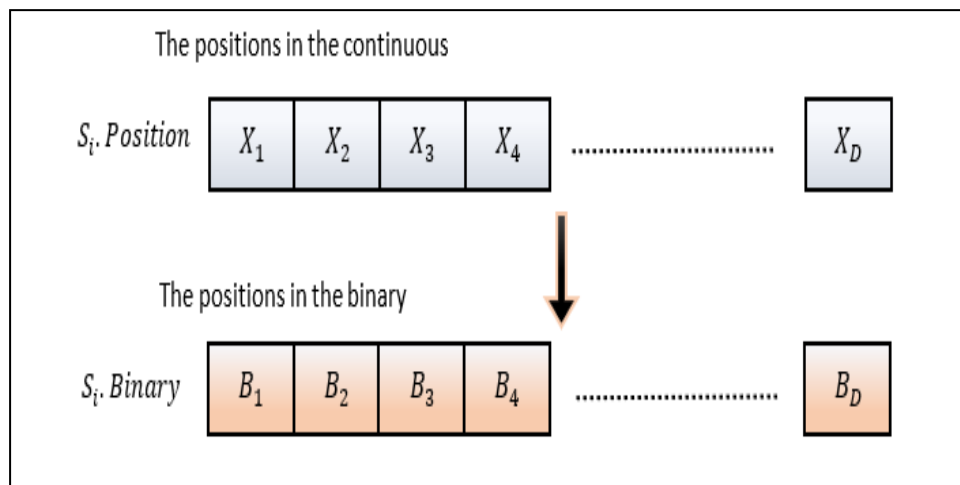
Equation (1) above consists of three main parts. In the first part, the overall entropy for the training set is calculated. In the second and third parts, the entropy of the feature is calculated. Therefore, the equation above could be simplified as follows:

$$IG(F) = H(T) - H(F) \quad (2)$$

Where H denotes the information entropy and T represents the training set. After calculating the IG values for all features in the dataset, the features should be sorted ascendingly. The features with low IG values are removed (unselected), while the rest are kept for the next stage. In order to determine the unselected features, a threshold value should be used. The second phase, which is the binary BH algorithm, is executed after the first phase is done. As stated previously, the BH algorithm is a population-based algorithm that consists of a number of stars, each of which contains a number of dimensions based on the nature of the optimization problem itself. In this study, the number of dimensions ( $D$ ) is equal to the number of features after applying IG. Moreover, each star consists of a unique representation of a solution, meaning that it contains two types of solutions. The first type of solution represents values in a continuous

form, while the second type represents the solution in a binary form that is converted from the first type.

Figure 1 below illustrates the solution representation of each star in the population.



**Figure 1:** Solution Representation for the stars

The major steps of BBH could be summarized as follows:

1-Read the inputs, which are:

- Size of Population, or number of stars.
- Number of Iterations.
- New features set after executing IG.

2-Train NBC based on the new subset of features and the training set.

3-Initialize all stars in the population randomly, as follows:

$$S_i. Position = (Upper - Lower) \times Rand(0,1) + Lower \tag{3}$$

Where  $S_i$  represents an individual star in the population, while upper and lower represent the boundaries of the search space. Equation (3) generates the positions of each star; the values of the positions are continuous, which should be converted into binary form in order to determine the selected and unselected features. The binary sequence is generated by comparing the values of the positions with the threshold 0.5, as follows:

$$S_i. Binary \begin{cases} 1 & \text{if } S_i. Position \geq 0.5 \\ 0 & \text{if } S_i. Position < 0.5 \end{cases} \tag{4}$$

Where 0 represents a removed or unselected feature and 1 represents a selected feature.

4- Calculate the fitness Function:

The proposed algorithm's fitness function is to reduce the classification performance error rate over the validation set of supplied training data, as indicated in Equation (5), while increasing the number of non-selected features (irrelevant features). A classifier should be used to calculate the fitness function. The accuracy was determined using the Nave Bayesian Classifier in this case.

$$Cost = 100 - Accuracy \tag{5}$$

where Accuracy denotes the classification accuracy rate of the classifier; in other words, the 5-fold cross validation. In this study, Naïve Bayesian Classifier (NBC) is used for calculating the classification accuracy.

#### 5-Determine the Black Hole

In each iteration, the best solution with the lowest error should be determined and set as the black hole (*BH*).

#### 6-Update the position of each star

As illustrated in Figure 1, two forms represent the solutions for each star. In order to move the stars in the searching space, the first array, which is in continuous form, should be updated via the following equation:

$$S_i.X_j^{new} = S_i.X_j^{old} + Rand(0,1) \times (BH - S_i.X_j^{old}) \quad (6)$$

Where *BH* represents the black hole or the best solution in the current iteration, and  $S_i.X_j^{new}$  and  $S_i.X_j^{old}$  represent the new and the old position of the star  $S_i$  respectively.

The updated position should be re-converted into binary form because the binary values, or the sequence of 0s and 1s, have changed. The continuous values should convert using equation (4). Then, the star should be re-evaluated using the fitness function (Step 4).

#### 7-Calculate and Check the Event Horizon (*R*)

In this step, the new positions of all stars are evaluated to determine whether they have crossed the event horizon or not. The event horizon (*R*) is calculated in each iteration based on the cost of the black hole (which should reset using Step 5) using the following equation:

$$R = \frac{BH.Cost}{\sum_{i=1}^N S_i.Cost} \quad (7)$$

The cost of each star in the population is compared with the value of *R*, and the star with a lower cost than *R* is eliminated and regenerated using initialization (Step 1).

#### 8-Checking the Stop Coding

The stop condition in the proposed BBH algorithm is the number of iterations, which is a fixed number. The algorithm stops when it reaches that number; otherwise, it executes the movement and regenerates steps again. To be more specific, if the number of loops is still lower than the number of iterations, then go to Step 5.

#### 9-Calculate the final results using the evaluation metrics, and print the final results.

The steps above can be illustrated in the following flowchart. The main flowchart of the proposed algorithm is given in Figure 2.

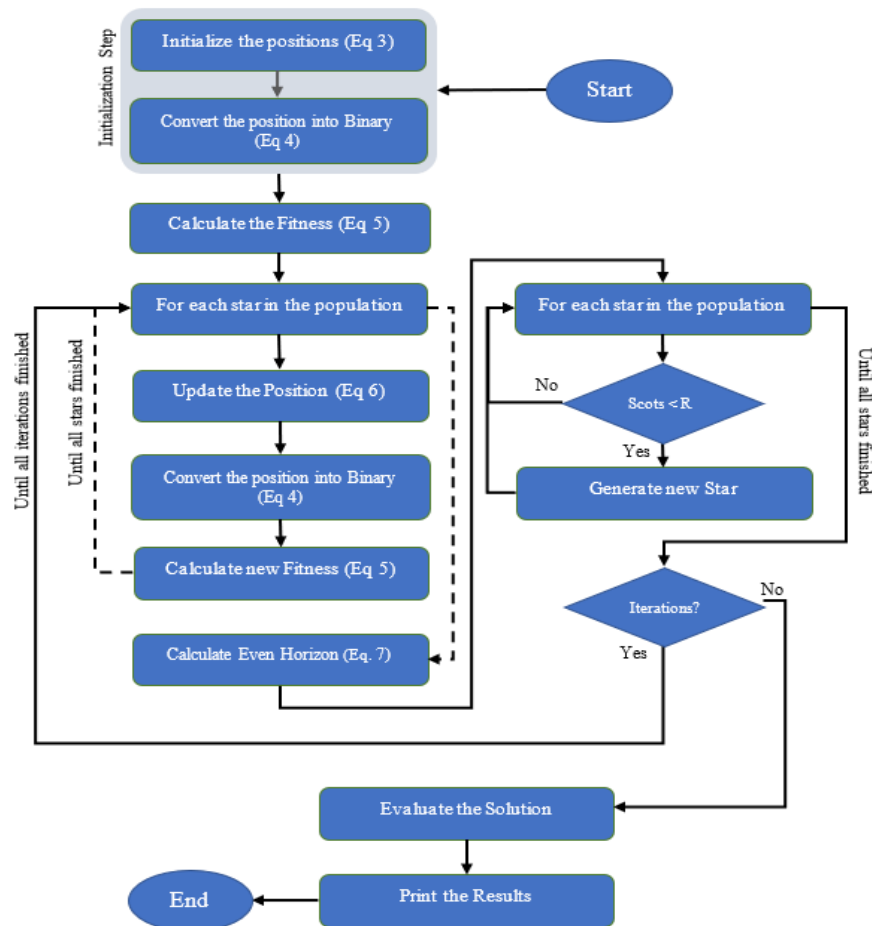


Figure 2: The flowchart of the proposed algorithm

#### 4. Results and evaluation

The dataset used for evaluating the proposed filtering system is SPAMBASE, which is a very popular dataset. It consists of around 4600 samples (emails), divided into two main classes: spam (1813  $\cong$  39%) and not-spam (2788  $\cong$  61%). Each sample or email has already been processed, and converted into feature vectors. In other words, the feature extraction step has already been completed, and the resulting vectors contained 57 attributes for each sample.

As explained in the previous section, the proposed filtering system consists of two main stages. In the first step, the information gain method, which is a filter method, is used to calculate the weights of all features, while the stage represents the BBH algorithm. The results of the first stage are displayed in Table 1.

**TABLE 1:** Results of Information Gain for all features

F	Value	F	Value	F	Value	F	Value	F	Value
1	0.0633940914	13	0.0587703815	25	0.1519977816	37	0.0861659930	49	0.0174822925
2	0.1152055475	14	0.0464861759	26	0.1713142710	38	0.0468924491	50	0.0293570741
3	0.0766860962	15	0.1484891557	27	0.1667884601	39	0.0586612127	51	0.0256490915
4	0.0989203670	16	0.2088795931	28	0.0967945596	40	0.0239568790	52	0.1468950830
5	0.0968304546	17	0.1054739180	29	0.1427346434	41	0.1071717166	53	0.1997369216
6	0.0783175396	18	0.0826097454	30	0.1259291588	42	0.0966195662	54	0.0770728390
7	0.3183780340	19	0.0635689405	31	0.1260543369	43	0.0625601933	55	0.0837693808
8	0.1242538724	20	0.1923898936	32	0.1136429636	44	0.0700218131	56	0.0875935402
9	0.1006605799	21	0.1322292620	33	0.0510928448	45	0.0366325652	57	0.0665749107
10	0.0763677165	22	0.0856459108	34	0.1016220546	46	0.0931757605		
11	0.1565211746	23	0.1993059586	35	0.1150374635	47	0.0767729685		
12	0.0385116955	24	0.2743964881	36	0.0420063978	48	0.0655275554		

Using the backward feature selection method, the features with weights lower than the threshold value are removed, while the rest of the features are kept for the next stage. The value of the threshold has been determined using the try-and-error method, which is equal to 0.06. There are 12 features with values lower than this threshold. These features are F12, F13, F14, F33, F36, F38, F39, F40, F45, F49, F50, and F51. The remaining features are used in the following stage (BBH algorithm).

As mentioned earlier in the previous section, each experiment has been executed 10 times, for different numbers of iterations and stars. Thus, there are ( $5 \times 3 \times 10 = 150$ ) run times. The results have been illustrated in Figure 3 (A, B, and C). Each figure presents the accuracy of all runs (10 times) for 100, 200, and 300 iterations and a specific number of stars. Furthermore, each figure depicts the overall accuracy average.

In general, the suggested algorithm outperformed the original in terms of accuracy, while BBH assisted NBC by selecting the most relevant features. Based on all 57 features, NBC's original accuracy was about 79.41%, whereas BBH's lowest result was around 79.41%. (88 percent ).

Figure 3 shows that the number of stars has a great impact on the searching process of the BBH algorithm; as the number of stars increased, the value of classification accuracy increased as well. The main reason behind this fact is that the possibility of finding better results is higher when more stars are utilized for the searching process. For example, if the population size is equal to 40 stars, then there are 40 possible solutions that are trying to reach better positions in a single iteration. On the other hand, these figures showed that the number of iterations has an effect on reaching the best solutions; as the number of iterations increased, the algorithm reached better classification accuracy.

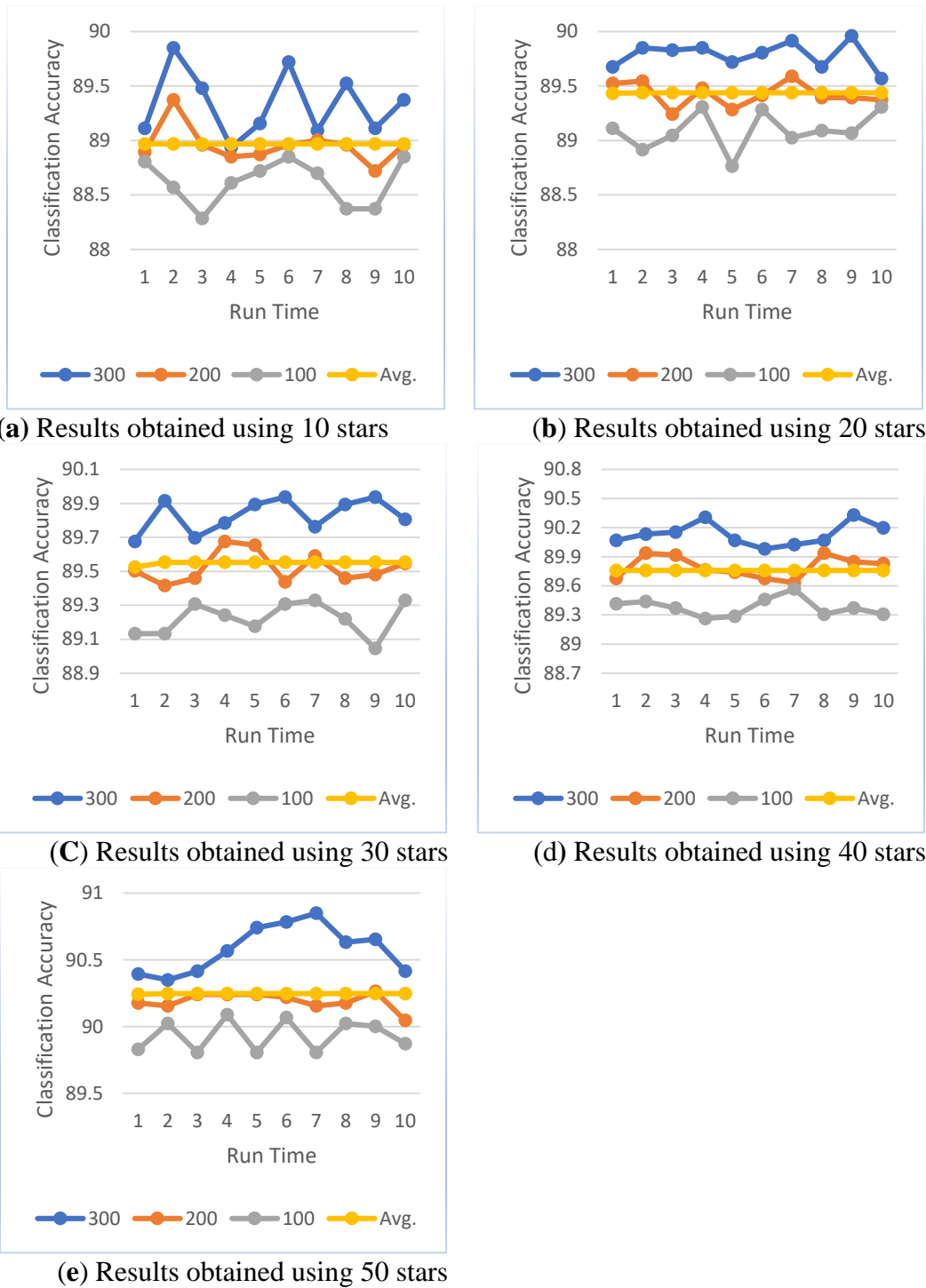


Figure 3: The Results

Table 2 below presents the other evaluation metrics, which are the precision, recall, and F-measure. It can be seen from the table below that the algorithm is stable and produces high precision and recall.



**Table 2:** Accuracy, Precision, Recall

Stars	Iterations	Features	Accuracy	PRECISION	RECALL	F-MEASURE
-	-	57	79.41751793	0.844	0.796	0.798
10	100	28	88.85024995	0.890	0.891	0.890
	200	26	89.37187568	0.888	0.886	0.886
	300	30	89.85003260	0.893	0.893	0.893
20	100	25	89.30667246	0.888	0.888	0.888
	200	24	89.58921973	0.877	0.875	0.876
	300	31	89.95870463	0.902	0.902	0.902
30	100	25	89.32840687	0.884	0.884	0.884
	200	25	89.67615736	0.894	0.895	0.894
	300	25	89.93697022	0.897	0.895	0.895
40	100	29	89.56748533	0.893	0.887	0.888
	200	27	89.93697022	0.897	0.897	0.897
	300	31	90.32818952	0.901	0.900	0.901
50	100	24	90.08911106	0.892	0.893	0.892
	200	31	90.26298631	0.899	0.897	0.897
	300	27	90.84981526	0.908	0.908	0.908

The best results obtained by BBH are compared with other email spam filtering approaches. These related works were chosen with care because they rely on the same dataset. Table 3 below presents the results. The algorithms used for the comparison are: support vector machine (SVM), K-nearest neighbor (KNN), ant colony optimization (ACO), genetic algorithm (GA), particle swarm optimization (PSO), negative selection algorithm (NSA), and distinguish feature selection (DFS) algorithm. The first three models are used to classify the SPAMBASE dataset based on all features (i.e., features = 57). while the rest of the models were enhanced using feature selection algorithms.

**Table 3:** Comparison Results

Method	Classifier	Accuracy (%)	Error Rate (%)	Ref
-	NBC	79.6	20.4	-
-	SVM	90.42	9.58	-
-	KNN	90.19	9.81	-
ACO	SVM	81.25	18.75	[44]
GA	NBC	77	23	[45]
ACO	NBC	84	16	[45]
PSO	NSA	82.62	17.38	[46]
DFS	SVM	71	29	[47]
BBH	NBC	<b>90.84</b>	<b>9.16</b>	<b>Proposed</b>

## 5. Conclusion

In this study, the Black Hole algorithm was used to select the most relevant features that will enhance the accuracy and prediction performance of the NBC. The Black Hole algorithm was combined with information gain before converting its positions to binary (0 represents unselected features and 1 represents selected features). The NBC was used in the proposed algorithm as a fitness function for evaluating the solutions. The proposed algorithm was able to

enhance the classification accuracy up to around 91% when the number of stars was equal to 50. For future studies, the proposed algorithm can be used for solving different feature selection problems, such as those in network intrusion detection systems or different medical datasets, such as those for heart disease or diabetes.

## 6. Acknowledgements

This work wouldn't have been possible without the support of Sinan Q. Salih. We are very grateful for his helpful information and for the valuable guidance.

## References

- [1] D. DeBarr and H. Wechsler, "Spam detection using Random Boost," *Pattern Recognit Lett*, vol. 33, no. 10, pp. 1237–1244, 2012, doi: 10.1016/j.patrec.2012.03.012.
- [2] Q. Wu, S. Wu, and J. Liu, "Hybrid model based on SVM with Gaussian loss function and adaptive Gaussian PSO," *Eng Appl Artif Intell*, vol. 23, no. 4, pp. 487–494, 2010, doi: 10.1016/j.engappai.2009.07.003.
- [3] S. M. Lee, D. S. Kim, J. H. Kim, and J. S. Park, "Spam detection using feature selection and parameters optimization," *CISIS 2010 - The 4th International Conference on Complex, Intelligent and Software Intensive Systems*, no. i, pp. 883–888, 2010, doi: 10.1109/CISIS.2010.116.
- [4] N. Jindal and B. Liu, "Analyzing and detecting review spam," *Proceedings - IEEE International Conference on Data Mining, ICDM*, pp. 547–552, 2007, doi: 10.1109/ICDM.2007.68.
- [5] M. Hall, "Correlation-based Feature Selection for Machine Learning," *Methodology*, vol. 21i195-i20, no. April, pp. 1–5, 1999, doi: 10.1.1.149.3848.
- [6] M. Chang and C. K. Poon, "Using phrases as features in email classification," *Journal of Systems and Software*, vol. 82, no. 6, pp. 1036–1045, 2009, doi: 10.1016/j.jss.2009.01.013.
- [7] M. Bayati and S. Jabbar, "Developing a Spam Email Detector," *Int J Eng Innov Technol*, vol. 5, no. 2, pp. 16–21, 2015.
- [8] H. Liu, X. Shi, D. Guo, Z. Zhao, and Yimin, "Feature selection combined with neural network structure optimization for HIV-1 protease cleavage site prediction," *Biomed Res Int*, vol. 2015, 2015, doi: 10.1155/2015/263586.
- [9] S. Li, P. Wang, and L. Goel, "Wind Power Forecasting Using Neural Network Ensembles with Feature Selection," *IEEE Trans Sustain Energy*, vol. 6, no. 4, pp. 1447–1456, 2015, doi: 10.1109/TSTE.2015.2441747.
- [10] M. Zhao, C. Fu, L. Ji, K. Tang, and M. Zhou, "Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes," *Expert Syst Appl*, vol. 38, no. 5, pp. 5197–5204, 2011, doi: 10.1016/j.eswa.2010.10.041.
- [11] J. Chen, H. Huang, S. Tian, and Y. Qu, "Feature selection for text classification with Naïve Bayes," *Expert Syst Appl*, vol. 36, no. 3 PART 1, pp. 5432–5435, 2009, doi: 10.1016/j.eswa.2008.06.054.
- [12] M. L. Zhang, J. M. Peña, and V. Robles, "Feature selection for multi-label naive Bayes classification," *Inf Sci (N Y)*, vol. 179, no. 19, pp. 3218–3229, 2009, doi: 10.1016/j.ins.2009.06.010.
- [13] G. Feng, J. Guo, B.-Y. Jing, and T. Sun, "Feature subset selection using naive Bayes for text classification," *Pattern Recognit Lett*, vol. 65, pp. 109–115, 2015, doi: 10.1016/j.patrec.2015.07.028.
- [14] A. Unler, A. Murat, and R. B. Chinnam, "Mr2PSO: A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification," *Inf Sci (N Y)*, vol. 181, no. 20, pp. 4625–4641, 2011, doi: 10.1016/j.ins.2010.05.037.
- [15] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005, doi: 10.1109/TPAMI.2005.159.
- [16] Sinan Q. Salih and A. A. Alsewari, "A new algorithm for normal and large-scale optimization problems: Nomadic People Optimizer," *Neural Comput Appl*, 2019, doi: 10.1007/s00521-019-04575-1.

- [17] B. Karimi, P. Mohammadi, H. Sanikhani, S. Q. Salih, and Z. M. Yaseen, "Modeling wetted areas of moisture bulb for drip irrigation systems: An enhanced empirical model and artificial neural network," *Comput Electron Agric*, vol. 178, Nov. 2020, doi: 10.1016/j.compag.2020.105767.
- [18] S. Q. Salih, A. A. Alsewari, and Z. M. Yaseen, "Pressure Vessel Design Simulation: Implementing of Multi-Swarm Particle Swarm Optimization," *Proceedings of the 2019 8th International Conference on Software and Computer Applications*, pp. 120–124, 2019.
- [19] Z. M. Yaseen, S. Naghshara, S. Q. Salih, S. Kim, A. Malik, and M. A. Ghorbani, "Lake water level modeling using newly developed hybrid data intelligence model," *Theor Appl Climatol*, vol. 141, no. 3–4, pp. 1285–1300, Aug. 2020, doi: 10.1007/s00704-020-03263-8.
- [20] S. Q. Salih and A. Rahman A. Alsewari, "Solving large-scale problems using multi-swarm particle swarm approach," *International Journal of Engineering & Technology*, vol. 7, no. 3, p. 1725, 2018, doi: 10.14419/ijet.v7i3.14742.
- [21] O. Kisi and Z. M. Yaseen, "The potential of hybrid evolutionary fuzzy intelligence model for suspended sediment concentration prediction," *Catena (Amst)*, vol. 174, no. May 2018, pp. 11–23, 2019, doi: 10.1016/j.catena.2018.10.047.
- [22] P. Moradi and M. Rostami, "Integration of graph clustering with ant colony optimization for feature selection," *Knowl Based Syst*, vol. 84, pp. 144–161, 2015, doi: 10.1016/j.knosys.2015.04.007.
- [23] A. Al-ani, "Ant Colony Optimization for Feature Subset Selection," vol. 1, no. 4, pp. 999–1002, 2007.
- [24] S. M. Vieira, J. M. C. Sousa, and T. A. Runkler, "Two cooperative ant colonies for feature selection using fuzzy models," *Expert Syst Appl*, vol. 37, no. 4, pp. 2714–2723, 2010, doi: 10.1016/j.eswa.2009.08.026.
- [25] C.-H. Cheng, T.-L. Chen, and L.-Y. Wei, "A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting," *Inf Sci (N Y)*, vol. 180, no. 9, pp. 1610–1629, 2010, doi: 10.1016/j.ins.2010.01.014.
- [26] M. M. Kabir, M. Shahjahan, and K. Murase, "A new local search based hybrid genetic algorithm for feature selection," *Neurocomputing*, vol. 74, no. 17, pp. 2914–2928, 2011, doi: 10.1016/j.neucom.2011.03.034.
- [27] C.-H. Lin, H.-Y. Chen, and Y.-S. Wu, "Study of image retrieval and classification based on adaptive features using genetic algorithm feature selection," *Expert Syst Appl*, vol. 41, no. 15, pp. 6611–6621, 2014, doi: 10.1016/j.eswa.2014.04.033.
- [28] M. Schiezero and H. Pedrini, "Data feature selection based on Artificial Bee Colony algorithm," *EURASIP J Image Video Process*, vol. 2013, no. 1, pp. 1–8, 2013, doi: 10.1186/1687-5281-2013-47.
- [29] V. Agrawal and S. Chandra, "Feature selection using Artificial Bee Colony algorithm for medical image classification," *2015 8th International Conference on Contemporary Computing, IC3 2015*, pp. 171–176, 2015, doi: 10.1109/IC3.2015.7346674.
- [30] B. Xue, M. Zhang, and W. N. Browne, "Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms," *Applied Soft Computing Journal*, vol. 18, pp. 261–276, 2014, doi: 10.1016/j.asoc.2013.09.018.
- [31] B. Xue, M. Zhang, S. Member, and W. N. Browne, "Particle Swarm Optimization for Feature Selection in Classification : A Multi-Objective Approach," pp. 1–16, 2012.
- [32] A. M. Taha, S. Der Chen, and A. Mustapha, "Multi-Swarm bat algorithm," *Research Journal of Applied Sciences, Engineering and Technology*, 2015, doi: 10.19026/rjaset.10.1839.
- [33] A. M. Taha, S.-D. Chen, and A. Mustapha, "Natural Extensions: Bat Algorithm with Memory.," *J Theor Appl Inf Technol*, vol. 79, no. 1, pp. 1–9, 2015.
- [34] A. M. Taha, S.-D. Chen, and A. Mustapha, "Bat Algorithm Based Hybrid Filter-Wrapper Approach," *Advances in Operations Research*, vol. 2015, 2015, doi: 10.1155/2015/961494.
- [35] H. H. Inbarani, M. Bagyamathi, and A. T. Azar, "A novel hybrid feature selection method based on rough set and improved harmony search," *Neural Comput Appl*, vol. 26, no. 8, pp. 1859–1880, 2015, doi: 10.1007/s00521-015-1840-0.
- [36] A. Hatamlou, "Black hole: A new heuristic optimization approach for data clustering," *Inf Sci (N Y)*, vol. 222, pp. 175–184, 2013, doi: 10.1016/j.ins.2012.08.023.
- [37] H. Mohammadzadeh and C. Author, "Case Study Email Spam Detection of Two Metaheuristic Algorithm for Optimal Feature Selection," 2020, doi: 10.20944/preprints202001.0309.v3.

- [38] B. Kuchipudi, R. T. Nannapaneni, and Q. Liao, "Adversarial machine learning for spam filters," in *ACM International Conference Proceeding Series*, Aug. 2020. doi: 10.1145/3407023.3407079.
- [39] H. Mohammadzadeh and F. S. Gharehchopogh, "Feature Selection with Binary Symbiotic Organisms Search Algorithm for Email Spam Detection," *Int J Inf Technol Decis Mak*, vol. 20, no. 01, pp. 469–515, Jan. 2021, doi: 10.1142/S0219622020500546.
- [40] K. V. Samarthrao and V. M. Rohokale, "A hybrid meta-heuristic-based multi-objective feature selection with adaptive capsule network for automated email spam detection," *Int J Intell Robot Appl*, vol. 6, no. 3, pp. 497–521, Sep. 2022, doi: 10.1007/s41315-021-00217-9.
- [41] S. Q. Salih, "A New Training Method Based on Black Hole Algorithm for Convolutional Neural Network," *Journal of Southwest Jiaotong University*, vol. 54, no. 3, pp. 1–10, 2019, doi: 10.1002/9783527678679.dg01121.
- [42] A. Hatamlou, "Solving travelling salesman problem using black hole algorithm," *Soft comput*, vol. 22, no. 24, pp. 8167–8175, 2018.
- [43] H. A. Abdulwahab, A. Noraziah, A. A. Alsewari, and S. Q. Salih, "An Enhanced Version of Black Hole Algorithm via Levy Flight for Optimization and Data Clustering Problems," *IEEE Access*, vol. 7, 2019, doi: 10.1109/access.2019.2937021.
- [44] P. Karthika, R. D & Visalakshi, "A Hybrid ACO Based Feature Selection Method for Email Spam Classification," *WSEAS Transactions on Computers*, vol. 14, pp. 171–177, 2015.
- [45] D. Karthika, P. Visalakshi, and T. Sankar, "Improving E-Mail Spam Classification using Ant Colony Optimization Algorithm," *Int J Comput Appl*, no. Ict, pp. 975–8887, 2015.
- [46] I. Idris et al., "A combined negative selection algorithm-particle swarm optimization for an email spam detection system," *Eng Appl Artif Intell*, vol. 39, pp. 33–44, 2015, doi: 10.1016/j.engappai.2014.11.001.
- [47] A. K. Uysal and S. Gunal, "A novel probabilistic feature selection method for text classification," *Knowl Based Syst*, vol. 36, pp. 226–235, 2012, doi: 10.1016/j.knosys.2012.06.005.