Cyber Terrorist Detection by using Integration of Krill Herd and Simulated Annealing Algorithms

Hassan Awad Hassan Al-Sukhni[1], Azuan Bin Ahmad[2], Madihah Mohd Saudi[3], Najwa Hayaati Mohd Alwi[4]
Faculty of Science and Technology (FST), Universiti Sains Islam Malaysia (USIM), Nilai, Negeri Sembilan, Malaysia

***Abstract*—This paper presents a technique to detect cyber terrorists suspected activities over the net by integrating the Krill Herd and Simulated Annealing algorithms. Three new level of categorizations, including low, high, and interleave have been introduced in this paper to optimize the accuracy rate. Two thousand datasets had been used for training and testing with 10-fold cross validation for this research and the simulations were performed using Matlab®. Based on the conducted experiment, this technique produced 73.01% accuracy rate for the interleave level; thus, outperforming the benchmark work. The findings can be used as a guidance and baseline work for other researchers with the same interest in this area.**

***Keywords*—Krill Herd; web content classification; cyber terrorists; simulating annealing**

I. Introduction

The cyber terrorist activities have increased lately through the web posing challenges to people around the world [1]. A cyber terrorist is defined by Al Mazari [2] as a criminal who uses computer technology and the Internet with the intention of causing violence or spreading ideology to threaten people. Furthermore, cyber terrorists are those who perform act of terrorism and profess motivations or justify their violent tactics according to their own goals and intentions [3]. Hence, this work defines terrorism as any act that is intended to harm others based on ideological motivations to justify their crimes.

It is important to reduce the number of features in order to classify the contents of cyber terrorists. The bag of words (BoW) consists of various numbers of words representations. It is one of the simpler and more preferred models as it represents web as a set of distinct words that are not compatible with each other by ignoring the order and meaning of the words. Sometimes a situation may arise when the number of words that are available in the web are considerable. A very high dimensionality might have an adverse effect on the feature space resulting in destructive influences. Besides, it degrades the performance of the entire system as well as the performance of the content classifier. Since data in content classification are of high dimensions, natural dimensionality reduction becomes a necessity to increase efficiency and accuracy [4]. The current study focuses on the dimensionality of the cyber terrorists.

This paper is presented based on the following sections. Section II discusses main related works in feature selection (FS) and cyber terrorist's domains. Section III evaluates the effect features of on three classifiers algorithms and the findings are presented in section IV. Section V concludes this paper and offers suggestions for future work.

II. Related Works

Traditional feature selection algorithms such as information gain and Chi-Squared are problem dependent and the dependencies of features are ignored [5]. This motivates the researchers to dwell into other FS methods like meta-heuristic algorithms and utilize them as feature learning for cyber terrorist's classifiers in order to be able to detect the terrorist activities in the dark web. Investigation and creation of novel approaches for dealing with the problems of feature selection dimensionality are still interesting areas for researchers mainly with regard to spam classifiers. Hence, employing the approaches of FS in [6] and [7] are discussed "(a) performance improvement such as predictive accuracy, speed of learning and quality of rules; (b) data visualization and simplification to visualize the data for model selection; and (c) dimensionality reduction and removing the noise and irrelevancy".

Some studies have been conducted on meta-heuristic techniques such as the local search and the population-based approach for the heuristics and hybrid meta-heuristics. A meta-heuristic technique is a technique that is based on computational approaches and is proven very effective as it optimizes the problem frequently through its improvement in which a candidate solution is employed. The help of this candidate solution deputed an improvement in terms of a given measure of quality [8]. The core idea of a simulated annealing designed and developed in [9] is on hil climbing-based approaches by using possibility for the escape of a local optima problem.

There are several researchers who have hybridized single-based methods with other approaches leading to improved effectiveness. For instance, [10] and [11] employed the Particle Swarm Optimization (PSO) for feature subset selection, utilizing the k-nearest neighbor as classifier, and BOW as extraction. On the other hand, they utilized filter-based approaches including Minimum Redundancy Maximum Relevance (mRMR), Information Gain (IG) and other five familiar wrapper-based approaches including ant colony optimization and genetic algorithm. Based on the results, it is revealed that the particle swarm as a feature selection method is better than information gain, term variance, fisher score and mRMR methods. Moreover, the particle swarm as feature selection is better than wrapper-based approaches utilizing all approaches of the k-nearest neighbor as the classifier. [12] Employed query expansion ranking as feature selection for sentiment analysis and compared it with state-of-art methods namely, document frequency difference, Chi square, information gain, and optimal orthogonal centroid. On the other hand, they compared four classifiers, namely Naïve

Bayes Multinomial, Support Vector Machines, Maximum Entropy Modeling, and Decision Trees. Their results showed that the proposed method of query expansion ranking as feature selection is better than state-of-arts. In addition, compared with four classifiers, the naïve Bayes multinomial is better than others as a classifiers algorithm are. Labani [13] also used a filter method named Multivariate Relative Discrimination Criterion for text classifier as feature selection. They proposed their method to reduce the redundant features in text categorization and they used three datasets to evaluate their proposed method. Then, they compared it with the state-of-art, namely Relative Discrimination Criterion, Information Gain, Gini index and Minimal-Redundancy-Maximal-Relevance. In addition, they used three classifiers algorithms, namely Multinomial naïve Bayes (MNB), Decision Tree (DT) and Multilayer Perceptron (MLP) classifiers. Based on their results, their proposed method of Multivariate Relative Discrimination Criterion as filter approach reduced the number of redundant features and increased the performance of text classifiers. On the other hand, their results did not mention which of three classifiers was the best. The next section evaluates the three classifiers algorithm to address the effect of features on the classifiers and outlines why krill heard is desired as a feature selection.

III. METHODOLOGY

A. Evaluating Three Classifier Algorithms

In order to illustrate the behaviour of three classifiers, BOW rule was applied as the extraction method.

$$\text{DB} = \sum_{i=\text{n}}^{\text{i}=1} \left(.\,(I_{i=n}^{i=1}).\ (BOWi) \right) \tag{1}$$

Where I is the first term, n is the number of terms extracted and I^i equals to Ideological Concepts; where I^1 is Salafi Jihadist ideology such as Al Qaeada, I^2 is Takfiri Ideology such as ISIS and I^3 is Spiritual Mentor such as: Usama Bin Laden, Abu Baker Al Baghdadi, and Al Mamakdsi.

The datasets were taken from *Tawhid* consisting of 800 webs. This experiment can be used for the fitness function as an evaluation measure. Table I provides the summary of the results for the three classifiers namely KNN, NB, and SVM by comparing the accuracies achieved using the BOW extraction method. The results indicate that the SVM classifier achieved the best results in DS2, with an accuracy rate of 69.94%. In addition, the NB classifier is the best classifier in DS3 and the KNN classifier is the best one in DS1. It is concluded that although the SVM outperforms the NB and KNN in terms of the accuracies achieved, the performance of document classification is still inconclusive. The number of better accuracies achieved by NB and KNN classifiers is near to SVM classifier results.

Evaluation of the results with the number of features used in each dataset taken into account shows that the NB classifier performs better when number of features is low. That is, the number of features in DS3 is 4636. On the contrary, for SVM, the accuracy result in DS2 equals to 69.94%, which has the highest number of features of 7096. Finally, for the KNN classifier, the effect of the number of features on the accuracies is inconclusive.

B. Employed Krill Herd WITH Simulated Annealing as Feature Selection

This section of the algorithm is used to describe the various techniques employed in the searching of the proposed Krill Herd. The searching technique is applied with the help of any feature selection algorithm that uses Krill Herd as feature selection (KHFS). Similar to other metaheuristic techniques, the proposed mechanism starts with an initial value of the population that is considered as the forced basis called the Krill density and the best individual that is chosen as a Krill. HSFS technique is weaker in terms of local searches and has a low convergence rate.

Due of this factor, it is easily available to run into various local optimum conditions rather than running it into the global optimum conditions [14]. In general, this is a huge drawback with harmony search algorithm. Such algorithm is used to employ various control parameters, which the KH algorithm actually avoids and does not make use of. The KH algorithm will have only one control parameter limit if common controlling parameters that are considered the maximum equation numbers and the size of the population of the Krill are excluded. HS algorithm has three control parameters namely phmcr, p par, and bw [15]. A fresh meta-heuristic algorithm like Krill Herd algorithm optimization is created to fill such gaps [16].

In this part of the study, the algorithms that are proposed to represent various webs were used as vector-space models. Therefore, as a common observation, it can be considered that the terms available as one-dimensional or multidimensional web space can be taken as di = (w_{i1}, w_{i2} . . . w_{in}). The proposed algorithm employs some representations where they codify the whole P of the features set in a vector of length m, where m is the number of the features; this is illustrated in Table 2. Each element of this vector is the label where the features are either selected or dropped. An example of solution representation is illustrated in Table II. In this case, 5 features {3, 4, 6, 7, 8 and 10} are selected. The others are dropped {1, 2, 5 and 9}.

TABLE. I. RESULTS FOR THE THREE CLASSIFIERS USING BOW

DS		**KNN**	**NB**	**SVM**	**Best**	
	# features				Best	Accuracy
DS1	4731	59.88	53.89	55.50	KNN	59.88 %
DS2	7096	55.24	53.60	69.94	SVM	69.94 %
DS3	4636	59.83	66.83	61.67	NB	66.83 %

TABLE. II. REPRESENT THE SELECTED FEATURES

1	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
0	***0***	***1***	***1***	***0***	***1***	***1***	***1***	***0***	***1***

1) Simulated annealing: The simulated annealing, as proposed by [17], can be considered a single heuristic solution, which is available on the basis of Hill Climbing methodology. The simulated annealing approach can be used to overcome the problem of the stagnation in the local Optima value. In the current study, the initial temperature is set to 2*|N|, where |N| represents the number of attributes for each dataset, and the cooling schedule is calculated as T = 0.93 * T (as adopted in [18] and [19]).

2) The proposed approach: Whenever binary optimization problems are taken into consideration, the feature selection is considered as a method which is wired to solve such problems. In the mechanism of the feature selection, the solutions are restricted to specific binary numbers namely {0, 1} values. If the KH algorithm is intended to work and compensate with this kind of feature selection method, a binary value of the version should be initially developed for this code. The solution required in this particular scenario is considered and taken as a single dimensional vector in which the length of the vector can be calculated based on the number of attributes of the original dataset. “1” or “0” represent every value in the vector (cell). Value “1” reveals that the corresponding attribute is chosen; otherwise, the value is set to “0”.

Searches space diversification and exploitation for finding out the best possible solutions intensification are two criteria that are contradictory in nature. They have to be considered whenever a metaheuristic method is defined and designed [20]. Based on the criteria that have been taken over metaheuristic, it can be broadly classified into two categories: the first one can be called population-based (e.g., swarm intelligence, evolutionary algorithms) algorithms and it can be considered that these particular algorithms are exploration oriented. However, the second one is the single-solution based (e.g. local search and simulated annealing) algorithms that are exploitation oriented. A proper balance is required in all the two stated properties above for an algorithm to have a good searching performance.

Combination of powerful properties for both algorithms KH and SA can be used to produce better results. Whenever both of them are combined and collectively worked out for a particular problem, better results are obtained. This hybridization intends to improve the exploitation property of the KH algorithm. In order to enhance the exploration in the same algorithm, tournament selection mechanism is utilized rather than the random selection. Core idea of feature selection is a multi-objective optimization solution that is offered for the problem at a place where there is Herd to contradict what to use that is substantially followed and solved. The major agenda for using the selection optimization technique is to demise the values of the features that are selected to obtain maximum classification accuracy. It can be considered that the number of values that are smaller further features in action, the better the chances of the solution for more classification, the better the solution is. The use of hybridization between these two steps is global search (KH) and local search algorithm (SA). In the proposed approach, a more sophisticated hybridization model is reflected in the system. A hybridization model between the two algorithms is also considered.

a) The Low Level of KH with SA: A hybrid approach that adopts the SA to replace the refining stage in the KH is presented here. The Hybrid algorithm integrates the explorative power of the KH with the speed of an SA algorithm in refining the solutions. In the hybrid KH, the algorithm includes two modules namely the KH module and the SA module. The KH finds the optimum region and then the SA takes over to find optimum features. Combining the two will strike the right balance between local exploitation and global exploration. The results obtained from the KH module would be helpful as the initial features selected from the SA module. The SA algorithm will be applied to the refinement and generation of the final results.

$$\text{FS-low-level}= \sum_{i=n}^{i=1}\big((\text{KH} \,.\, \text{SA}) \,.\, (L)\ \text{Si},\ \text{Ni}\big) \quad (2a)$$

Where S_i is used feature, *Ni* is unused feature, *n* is number of the features and *L* is combination of KH with SA low level.

b) The Interleaved Hybridization: In this hybrid algorithm, the local approach is embedded into the KH. After each iteration, the SA adopts the best vector from the N_{pop} as the starting point. N_{pop} is updated if the locally optimized vectors show a better fitness value than those in N_{pop} and this procedure is repeated until they come to a stopping condition.

$$\text{FS-interleave-level}= \sum_{i=1}^{n}\big((\text{KH} \,.\, \text{SA}) \,.\, (\text{I})\ \text{Si},\ \text{Ni}\big) \quad (2b)$$

Where S_i is used feature, *Ni* is unused feature, *n* is the number of features and *L* is combined KH with SA Interleaved level.

c) The High Level of KH with SA: o enhance the algorithm, a one-step SA algorithm is proposed. A fresh feature selection solution is then generated by applying the KH operations, and the following process is applied to the new solution. In this algorithm, the explorative power of KH and the fine-tuning power of SA algorithms are interleaved in each iteration to obtain a high quality of features selection.

$$\text{FS-high-level}= \sum_{i=1}^{n}\big((\text{KH} \,.\, \text{SA}) \,.\, (\text{H})\ \text{Si},\ \text{Ni}\big) \quad (3)$$

Where S_i is used feature, *Ni* is unused feature, *n* is the number of features and *L* is combined KH with SA High level.

To maintain a proper balance between all particular selected features that are available as part of the solution in each of the minimum solutions and to provide the maximum accuracy for the particular feature selection, the fitness function in Eq. (4) is utilized in both KH and SA algorithms to assess search agents.

$$fitness = \alpha\gamma_R\,(D) + \beta\frac{|R|}{|N|} \quad (4)$$

Where γ_R (D) represents the classification error rate of a given classier (each of three classifiers is used here). Furthermore, |R| is the cardinality of the selected subset and |N| is the total number of features in the dataset, α and β are two parameters corresponding to the importance of classification quality and subset length, $\alpha \in [0, 1]$ and β = (1-α) adopted from [21] and [22].

C. Parameter Settings

A better approach can be applied based on approaching the three classifications in the algorithm. For this particular, all datasets are divided cross validation. Finally, the cross validations are divided in the same manner as in [23] for assessment. In *K*-fold cross-validation, *K-1* folds are utilized for validation and training and the remaining folds can be used for testing purposes.

A total of *M* iterations can be applied for this process, and finally every single optimizer unit can be evaluated K*M times for each dataset. As a matter of resumption of the data used for training, the validation should be equal in size and all the parameters must be set as follows: the best results can be obtained whenever the maximum numbers of iterations are 100. The size of the population should be set to 10. A total of five iterations can be done for every algorithm and finally the random seeds must be used. All the parameters of SA are similar to those utilized in the previous subsection and can be the same as they are created in the previous section.

D. Empirical Study of the Impact of Different KH Parameters on Parameters on Convergence behaviour

This section aims to study the evolution of solution of the algorithms over generations under a number of settings of important parameters since such factors are vital for the effectiveness and precision of the algorithm. These are diverse configurations of four parameters i.e. Number of Krill's (NK) and Foraging Motion *Vf*; *NK* is the number of initial population. Under this condition, this study highlights the effect of single parameter changes. In particular, this section tests the following different scenarios as revealed in Table III.

Furthermore, every single designed case was executed twenty times with the repetition numbers set to fifty for all runs. Based on experiments, the case S7 was chosen to carry out tests in this section, the parameter is set to *NK=6*, and Foraging Motion, *Vf*, is 0.2.

TABLE. III. CONVERGENCE SCENARIOS

Scenarios	***NK***	***Vf***
S1	***1***	***0.1***
S2	***10***	***0.3***
S3	***15***	***0.4***
S4	***20***	***0.5***
S5	***30***	***0.7***
S6	***5***	***0.8***
S7	***6***	***0.2***

IV. RESULTS AND FINDINGS

Three different dataset having different characteristics in our experiments were used to obtain a fair comparative analysis and evaluation of the performance of the algorithms. The dataset collected from a famous Islamic terrorist websites, which are Alemarah1 (Islamic Emirate of Afghanistan, 2019). Islamion (Dabiq, 2019) and Tawhid(Islamic State Media,2019) websites. The dataset being tested consisted of two categories of topics, Islamic terrorists and non-Islamic terrorists. These topics of news constituted 600 'news' documents that were to be used as training and testing dataset for Alemarah1 news and Islamion. The Tawhid website used 800 'news' documents as training and testing using cross-validation methods to split training and testing. Table IV shows a summary description of the number of features in each dataset benchmarks that will be used in the experiments with the extraction method BOW.

Table V presents a summary of the results of the KNN accuracies and the number of feature results using BOW as the extraction method and PSO, HS, GA, KH optimization feature selection methods. It shows that a high accuracy of 62.8% is achieved with PSO optimization feature selection method for dataset DS3. The second rank for the best optimizations feature selection methods was the KH based on dataset DS1 with an accuracy of 60.9%. The table shows that the GA optimization feature selection method scored in DS2 with 58.9% accuracy. In addition, the minimum number of features with accuracy was related to KH in DS1 with the number of features 3157 out of 4731.

Table VI shows a summary of the results of the NB accuracies and the number of features results using BOW as the extraction method and PSO, HS, GA, KH optimization feature selection methods. It shows that the accuracies achieved with KH optimization feature selection method were the best resulting an accuracy of 67.3% for dataset DS3. The second rank for the best optimizations feature selection methods were for KH based on the dataset DS1 with an accuracy of 56.9%. The table shows that the HS optimization feature selection method in DS2 scored 54.03%. Furthermore, the best minimum number of features with accuracy was related with PSO in DS1 with the number of features 3150 out of 4731.

TABLE. IV. A SUMMARY DESCRIPTION OF THE NUMBER OF FEATURES IN EACH DOCUMENT DATASET

Document Set (DS)	**Source**	**# of webs**	**# of features**
DS1	Islamion	600	4731
DS2	Tawhid	800	7096
DS3	Alemarah1 news	600	4636

TABLE. V. THE KNN ACCURACIES AND NUMBER OF FEATURES RESULTS USING BOW AS EXTRACTION METHOD AND OPTIMIZATION FEATURE SELECTION METHOD

DS		***PSO***		***HS***		***GA***		***KH***	
	Original # of features	*# of features*	*Accuracy*	*# of features*	*Accuracy*	*# of features*	*Accuracy*	*# of features*	*Accuracy*
DS1	***4731***	***3654***	***60.46***	***3541***	***59.3***	***3832***	***59.5***	***3157***	***60.9***
DS2	***7096***	***6378***	***56.9***	***6921***	***57.21***	***6835***	***58.9***	***6457***	***57.6***
DS3	***4636***	***3256***	***62.8***	***3025***	***60.3***	***3293***	***60.9***	***3560***	***61.74***

TABLE. VI. THE NB ACCURACIES AND NUMBER OF FEATURES RESULTS USING BOW AS EXTRACTION METHOD AND OPTIMIZATION FEATURE SELECTION METHOD

DS		PSO		HS		GA		KH	
	Original # of features	***# of features***	***Accuracy***	***# of features***	***Accuracy***	***# of features***	***Accuracy***	***# of features***	***Accuracy***
DS1	4731	3150	54.35	3298	55	3818	55.2	3880	56.9
DS2	7096	5763	53.98	5643	54.03	4980	53.9	5870	53.9
DS3	4636	3653	66.9	3720	66.89	3911	67	3795	67.3

Table VII shows a summary of the results of the SVM accuracies and the number of features results using BOW as extraction method and PSO, HS, GA, KH optimization feature selection methods. It shows that the best accuracy was achieved with KH optimization feature selection methods equalling 71.9% for dataset DS2. The second rank for the best optimizations feature selection methods was KH based on dataset DS3 with an accuracy of 65.1%. The table shows that the KH optimization feature selection method in DS1 scored an accuracy of 59.3%. In addition, the best minimum number of features with accuracy was related to KH in DS1 with the number of features 3126 out of 4731.

Table VIII shows a summary of the results of the SA accuracy and the number of features results using BOW as extraction method and three classifiers algorithm. It shows that the SVM classifier yielded the best accuracy equalling to 71.13% for dataset DS2. The second rank for the best classifier was for NB based on dataset DS3 with an accuracy of 66%. The table shows that the KNN classifier in DS1 scored 60.41%. In addition, the best minimum number of features with accuracy was related to KNN and NB in DS1 and DS3 with number of features 2943 out of 4731 and 3420 out of 4636. The best result was achieved in dataset DS2; that is 5720 out of 7096, with KNN, but it does not offer a high accuracy. For example, while the minimum of number of features in DS2 was 5720 out of 4731 using KNN; the best accuracy was obtained when using SVM with number of features of 5796 which is higher than that of KNN. The best number of features in terms of accuracy was in DS1 using KNN.

TABLE. VII. THE SVM ACCURACIES AND NUMBER OF FEATURES RESULTS USING BOW AS EXTRACTION METHOD AND OPTIMIZATION FEATURE SELECTION METHOD

DS		PSO		HS		GA		KH	
	Original # of features	***# of features***	***Accuracy***	***# of features***	***Accuracy***	***# of features***	***Accuracy***	***# of features***	***Accuracy***
DS1	4731	4016	55.95	3951	58.9	3852	59.34	3126	59.3
DS2	7096	6128	70.94	6629	71.31	6214	70.59	6721	71.9
DS3	4636	3601	63.67	4300	64.22	3746	63.22	3529	65.1

TABLE. VIII. THE SA ACCURACIES AND NUMBER OF FEATURES RESULTS USING BOW AS EXTRACTION METHOD ON THE BENCHMARK DATASET

DS		SVM		NB		KNN	
	Original # of features	***New # of features***	***Accuracy***	***New # of features***	***Accuracy***	***New # of features***	***Accuracy***
DS1	4731	3015	57	3205	54.92	2943	60.41
DS2	7096	5796	71.13	5891	52.76	5720	55.95
DS3	4636	3583	62.8	3420	66	3509	60.17

Based on Tables V to VIII, the best results were achieved in SVM with KH with a score of 70.9% and the number of features 6721 out of 7096. In general, the KH is better than other optimizations as feature selection using SVM, but not with all results. On the other side, the SVM is better than NB and NB is better than KNN. The next section evaluates the combination of the power of the proposed KH approach and the power of SA as a feature selection.

A. The Low Level of Krill Herd with Simulated Annealing

Table IX provides a summary of the results of the three classifiers accuracies and the number of feature results using BOW as an extraction method using the low level of KH with SA optimization feature selection methods. It shows that the accuracies achieved with SVM were the best with an accuracy of 72.6% for the dataset DS2. The second rank was to NB based on the dataset DS3 with an accuracy of 68.42%. The table also shows that the SVM in DS1 scored 59.86%. In addition, the best minimum number of features in terms of accuracy was related to SVM in DS1 with a number of features 2960 out of 4731.

B. The Interleaved of Krill Herd with Simulated Annealing

Table X provides a summary of the results of accuracies of the three classifiers and the number of feature results using BOW as the extraction method and using interleaved hybridization optimization as feature selection method. It shows that the highest accuracy, 73.01%, is achieved with SVM for dataset DS2. The second and third ranks were for SVM based on dataset DS3 with 71.98% and DS1 with 60.8% accuracies respectively. In addition, the best minimum number of features in terms of accuracy was related to SVM in DS3 with a number of features 2397 out of 4636.

TABLE. IX. THE LOW LEVEL OF KH WITH SA

DS		SVM		NB		KNN	
Dataset	***Original # of features***	***# of features***	***Accuracy***	***# of features***	***Accuracy***	***# of features***	***Accuracy***
DS1	4731	2458	60.8	3019	57.93	2969	71.2
DS2	7096	4389	73.01	5211	55.68	5701	59.1
DS3	4636	2397	71.98	3102	70.1	3114	70.9

TABLE. X. THE INTERLEAVED HYBRID

DS		SVM		NB		KNN	
	Original # of features	***# of features***	***Accuracy***	***# of features***	***Accuracy***	***# of features***	***Accuracy***
DS1	4731	2458	60.8	3019	57.93	2969	71.2
DS2	7096	4389	73.01	5211	55.68	5701	59.1
DS3	4636	2397	71.98	3102	70.1	3114	70.9

C. The High Level of Krill Herd with Simulated Annealing

Table XI presents a summary of the results of the three classifiers accuracies and the number of features results using BOW as an extraction method using a high level of KH with SA optimization feature selection methods. It shows that an accuracy of 72.6% was achieved with SVM for dataset DS2. The SVM based method on the dataset DS3 ranked second with an accuracy of 66.2%. The table shows that KNN in DS1 scored 60.84%. In addition, the best minimum number of features in terms of accuracy was related to SVM in DS1 with a number of features 3006 out of 4731.

Table XII provides a summary of the results for the six optimizations as feature selections and the number of features results using BOW as an extraction method using high SVM methods. It shows that the minimum of feature selected was better when using interleave KH with SA.

TABLE. XI. THE HIGH LEVEL OF KH WITH SA

DS		SVM		NB		KNN	
	Original # of features	***# of features***	***Accuracy***	***# of features***	***Accuracy***	***# of features***	***Accuracy***
DS1	4731	3006	59.86	3725	57.6	3097	60.84
DS2	7096	6219	72.6	5490	54.1	6237	58.75
DS3	4636	3491	66.2	3915	68.42	3397	60.91

TABLE. XII. SUMMARY OF THE NEW NUMBER OF FEATURES USING SVM CLASSIFIER

		interleave	**KH**	**SA**	**GA**	**PSO**	**HS**
Data set	***Original # of features***	***New # of features***	***New # of features***	***New # of features***	***New # of features***	***New # of features***	***New # of features***
DS1	4731	<u>2458</u>	3126	3015	3852	4016	3951
DS2	7096	<u>4389</u>	6721	5796	6214	6128	6629
DS3	4636	<u>2397</u>	3529	3583	3746	3601	4300

V. CONCLUSION

The major portions that are covered in this particular study are the findings of all the near optimal features, the selection of all the optimal features. With respect to the fitness function, which is given in the criteria, it is imperative that all the values of the features are available in a specific classifier for different classes and category types. Considering all algorithms, it is shown that SVM produces the best performance and NB outperforms KNN. Based on the behaviour of all the classifiers that were obtained, the meta-heuristic such as the Harmony Search feature selection is very helpful to find out and fill the gaps of the problem of misclassification. The reaction for the algorithm HS is depicted in this study and based on selecting the fitness function for evaluation, it is concluded that the high level of combining KH with SA gives the worst performance compared to the low level of combining KH with SA and interleaves of combining KH with the SA proposed feature selection. Thus, this study proposes three combinations of combining KH with SA algorithm to address this problem. The impact of NK and Vf KH parameters is tested, and the empirical studies demonstrate that the NK parameter is set to 6, and the Vf is 0.2. The first finding that considers the proposed KH was better than HS as the feature selection, and the interleave of combined KH with SA show the best performance. After looking at the observations, and with the aim to find out the best advantage of KH, an extension of the KH algorithm is done with the help of the SA algorithm. It is concluded that the best performance is achieved when KH is interleaved with SA.

This study suggests some directions for future research. Upcoming studies should try to find out the results for the new stages that are available in the extraction methods called the pragmatic of 'Bag-of-Narratives', which focus on conweb-aware and intent-driven methods. These pragmatic curves play an important role for analysing tasks such as sentiment analysis known as a concept in which a negative connotation is generally taken into account (e.g., small seat, might turn out to be positive, if the intent is for an infant to be safely seated in it. The current study has some limitations, one important of which is the difficulty in assessing or evaluating the accurate performance of the parameters of KH. Therefore, future researches could attempt to develop a new way to find the best parameters settings.

ACKNOWLEDGMENT

This research paper supported by the university Sains Islam Malaysia Grant PPPI/UGC_0119/FST/051000/13519.

REFERENCES

[1] Weimann, G., 2016. Terrorist migration to the dark web. Perspectives on Terrorism, 10(3).

[2] Al Mazari, A., Anjariny, A.H., Habib, S.A. and Nyakwende, E., 2018. Cyber terrorism taxonomies: Definition, targets, patterns, risk factors, and mitigation strategies. In Cyber Security and Threats: Concepts, Methodologies, Tools, and Applications (pp. 608-621). IGI Global.

[3] Holbrook, Donald (2014). The Al-Qaeda Doctrine. London: Bloomsbury Publishing. pp. 30ff, 61ff, 83ff. ISBN 978-1623563141.

[4] Rogati, M. and Yang, Y., 2002, November. High-performing feature selection for text classification. In Proceedings of the eleventh international conference on Information and knowledge management (pp. 659-661). ACM.

[5] S. Wang, J .Tang, H. Liu. Embedded unsupervised feature selection. InTwenty-ninth AAAI conference on artificial intelligence 2015 Feb 10.

[6] M. Mafarja, S .Abdullah." Fuzzy modified great deluge algorithm for attribute reduction". InRecent Advances on Soft Computing and Data Mining 2014 (pp. 195-203). Springer, Cham.

[7] M. Mafarja, S .Abdullah." A fuzzy record-to-record travel algorithm for solving rough set attribute reduction". International Journal of Systems Science. 2015 Feb 17;46(3):503-12.

[8] EG .Talbi. Metaheuristics: from design to implementation. John Wiley & Sons; 2009 May 27.

[9] S .Kirkpatrick, CD. Gelatt, MP .Vecchi. "Optimization by simulated annealing". science. 1983 May 13;220(4598):671-80.

[10] Moradi, P. & Gholampour, M. 2016. A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy. Applied Soft Computing, 43: 117-130.

[11] Maldonado, S. & Huillier, G. 2013. SVM-based feature selection and classification for email filtering. In Pattern Recognition-Applications and Methods, pp. 135-148. Springer Berlin Heidelberg.

[12] Parlar, T., Özel, S.A. and Song, F., 2018. QER: a new feature selection method for sentiment analysis. Human-centric Computing and Information Sciences, 8(1), p.10.

[13] Labani, M., Moradi, P., Ahmadizar, F. and Jalili, M., 2018. A novel multivariate filter method for feature selection in text classification problems. Engineering Applications of Artificial Intelligence, 70, pp.25-37.

[14] Yang, F. Et. al. 2009. An efficient hybrid data clustering method based on K-harmonic means and Particle Swarm Optimization. Expert Systems with Applications 36(6): 9847-9852.

[15] Akay, B. & Karaboga, D. 2009. Parameter Tuning for the Artificial Bee Colony Algorithm. Computational Collective Intelligence. Semantic Web, Social Networks and Multiagent Systems, 5796. pp. 608-619. Springer Berlin Heidelberg.

[16] Gandomi, H. & Alavi, H. 2012. Krill herd: a new bio-inspired optimization algorithm. Communications in Nonlinear Science and Numerical Simulation, 17(12): 4831-4845.

[17] Kirkpatrick, S., C.D. Gelatt, & M.P.1983. Vecchi, Optimization by simulated annealing. Science, 1983. 220(4598): 671–680.

[18] Jensen, R., & Shen, Q. (2003, September). Finding rough set reducts with ant colony optimization. In Proceedings of the 2003 UK workshop on computational intelligence (Vol. 1, No. 2, pp. 15-22).

[19] Majdi M., & Seyedali M. .2017. Hybrid Whale Optimization Algorithm with Simulated Annealing for Feature Selection, Neurocomputing.

[20] Talbi, E-G. 2009. Metaheuristics From Design to Implementation. 1stEd Hobokeb, NJ, USA: John Wiley & Sons.

[21] MM .Mafarja, S .Mirjalili." Hybrid Whale Optimization Algorithm with simulated annealing for feature selection". Neurocomputing. 2017 Oct 18;260:302-12.

[22] Majdi M., & Seyedali M..2017. Hybrid Whale Optimization Algorithm with Simulated Annealing for Feature Selection, Neurocomputing.

[23] Friedman, J., T. Hastie, & R. Tibshirani.2001.The elements of statistical learning 1, Springer series in statistics Springer, Berlin.

[24] Alemarah1, Islamic Emirate of Afghanistan, (2019), retrieved from http://www.alemarah-english.org.

[25] Islamion, (2019), Dabiq, retrieved from http://islamion.com/tag/%D8%AF%D8%A7%D8%A8%D9%82.

[26] Tawhid, (2019), Islamic State Media, Retrieved from https://tawhid.website/?categories=islamic-state-media.