Optimization Algorithms to Solve Feature Selection Problem: A Review

Laith Abualigah[1], Akram Jamal Ali Aldulaimi[1], Mohammad Al Shinwan[1], Ahmad M. Khasawneh[1],
Hamzeh Alaboo[2], Mofleh Al Diaba[3], Mohammad Shehab4
[1]Faculty of Computer Sciences and Informatics, Amman Arab University, Amman, Jordan.
lx.89@ymail.com
[2]College of Computing and Informatics, Saudi Electronic University, Saudi Arabia
[3]Department of Computer Science, Al Albayt University, Al Mafraq- Jordan
4Computer Science Department, Aqaba University of Technology, Aqaba, Jordan

ABSTRACT

Feature Selection (FS) method is one of the most important data pre-processing steps in data mining domain, it is used to find the essential features subset in order to make a new subset of informative features. The model that used the informative subset such that a classification model built only with this subset would get better predictive accuracy than the model that used a complete set of features. In this paper, we provide a summary of almost all of the methods present in the literature of 2018. The goal is to provide a general presentation to variant optimization algorithms that can be applied to a wide array of machine learning problems. We converge on Filter, Wrapper and Embedded methods. Finally, we analyzed the results of several feature selection techniques on standard datasets to illustrate the applicability of feature selection techniques.

Key words: Feature selection, Optimization; Neta-heuristics, Algorithms, classification, Review.

1. INTRODUCTION

Feature selection problems is a real word problem in order to solve these problems, the researchers worked to find approaches can solve these problems well during get better performance from used algorithms mimics the inbred behavior of the animal in nature when it searches about the food [8, 26]. Therefore, these techniques are used for selecting the features that are depending on the workspace search that generates suitable and optimal feature subsets [41], [42], [34]. The feature is identified as an attribute, property, dimension, and characteristic. The increasing request of winning benefits of high-dimensional data kept forming a new challenges for data mining. In order to ensure the classification accuracy, considering the base of the original attributes preselected by the evaluation function. The optimization algorithms are used to enhance the classification effect [34, 44, 45].

Feature selection (FS) is one of the solutions that work to make a predictive that reduces the calculation mistakes of classifiers over choice significant attributes and removal redundant, unrelated and noise attributes in original data set [1]. Feature selection methods usually classified into three broad categories are wrapper, filter, and embedded models [21], [39], [46], [47].

Filter-based methods are rankers; the features are evaluated according to measures directly from the data without predictors and using rank ordering of the features. It provides no number regarding the most remarkable features [48] ,[49]. A filtering approach is good generalization ability and low computational cost; this approach can also handle with the high-scale dataset. The wrapper-based approach uses the learning methods as a classifier (as part of the fitness function) to evaluate the usefulness of subset features and subsequently obtaining the better performance of prediction. Embedded methods works with linear classifiers such as SVM are embedded in the algorithm as expanded functionality. It is also able to capture dependencies at a lower computational cost than other methods [1], [38], [50], [51].

This paper is organized as follows: A general description of the feature selection problems is presented in Section 2. The published papers are reviewed in Section 3. The general research methodology to solve the feature selection problem is showed in Section 4. The feature selection problem for classification technique is reviewed in Section 5. Datasets and evaluation criteria are stated in Section 6. The results are discussed in Section 7 and finally, the paper is concluded in Section 8.

2. FEATURE SELECTION PROBLEMS

Metaheuristic algorithms process display two inconsistent criteria are the exploration of the search space and exploitation to determining the best solution. In the area of this study, there is not the metaheuristic-based method is capable to solve every problem of FS. However, enhancements or modifications can be made to introduce a

new version of metaheuristic algorithms to enhance the performance by balancing between exploration and exploitation of the search space. This stimulus underlies utmost our tries to create a model of predictive based on hybridization approaches for solving subsets of problems, through reducing the number of features, weak relevant and irrelevant features. Practically, a subset that optimal is likely to contain only strongly relevant features. These approaches made the improvement in classification accuracy by comparing different approaches with other methods. Also, discusses the improvement that can be happened. Most optimization algorithms mimic in its work inbred behavior of animals when they search about the food in nature.

3. RELATED WORKS

Feature selection (FS) technique is one of the answers that employed to make a predictive accuracy of the search space problem. Search approaches can be summarized as shadows: thorough search, heuristic search, probabilistic search, and involuntary hybrid exploration algorithms. Metaheuristic algorithms plan is much quicker, time overwhelming and only search a particular path to finding the optimal solution [1].

Metaheuristic search typically practical to a real-world problem, and to exact varied series of computer science [5]. Also, a heuristic is suitable to treat other parts of huge data, such as diversity, and speed [6]. Heuristics-based exploration plan displays two criteria are exploitation to govern neighbor and the search area [7], [52], [53].

Different metaheuristics methods are applied to treat feature selection and literature can be found in [8]. Genetic algorithm (GA) is the furthermost inspect metaheuristics [9]. Population and single-based metaheuristic algorithms are proposed. In metaheuristic algorithm that single based such as, Hill climbing, simulated annealing [10], [11], [12], have been used, scatter search random search [13], harmony search [14], hill climbing has main disadvantages, is very tricky for opening solutions and it often times dropping in local optima.

In population-based, every method has been used from traditional memetic algorithms [15], genetic programming [16], particle swarm optimization (PSO) [17], and bat algorithm [18].The subset features have been selected by using the spider monkey optimization approach. The primary population algorithms have been given for the dataset, and the assessment of the fitness calculation was done using the SVM for classification accuracy. In order to continue or stop the process, a stopping criterion is tested. The best subset final of attributes with large accuracy of classification is defining as the better optimal results [26].

One of the recent metaheuristic whale optimization algorithms is a metaheuristic algorithm that mimics the whale hunting appliance [27]. Coral reefs optimization as a hybrid binary algorithm and simulated annealing for attribute selection (BCROSAT) can discovery the maximum accuracy and selects the minimum number of features for most data sets that are used [28]. The enhanced binary genetic algorithm and feature granulation (IBGAFG) to get feature space granular and find the informative subset of feature [4]. Instance selection is a method that reduces the size of the original training data, instance selection, and feature extraction as the combination reduces the large volume of computation time of training the classifier [29].

The hybridization between the particle swarm optimization and slap swarm algorithm creates a method called (SSAPSO), where the efficiency of the exploitation and exploration steps is enhanced [30]. For global optimization and feature selection algorithm a novel chaotic salp swarm algorithm is practical for two problems; FS problem and global optimization problem [31]. FS is the procedure to identify statistical most relevant features to improve the predictive capability of the classifiers [32]. A method for FS to enhance clustering of a document by particle swarm optimization where this approach focuses on enhancements to the current implementation of Bayesian calibration to building energy simulation [33].

The feature selection technique of water wave optimization (WWO) where this technique uses to build the text feature selection technique based on water wave optimization, which is (WWOTFS) [34]. A hybrid approach based on binary chemical reaction optimization and tabu search for FS, one of the four basic reactions are performed, in the iteration, the best solution may be check, and tabu search employs to search neighbors, which is the process of a local search [35].

An enhanced FS algorithm, Merging the ant colony optimization algorithm and FS is given FACO approach [36]. Now, with the growth in data of network and feature sets, security of the network is threatened by extra networks attacks, such as APT and DDoS attacks. To speedily detect anomalies of networks, a classification technique is extensively used in the anomaly field of data discovery. However, there are massive redundant and irrelevant features in the dataset, which are considered difficulties avoiding the algorithm of classification from creating an efficiency anomaly detection classifiers. To enhance the performance of classification for classifiers, the ant colony optimization method accepts to search the optimal subset of feature, and it selects the redundant features independent of the classifier, which can efficiently reduce the complexity time of algorithms for classification and enhance the accuracy of classification [36].

The first hybrid approach is proposed, Lee and Oh presented aggregation between algorithms local search methods, were inserted inside genetic algorithm to improve the space of search by search the most hopeful area discovered by genetic algorithm procedure [36].

Lately, hybrid metaheuristic technique displayed high-performance in solve data hard combinatorial optimization problem. Combination approaches such as hybrid between particle swarm optimization and genetic algorithm [19], and ant colony optimization with genetic algorithm [20] have also been suggested, also, there are some approaches that work local search algorithm as inside operative to balance between exploration and exploitation [19], Simulated Annealing (SA) with Genetic Algorithm [22]. [23] proposed a new wrapper-based approach for hybrid SA with crossover operator. Moreover, in [24], proposed GA was hybridized with PSO algorithm which used the SVM as a classifier. For a new hybrid technique, we refer to [25].

4. RESEARCH METHOODOLOGY

At first, to accomplish feature selection task suitably [54,55,56], we need three steps to do it:

A. *Dataset partitioning:*

(Dataset was split into three subsets: the first set of features for training, the second set for validation, and the third and final set for testing. The training dataset (first one) is employed to train the classifier that utilized while the validation dataset is employed to evaluate the performance of classifier and is applied inside the optimization fitness (algorithm). The test dataset (final dataset) for the evaluation of the whole feature selection and classifier method.

B. *Decreasing dimensionality of datasets:*

(Dataset is decreased by using techniques of measurement in order to determine features that important in the original dataset for classifying a target).

C. *Assigning values for specifying parameters:*

That using an approach (there are numbers of user parameters that specified, the values of which need to be determined for any method is applying to function suitably for the feature selection tasks) [2].

Then the feature selection technique generates informative feature subsets through select important feature by eliminating irrelevant, redundant and noisy features from the original dataset. The candidate feature subsets are evaluated by using the evaluation criteria (classification accuracy, statistical worst, statistical mean and statistical best fitness,

standard deviation and the average number of features) in order to get the best feature subsets. Then the result will be maximal accuracy of classification and minimal size of features. Figure 1 explains an overview of the research methodology:
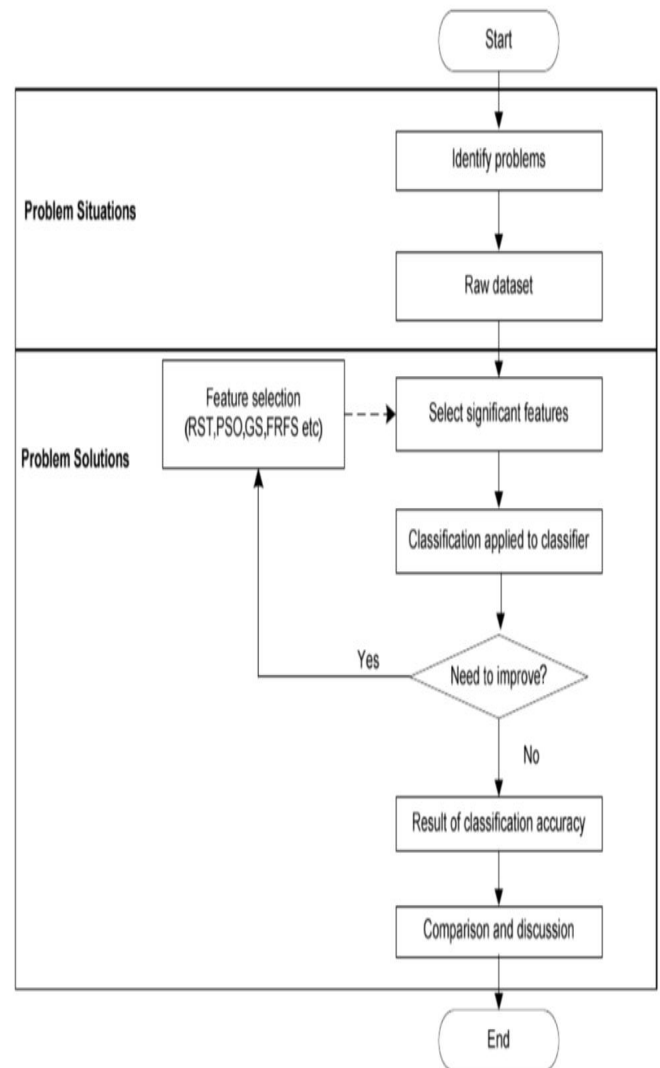

Figure 1: The methodology of research in general

5. FEATURE SELECTION FOR CLASSIFICATION

Feature selection is substantially important in machine learning, pattern classification, information retrieval, data analysis, and data mining applications. [3] Utilized impressive classification results with classifier alone; during the method does well for a dataset which does not have many features and classification works using many features were yonder their consideration. So, proposed other methods such as FS may be needed if it contained thousands of features in order to select informative feature subsets before giving them to the classifier [3]. As shown in Figure 2. The classification procedure with FS includes the input variable (represent

attributes for full dataset) and the final output variable is the pattern of classification based on features that selected from the previous feature selection process
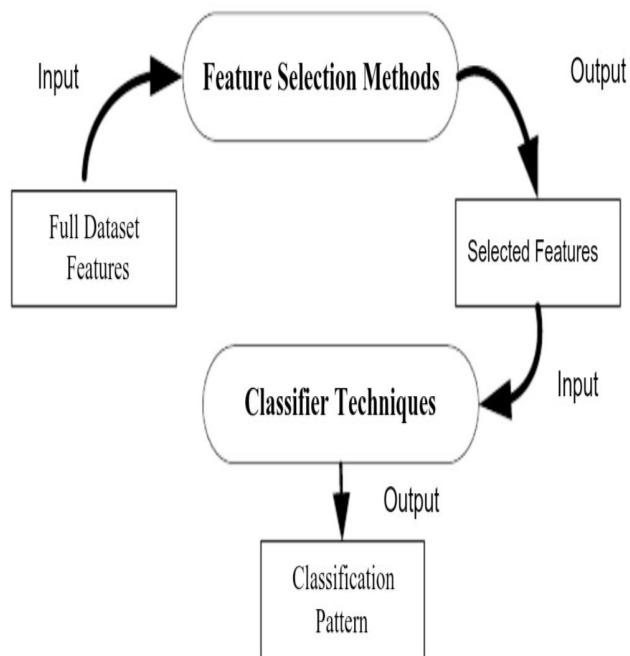

Figure 2: Feature selection techniques for classification

6. DATASETS AND EVALUATION CRITERIA

The empirical results were evaluated on different datasets from many sources such as California university (UCI) machine learning repository [37]. Also, the researchers could get datasets from the Kent Ridge Biomedical Dataset at the website: http://ieo.ugr.es/Elvira/DBCR Repository. In addition, the rest of the datasets are microarray datasets, which are multi-class data and can be downloaded from http:// csse.szu.edu.cn/staff/Zhuzx/Dataset.html [28].
Further that could gain data by collected through the RESSET database and the website is http://www2.resset.cn/product [4]. Also, there are other resources of datasets are used by researchers for this purpose.

To evaluate the dissimilar feature selection methods with other methods, the researchers used the different following evaluation criteria such as standard deviation, statistical worst, statistical mean and statistical best fitness, classification accuracy and average selection size.

7. RESULTS AND DISCUSSION

The approaches in the literature were utilized to select an optimal subset of features to increase classification execution and decrease the number of selected informative features. The datasets were from different sources and universally

accredited in this field so that these datasets were selected to have various numbers of features and instances as representatives from different kinds of issues that the proposed algorithms tested on it.

For each dataset, the objects are randomly divided into three sets namely training, validation, and testing sets in a cross-validation manner. The training dataset is used for training the used classifier while the validation dataset is used for assessing the classifier performance and is applied inside the optimization fitness. The testing dataset is kept hidden for both the classifier and the optimizer for the evaluation of the whole feature selection and system of classification [40]. These classifications are randomly initializing with solutions in the space of characteristic and are applied to decrease the fitness function in but a solution with all the features selected is forced to be one of the first solutions.

As for the comparison between these optimization algorithms, the researchers applied accurate criteria for evaluation and accepted in this area: standard deviation, analytical worst, statistical mean and statistical best fitness, classification accuracy and average selection size. Despite all these methods, but there is no specific classification to solve all feature selection problems so the range of research remains broad and open to the suggestions of researchers in the future to submit new proposals to solve feature selection problems.

8. CONCLUSION

Based on a review of the existing literature, all the researchers introduced feature selection approaches in order to solve this real word problem (feature selection problem) through suggesting new methods (based on single algorithm or hybrid between tow metaheuristic algorithms) and the results were accepted. Through this process we got tow contradictory goals are maximal accuracy of classification and minimal size of features. Also, when the feature selection problems are solved this process will give rise to speed learning, ease of rules, visualizes the data, the comprehensibility of the concepts and predictive accuracy. Reduced dimensionality by feature selection methods can also help to get a better understanding of assumed problems in data mining, machine learning, and patterns recognition etc. Finally, we concluded that the optimization algorithms are the powerful technique to solve the feature selection problem and there is no suit algorithm for all problems.

REFERENCES

[1] Liu, H., amd Motoda, H. " Feature selection for knowledge discovery and data mining" ,Springer Science & Business Media, Vol. 454, 2012.

[2] Md. Monerul kabir, md. Shahjahan and kazuyuki Murase, " A new hybrid ant colony optimization algorithm for feature selection" . Expert systems with applications, Elesevier, 2012; 3747-3763 .

https://doi.org/10.1016/j.eswa.2011.09.073

[3] N. B. Omar, F.B. Jusoand, M. S. Bin Othman and Roliana Binti Ibrahim, "Review of Feature Selection for Solving Classification Problems," JISRI,64-70.

[4] H. Dong, T. Li, R. Dang and J. Sun, "A novel hybrid genetic algorithm with granular information for feature selection and optimazation,". Applied Soft Computing, Elesevier, 2018; 33-46.

[5] Du, K. L., and Swamy, M. N. S. "Search and optimization by metaheuristics,". Springer, New York City; 2016: 434.

[6] Dhaenens, C., and Jourdan, L. "Metaheuristics for big data,". John Wiley and Sons, 2016.

[7] Talbi, E. G. "Metaheuristics: from design to implementation ,". John Wiley & Sons,2009; Vol. 74.

[8] Diao, R., and Shen, Q. "Nature inspired feature selection meta-heuristics. Artificial Intelligence Review,". 2015; 44(3): 311-340.

[9] Bala, J., Huang, J., Vafaie, H., DeJong, K., and Wechsler, H. "Hybrid learning using genetic algorithms and decision trees for pattern classification,". In IJCAI (1), August , 1995: 719-724.

[10] Caruana, R., and Freitag, D. Greedy attribute selection. In Machine Learning Proceedings 1994: pp. 28 36.

[11] Pacheco, J., Casado, S., and Núñez, L. "A variable selection method based on Tabu search for logistic regression models. European Journal of Operational Research,".2009; 199(2): 506-511

[12] Wang, Y., Li, L., Ni, J., and Huang, S. "Feature selection using tabu search with long-term memories and probabilistic neural networks,". Pattern Recognition Letters, 2009; 30(7): 661-670.

https://doi.org/10.1016/j.patrec.2009.02.001

[13] Stracuzzi, D. J., and Utgoff, P. E. "Randomized variable elimination. Journal of Machine Learning Research,". 2004; 5(Oct): 1331-1362.

[14] Diao, R., and Shen, Q. Feature selection with harmony search. IEEE Transactions on Systems, Man, and Cybernetics,". Part B (Cybernetics), 2012; 42(6): 1509-1523.

[15] Zhu, Z., Ong, Y. S., and Dash, M. "Wrapper–filter feature selection algorithm using a memetic

framework. IEEE Transactions on Systems,". Man, and Cybernetics, Part B (Cybernetics), 2007; 37(1): 70-76.

[16] Muni, D. P., Pal, N. R., and Das, J. "Genetic programming for simultaneous feature selection and classifier design,". IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2006; 36(1): 106-117.

[17] Alba, E., Garcia-Nieto, J., Jourdan, L., and Talbi, E. G. "Gene selection in cancer classification using PSO/SVM and GA/SVM hybrid algorithms,". In Evolutionary Computation, 2007. CEC 2007. IEEE Congress on 2007, September: (pp. 284-290).

[18] Nakamura, R. Y., Pereira, L. A., Costa, K. A., Rodrigues, D., Papa, J. P., and Yang, X. S. "BBA: a binary bat algorithm for feature selection. In Graphics, Patterns and Images (SIBGRAPI),". 2012 25th SIBGRAPI Conference on IEEE. 2012, August: pp. 291-297.

[19] Oh, I. S., Lee, J. S., and Moon, B. R. "Hybrid genetic algorithms for feature selection". IEEE Transactions on pattern analysis and machine intelligence, 2004; 26(11): 1424-1437.

[20] Atyabi, A., Luerssen, M., Fitzgibbon, S., and Powers, D. M." Evolutionary feature selection and electrode reduction for EEG classification,". In Evolutionary Computation (CEC), 2012 IEEE Congress on IEEE. 2012, June: pp. 1-8.

[21] Nemati, S., Basiri, M. E., Ghasem-Aghaee, N., and Aghdam, M. H. "A novel ACO–GA hybrid algorithm for feature selection in protein function prediction. Expert systems with applications,".2009; 36(10): 12086-12094.

[22] Vasant, P. "Hybrid simulated annealing and genetic algorithms for industrial production management problems. International Journal of Computational Methods,". 2010; 7(02): 279-297.

[23] Wu, J., Lu, Z., and Jin, L. "A novel hybrid genetic algorithm and simulated annealing for feature selection and kernel optimization in support vector regression. In Information Reuse and Integration (IRI),". 2012 IEEE 13th International Conference on IEEE. 2012, August: pp. 401-406.

https://doi.org/10.1109/IRI.2012.6303037

[24] Talbi, E. G., Jourdan, L., Garcia-Nieto, J., and Alba, E. "Comparison of population based metaheuristics for feature selection: Application to microarray data classification,". In Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on IEEE. 2008, March: (pp. 45-52).

[25] Yusta, S. C. "Different metaheuristic strategies to solve the feature selection problem. Pattern Recognition Letters,". 2009; 30(5): 525-534.

[26] Rani and Ramyachitra, "Microarray Cancer Gene Feature Selection Using Spider Monkey Optimization Algorithm and Cancer Classification using SVM,".8th International Conference on Advances in Computing and Communication (ICACC), 2018; pp. 108 116.

[27] Mohammad M. , N. and Aljarah, "Improved whale optimization algorithm for feature selection in Arabic sentiment analysis,". Springer, 2018.

[28] Chaokun, J. and Luo, A. "Hybrid bainary coral reefs optimazation algoriyhm with samulated annealing for feature selection in high dimentional bieomedical datasets,". Eleseiver, 2018: 102-111.

[29] Suganthi and Karunakaran," Instance selection and feature extraction using cuttle fish optimization algorithm and principal component analysis using decision tree,". Springer, 2018.

[30] Ibrahim, A. and Abd Elaziz, S. "Improved salp swarm algorithm based on particle swarm optimization for feature selection,". Journal of Ambient Intelligence and Humanized Computing, 2018.

[31] Sayed, and Ghada, M. "Anovel chaotic salp swarm algorithm for global optimization and feature selection,". Springer, 2018.

[32] Mallenahalli and Sarma, "A Tunable Particle Swarm Size Optimization Algorithm for Feature Selection,". IEEE, 2018.

[33] Abualigah and Khader, H. "A new feature selection method to improve the document clustering using particle swarm optimization algorithm,". Journal of computational science, 2017.

[34] Chen, Y. Q. and Hu, L. "Text Feature Selection Based on Water Wave Optimization Algorithm,". Tenth International Conference on Advanced Computational Intelligence (ICACI), IEEE, 2018: pp. 546 551.

[35] Chaokun, M. and Luo, J. "A Hybrid Algorithm Based on Binary Chemical Reaction Optimization and Tabu Search for Feature Selection of High-Dimensional Biomedical Data,". 2018: 733-743.

[36] Peng, C. , S. and Hu, Z. "An Improved Feature Selection Algorithm Based on Ant Colony Optimization,". IEEE, 2018: 69203-69209.

[37] Bache, K., and Lichman, M. "UCI Machine Learning Repository Online,". 2015.

[38] Sikora, R., and Piramuthu, S. "Framework for efficient feature selection in genetic algorithm based data mining,". European Journal of Operational Research, 2007; 180(2): 723-737.
https://doi.org/10.1016/j.ejor.2006.02.040

[39] Guyon, I., Gunn, S., Nikravesh, M., and Zadeh, L. A. (Eds.). "Feature extraction: foundations and applications,". Vol. 207. Springer, 2008.

[40] Hafez, Z., E., and Emary, H. "Sine Cosine Optimization Algorithm for Feature Selection,". IEEE, 2016.

[41] Abualigah, L. M., Khader, A. T., Al-Betar, M. A., & Alomari, O. A. (2017). Text feature selection with a robust weight scheme and dynamic dimension reduction to text document clustering. Expert Systems with Applications, 84, 24-36.

[42] Abualigah, L. M., & Khader, A. T. (2017). Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering. The Journal of Supercomputing, 73(11), 4773-4795.

[43] Abualigah, L. M. Q., & Hanandeh, E. S. (2015). Applying genetic algorithms to information retrieval using vector space model. International Journal of Computer Science, Engineering and Applications, 5(1), 19.

[44] Abualigah, L. M., Khader, A. T., Hanandeh, E. S., & Gandomi, A. H. (2017). A novel hybridization strategy for krill herd algorithm applied to clustering techniques. Applied Soft Computing, 60, 423-435.

[45] Abualigah, L. M., Khader, A. T., & Al-Betar, M. A. (2016, July). Multi-objectives-based text clustering technique using K-mean algorithm. In 2016 7th International Conference on Computer Science and Information Technology (CSIT) (pp. 1-6). IEEE.

[46] Abualigah, L. M. Q. (2019). Feature Selection and Enhanced Krill Herd Algorithm for Text Document Clustering. Studies in Computational Intelligence.

[47] Abualigah, L. M., Khader, A. T., Al-Betar, M. A., Alyasseri, Z. A. A., Alomari, O. A., & Hanandeh, E. S. (2017, May). Feature selection with β-hill climbing search for text clustering application. In Information and Communication Technology (PICICT), 2017 Palestinian International Conference on (pp. 22-27). IEEE.

[48] Abualigah, L. M., Khader, A. T., & Hanandeh, E. S. (2018). A hybrid strategy for krill herd algorithm with harmony search algorithm to improve the data clustering**1**. Intelligent Decision Technologies, (Preprint), 1-12.

[49] Abualigah, L. M., Khader, A. T., & Hanandeh, E. S. (2018). Hybrid clustering analysis using improved krill herd algorithm. Applied Intelligence, 1-25.

[50] Abualigah, L. M., Khader, A. T., AlBetar, M. A., & Hanandeh, E. S. (2017). Unsupervised text feature selection technique based on particle swarm optimization algorithm for improving the text clustering. EAI Google Scholar.

[51] Abualigah, L. M., Khader, A. T., & Hanandeh, E. S. (2018). A combination of objective functions and hybrid Krill herd algorithm for text document clustering analysis. Engineering Applications of Artificial Intelligence, 73, 111-125.

https://doi.org/10.1016/j.engappai.2018.05.003

[52] Abualigah, L. M., Khader, A. T., & Hanandeh, E. S. (2018). A novel weighting scheme applied to improve the text document clustering techniques. In Innovative Computing, Optimization and Its Applications (pp. 305-320). Springer, Cham.

[53] Abualigah, L. M., Khader, A. T., Al-Betar, M. A., Alyasseri, Z. A. A., Alomari, O. A., & Hanandeh, E. S. (2017). Feature selection with-hill climbing search for text clustering application.

[54] Alomari, O. A., Khader, A. T., Mohammed, A. A. B., Abualigah, L. M., Nugroho, H. E. R. U., Chandra, G. R., ... & AFFENDEY, L. S. (2017). MRMR BA: A hybrid gene selection algorithm for cancer classification. J Theoretical Appl Inf Techn, 95(12), 15.

[55] Alomari, O. A., Khader, A. T., Al-Betar, M. A., & Abualigah, L. M. (2017). Gene selection for cancer classification by combining minimum redundancy maximum relevancy and bat-inspired algorithm. International Journal of Data Mining and Bioinformatics, 19(1), 32-51.

[56] Al-Sai, Z. A., & Abualigah, L. M. (2017, May). Big data and E-government: A review. In Information Technology (ICIT), 2017 8th International Conference on (pp. 580-587). IEEE. https://doi.org/10.1109/ICITECH.2017.8080062