

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/344487870>

A parallel hybrid krill herd algorithm for feature selection

Article in International Journal of Machine Learning and Cybernetics · March 2021

DOI: 10.1007/s13042-020-01202-7

CITATIONS

19

READS

327

6 authors, including:



Laith Abualigah

Amman Arab University

170 PUBLICATIONS 4,788 CITATIONS

[SEE PROFILE](#)



Bisan Alslibi

Universiti Sains Malaysia

21 PUBLICATIONS 136 CITATIONS

[SEE PROFILE](#)



Mohammad Shehab

The World Islamic Science and Education University

47 PUBLICATIONS 814 CITATIONS

[SEE PROFILE](#)



Mohammad Al Shinwan

Amman Arab University

32 PUBLICATIONS 380 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Multi-objective Optimization Methods [View project](#)



Machine Learning techniques for relating swelling Potentiality using index geotechnical properties [View project](#)



A parallel hybrid krill herd algorithm for feature selection

Laith Abualigah¹ · Bisan Alsalibi² · Mohammad Shehab³ · Mohammad Alshinwan¹ · Ahmad M. Khasawneh¹ · Hamzeh Alabool⁴

Received: 12 December 2019 / Accepted: 18 September 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

In this paper, a novel feature selection method is introduced to tackle the problem of high-dimensional features in the text clustering application. Text clustering is a prevailing direction in big text mining; in this manner, documents are grouped into cohesive groups by using neatly selected informative features. Swarm-based optimization techniques have been widely used to select the relevant text features and shown promising results on multi-sized datasets. The performance of traditional optimization algorithms tends to fail miserably when using large-scale datasets. A novel parallel membrane-inspired framework is proposed to enhance the performance of the krill herd algorithm combined with the swap mutation strategy (MHKHA). In which the krill herd algorithm is hybridized the swap mutation strategy and incorporated within the parallel membrane framework. Finally, the k-means technique is employed based on the results of feature selection-based Krill Herd Algorithm to cluster the documents. Seven benchmark datasets of various characterizations are used. The results revealed that the proposed MHKHA produced superior results compared to other optimization methods. This paper presents an alternative method for the text mining community through cohesive and informative features.

Keywords Feature selection · Document clustering · Parallel membrane computing · Krill herd algorithm · Local search · Optimization problem

1 Introduction

The fast evolvement of Internet technology and the explosive growth of online text document content have raised the challenge of dealing with the dynamic size of text documents (TD) [1]. Text clustering (TC) is among the utmost promising unsupervised learning techniques utilized for categorizing (clustering) huge texts toward a subset of analogous clusters [2, 3]. Sharing a high level of homogeneity documents are grouped in one cluster in which different clusters exhibit a high level of heterogeneity of information (documents) [4].

A frequent dominant pattern, called vector space model (VSM), is utilized in the text mining domain such as text classification, summarization, clustering, and feature selection [5, 6]. VSM describes the components of any document as a vector of features with its weighting scores. In this pattern, the weighting score is assigned for each feature according to its importance, which is outlined as a single-dimensional space. Unfortunately, the effectiveness of traditional text analysis algorithms may significantly be degraded on large-scale text collection because of its high-dimensional vector space and the irrelevant, redundant, and uninformative features [7]. To effectively and efficiently cluster a massive text collection, feature selection (FS) techniques can be utilized to find the most informative features [8].

Feature selection approaches have been popularly used in different workable applications such as text clustering, gene selection, cancer classification, image segmentation, unconstrained face recognition, and data mining [9, 10]. The primary purpose of feature selection, as a technique, is to take off unnecessary, noisy, and immaterial features; thus, giving better performance and minimal computational time. Feature selection approaches are divided into filter methods

✉ Laith Abualigah
Aligah.2020@gmail.com

¹ Faculty of Computer Sciences and Informatics, Amman Arab University, Amman 11953, Jordan

² School of Computer Sciences, Universiti Sains Malaysia, Pinang, Malaysia

³ Computer Science Department, Aqaba University of Technology, Aqaba, Jordan

⁴ College of Computing and Informatics, Saudi Electronic University, Abha, Saudi Arabia

and wrapper methods. Filter methods take advantage of analytical and natural features to individually choose optimal features according to the association within features [11]. Then again, a supervised classification technique is applied to find the optimal features in wrapper methods [12]. Mostly, in multi-dimensional cases, an accurate exploration is taken more cost computationally [13, 14]. A promising strategy is to model and present the problem of selecting the better features as a multi-objective problem, where the objectives are the effectiveness algorithm and the number of features chosen [15, 16].

Nature-inspired algorithms are mostly meta-heuristics optimization inspired by nature. Sources of the algorithms inspiration have divided into main three kinds: evolutionary-based algorithms, swarm-based algorithms, and physics-based algorithms [17–19]. Two different features that are characteristic in all these techniques are exploration and exploitation. As stated above, swarm-based optimization algorithms simulate the group behavior of natural creatures such as the grey wolf, bat, antlion, particle, moth, verse, krill, etc.. Lately, several new evolutionary optimization algorithms are introduced and given a promising performance in solving the feature selection problem [15, 20, 21].

The contributions of this paper are listed below:

- A novel parallel membrane-inspired framework is proposed for solving the problem of feature selection in large-scale text clustering applications. Membrane computing is naturally a parallel and distributed computing framework that is used in a wide variety of computer science areas. It abstracts sophisticated computational paradigms based on the construction and functioning of living cells. This framework enabled the collaboration of the proposed hybrid krill herd algorithm to explore the search space effectively and to ensure the fast convergence in a reasonable time.
- A hybrid bio-inspired algorithm with a local search strategy, krill herd algorithm with swap mutation, is used as sub-algorithms in the Membrane computing (MC) framework to solve the feature selection problem. Krill herd is a recent optimization algorithm successfully used to solve several different hard optimization problems. Swap mutation strategy is a powerful local search strategy used to avoid the problem of stuck in local optimum. In the Swap mutation strategy, two random positions in the search space are selected and replaced with the selected areas. However, stuck in local optima is a drawback since optimization algorithms usually do not have significant randomization behaviors while solving optimization problems.

- The k-means clustering, which is a traditional clustering technique, is employed to partition a massive collection of documents using the resulted features achieved by the proposed feature selection algorithm.

For the experimental in this paper, seven benchmark datasets of various complexities and sizes are assessed. The sensitivity investigation is conducted to analyze the impact of the krill herd algorithm parameters during the convergence behavior. Examination with the k-mean technique revealed that the proposed MHKHA produced the optimal outcomes for most experimented datasets compared to other recently published methods in the literature according to the evaluation measures (i.e., F-measure, precision, recall, and accuracy). As well as, the proposed method yielded the best outcomes with all the similar techniques examined by the significance test (i.e., *t*-test). Thus, this presents a new option for the text mining community through cohesive and informative features to deal with big data.

The remaining of this paper is divided into six sections. The feature selection problem is illustrated in Sect. 2. Section 3 provides a summary of the proposed algorithm procedures. Section 4 provides experimental results and discussion. Finally, conclusion and future work directions are given in Sect. 5.

2 Feature selection problem

2.1 Mathematical model

The feature selection problem is represented as an mathematical optimization dilemma by a unique paradigm to obtain new representative features. Besides, it reduces non-informative features (irrelevant feature). The mathematical notations of the feature selection problem are explained as follows:

This is a full description of the feature selection problem, which is written in three points as follows:

- Given F as a huge collection of text features that belong to a document in the dataset (D) $F_i = \{f_{i,1}, f_{i,2}, \dots, f_{i,j}, \dots, f_{i,(t-1)}, f_{i,(t)}\}$. Where t in $f_{i,t}$ shows how many unique features in the dataset (D), j presents the feature j , and i is the document number.
- Let $NF_i = \{nf_{i,1}, nf_{i,2}, \dots, nf_{i,j}, \dots, nf_{i,(m-1)}, nf_{i,(m)}\}$, which presents a distinct subset of illuminating features (useful) taken by the proposed algorithm (i.e., MHKHA) with a new smaller dimension space, m presents the new length

of the obtained features by the proposed feature selection algorithm.

- Note, $nf_{ij} \in \{0, 1\}$, $j = 1, 2, 3, \dots, (m - 1), m$. If $nf_{ij} = 1$ indicates that the feature number j is elected in the given document (document number i), if $nf_{ij} = 0$, it indicates that the feature number j does not elected in the given document (document number i) [3, 22].

2.2 Text pre-processing

Generally, to deal with any of the document analysis techniques (i.e., text retrieval, text segmentation, text document clustering, text feature selection, and others), a new data presentation is needed to be shown by a mathematical matrix. Which is used to convert document features (document contents) into a standard format (numerical matrix) to become flexible and easy to use in the underlying clustering algorithm [6, 23]. The steps used in converting the document contents into the numerical matrix are named the pre-processing steps. These procedures are arranged into three main levels: Tokenization preperation, remove all stop words, and find the terms stemming [22, 24].

2.3 Term weighting (TF-IDF)

One of the primary measurements is the term frequency alongside with inverse document frequency (TF-IDF). Which is utilized for computing the feature weighting score (weight). Every document contents are presented as a vector of scores (see Eq. (1)):

$$d_i = (w_{i,1}, \dots, w_{i,j}, \dots, w_{i,(t-1)}, w_{i,t}) \quad (1)$$

The term weighting score is computed for all terms using the term frequency (TF), dealing with each document components alone combined with other factors as mentioned in Eq. (2). If the term frequency of a term is large and it appeared in some (few) documents, this term is beneficial to differentiate amongst the documents features [3]. The term weighting is computed by Eq. (2).

$$w_{ij} = tf(i, j) * idf(i, j), \quad (2)$$

where,

$$idf(i, j) = \log(n/df(j)), \quad (3)$$

w_{ij} denotes to the weight score (value) of the j , h term in the document number i , and $tf(i, j)$ represents the frequency

value of the term number j in the document number i . $idf(i, j)$ denotes to a sensitive factor utilized to increase the importance of the low-frequency term, which has been shown in some documents as Eq. (4). $idf(i, j)$, where n denotes to all documents involved the original dataset, and $df(j)$ represents how many documents contain the term number j . The following expression is the standard format that used to present the documents contents in a common way (mathematical matrix) using the vector space model (VSM) [22].

$$VSM = \begin{bmatrix} w_{1,1} & \dots & w_{1,(t-1)} & w_{1,t} \\ \vdots & \ddots & \vdots & \vdots \\ \dots & \dots & \dots & \dots \\ w_{(n-1),1} & \dots & \dots & w_{(n-1),t} \\ w_{n,1} & \dots & w_{n,(t-1)} & w_{n,t} \end{bmatrix} \quad (4)$$

2.4 Solution representation

In the introduced feature selection method, parallel membrane-driven hybrid krill herd algorithm (MHKHA), each a candidate solution denotes to a new smaller number of features for a document as the given in Solution 1. The solutions of the feature selection algorithm is a set of krills that is designated as one vector, every vector owns several elements according to the total number of extracted features in the given document and each location in that vector corresponding to one feature. For instance, the j_{th} position in any krill (any solution) indicates to the situation of the j_{th} feature if it is chosen as an informative feature or not. The feature selection algorithm begins with a set of the random first population and develops its solutions to realize the best obtained solution [25].

As shown in Table 1, If the j_{th} position holds a value corresponding to 1, it indicates that the j_{th} one is elected to be an informative feature. But, If the j_{th} position holds a value corresponding to 0, it indicates that the j_{th} feature is not elected (noninformative feature). Else, If the j_{th} position holds a value corresponding to -1, it indicates that the j_{th} feature number is not included in the initial document.

2.5 Fitness function

The fitness function is a kind of external evaluation measures, which is utilized to assess the candidate solutions generated using the proposed algorithm (MHKHA). In each cycle, the fitness function is measured to judge if there is an enhancement (Increase its fitness function value) found

Table 1 Solution representation

X	1	0	-1	1	-1	1	1	0	-1	1
---	---	---	----	---	----	---	---	---	----	---

in the fitness function cost of the current solution to make a decision either take or reject it. Eventually, the agent with the high fitness value is considered as the optimal solution archived by the algorithm so far [3].

The mean absolute difference (MAD) measurement is utilized in this paper as a fitness function to evaluate the solutions in MHKHA for solving the feature selection problem. This measure used the weighting score (i.e., TF-IDF scheme) for assessing the features (solution positions) [26]. The MAD measure is utilized in the text mining domain especially in this area to give a weight score for each term by determining the variation value between each feature with the computed mean value, which is computed by using Eq. (5).

$$MAD_{(Xi)} = \frac{1}{a_i} \sum_{j=1}^t |x_{ij} - \bar{x}_i|, \quad (5)$$

where,

$$\bar{x}_i = \left(\frac{1}{a_i} \right) \sum_{j=1}^t x_{ij}, \quad (6)$$

$MAD_{(Xi)}$ denotes to the fitness value of the agent i in the candidate solutions, x_{ij} denotes to the weight score of the

feature j in the document i , and a_i is the number chosen features for the document i . t is the number of all individuals feature at the level of all documents in the given dataset (D). \bar{x}_i denotes to the variation value of the vector i .

3 The proposed MHKHA

3.1 Overview of the proposed method

Newly, the general text pre-processing proceedings are utilized to choose optimal features to decrease the running cost and improve the execution of the document process (i.e., text clustering and text feature selection) [27]. Consequently, in this paper, we introduced a novel approach for obtaining an accurate subset of more informative features. The results of the presented method can improve the text process during any text analysis steps. Algorithm 1 shows the actions of the proposed method (feature determination for text clustering purpose) by using the proposed dynamic membrane-driven for the hybrid krill herd algorithm with the swap mutation strategy for tackling the feature selection problem and enhance the effectiveness of the documents clustering problem.

Algorithm 1 The proposed feature selection

- 1: **Input:** Set of a huge documents.
 - 2: **Output:** Documents Clusters.
 - 3: **Algorithm**
 - 4: **First step-(Text preprocessing)**
 - 5: Text preprocessing steps as in Section 2.2.
 - 6: Convert the document features into a numerical matrix.
 - 7: **Second step-(Membrane-driven hybrid krill herd algorithm for feature selection)**
 - 8: Apply the proposed Membrane-driven hybrid KHA with Swap mutation.
 - 9: Krill herd algorithm
 - 10: Swap mutation strategy
 - 11: Return a new subset of selected features.
 - 12: **Third step-(K-mean text clustering)**
 - 13: Apply the k-mean technique on the selected features [3, 28].
 - 14: Convert text documents presentation into a numerical matrix.
 - 15: Return the documents clusters.
-

As shown in Algorithm 1, in the initial round of the proposed approach (text pre-processing steps), the pre-processing step is utilized to draw the text features in a mathematical manner (numerical matrix), next transformed the document features into a pattern of features weight. The main aim of this step is to deal with the features in the analytical calculations.

In the second round, membrane-driven hybrid krill herd algorithm with swap mutation is proposed for addressing the feature selection problem for excluding uninformative features from the original text to enhance the production of the next step (text clustering). In this step, the membrane-driven framework is combined with the hybrid krill herd algorithm to select the more related features from the given datasets. Then the hybrid algorithm is introduced by adding the swap mutation as a local search strategy into the krill herd algorithm to adjust its solutions. Generally, the feature selection procedure is recognized as a necessary pre-processing action in numerous domains such as pattern recognition, text mining, machine learning, big data, and continue.

Finally, in the third round, the k-means technique is employed to cluster the documents into similar groups by using the obtained subset of more informative features from the proposed feature selection algorithm (MHKHA).

Since krill herd algorithm has been suggested for solving discrete problems, it is combined with many other techniques components to improve its ability in solving these problems. In this paper, we introduced a hybrid krill herd algorithm with a local search strategy under a novel membrane-inspired feature selection method to enhance the performance of krill herd algorithm, namely, dynamic

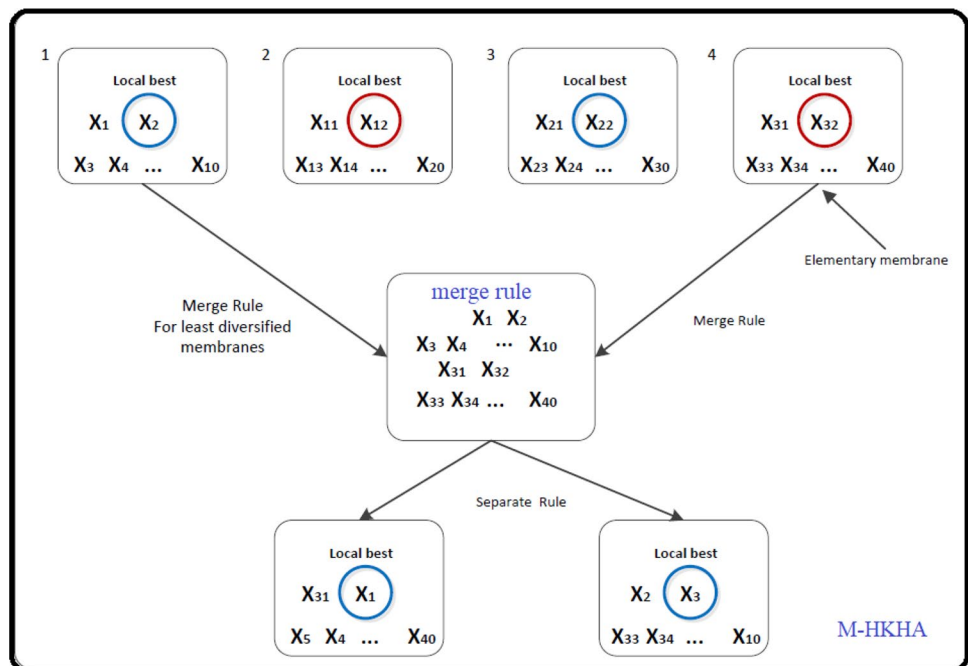
membrane-driven hybrid krill herd algorithm with a local search strategy for addressing feature selection problem (MHKHA). Where a binary hybrid krill herd algorithm with swap mutation strategy under the framework of membrane computing is applied.

The proposed algorithm (MHKHA) consists of five central steps: dynamic membrane-driven construction, motion calculations, genetic operators, evaluation, local search strategy (i.e., swap mutation). These procedures will be described in the next subsections.

3.2 Dynamic membrane-driven

Existing works on membrane-inspired evolutionary algorithms have proved that this method achieved success in addressing many soft and hard real-world and other optimization problems. However, most of them use a fixed number of membranes and thus static membrane structure where the population diversity is maintained through the membrane distribution itself. To tackle these issues, this paper introduced a novel parallel membrane-driven hybrid krill herd algorithm and swap mutation for solving the problem of the text feature selection. The use of different algorithms in different membranes can greatly help in maintaining population diversity throughout the search process. Furthermore, the number of membranes is dynamically adjusted during the optimization process, enabling better stability between exploration and exploitation search operations. The main highlights of the proposed method can be summarized as the following:

Fig. 1 Illustration of merge and separation rules inside membranes



- Membrane computing framework with a dynamic structure: If no change in the fitness triggered for a specified number of iterations, the two least diverse membranes will be merged and their sub-populations will undergo swap mutation operator to further enhance their population diversity.
- Optimization algorithm: Different optimization techniques, namely, krill herd and swap mutation, is used as a sub-algorithm in different membranes.

The flowchart of the suggested process is presented in Fig. 1. It begins by determining the arrangement of the layer (membrane) arrangement, inclusive of the original number of primary membranes and the population estimate. Alongside this, the membrane composition will be initialized, where a particular number of first layers will be located inside the skin membrane. Next, the solutions will be scholastically allocated in the primary segments. In each central membrane, the introduced hybrid krill herd with swap mutation algorithm is working simultaneously (parallel) to select the optimal subset of features, as illustrated in Fig. 1.

Later, a set number of improvement repetitions, knowledge regarding the most suitable local solution from all layers will be delivered to the surface layer. Then, the domestic most appropriate solution will be analyzed, and the global best solution will be transported from the surface to other layers by employing universal communication control. In case that the best fitness cost has not been developed for a particularized number of improvement iterations, population diversity in each elementary membrane will be calculated. The two least diversified layers will be merged using the merge rule, and their solutions will undergo swap mutation operators to enhance the population diversity further. After that, the merged membrane will be separated into two layers, and the evolutionary process of the proposed hybrid krill herd with a swap mutation algorithm will be performed in each membrane.

The rule will be repeated until the finish condition is reached. The main procedures are described in further detail as below:

Initialization: in this step, the configuration of the suggested membrane arrangement is initialized involving the membrane arrangement, the number of membranes m , the number of symbol-objects (solutions/agents), the repetition of communication (n), and the higher and lower bound of possible values.

Membrane structure construction: in the first setup of the proposed arrangement in this framework, $m - 1$ primary

layers will be located in the surface layer, which is described as:

$$[[]_2 []_3 []_4, \dots []_m]_1. \quad (7)$$

Following the P arrangement with current membranes notations, the introduced algorithm is determined by these next notations:

Evolution rules: elementary membranes with even number will apply the evolutionary process of the krill herd algorithm. Elementary membranes with an odd number will apply the evolutionary process of the swap mutation.

Communication rules: the integrated communication commands are required to help the sharing of objects over various membranes. The motivation behind these communications is that genetic transfer of information across membranes can aid in genetic leaps out of possible local optima. Especially, these rules are subject to transfer the most reliable local solution indicated by X_{lbm} from each $m - 1$ primary layers into the surface layer and when the global most suitable solution shown as X_{gb} from the skin layer return to each $m - 1$ essential layers. Send-In and send-out transfer rules can be expressed as follows:

$$[a^n, X_{lbm}]_i \rightarrow [a^0]_i X_{lbm}, i \in 2, 3, \dots m \quad (8)$$

$$a^j \rightarrow a_i^{j+1}, i \in 2, 3, \dots m, j \in 0, 1, \dots n \quad (9)$$

In Eq. (8), the procedure transfers the most suitable solution in the local area from each position of the elementary membranes, over the given $m - 1$, to the skin layers. a is used to initiate the local communication practice where it is started to a^0 and increase in each repetition by using the given rule as in Eq. (9). The number of connections is expressed by n . After performing the given rule in Eq. (8), a will be started.

$$X_{gb} []_i \rightarrow [X_{gb}]_i, i \in 2, 3, \dots m \quad (10)$$

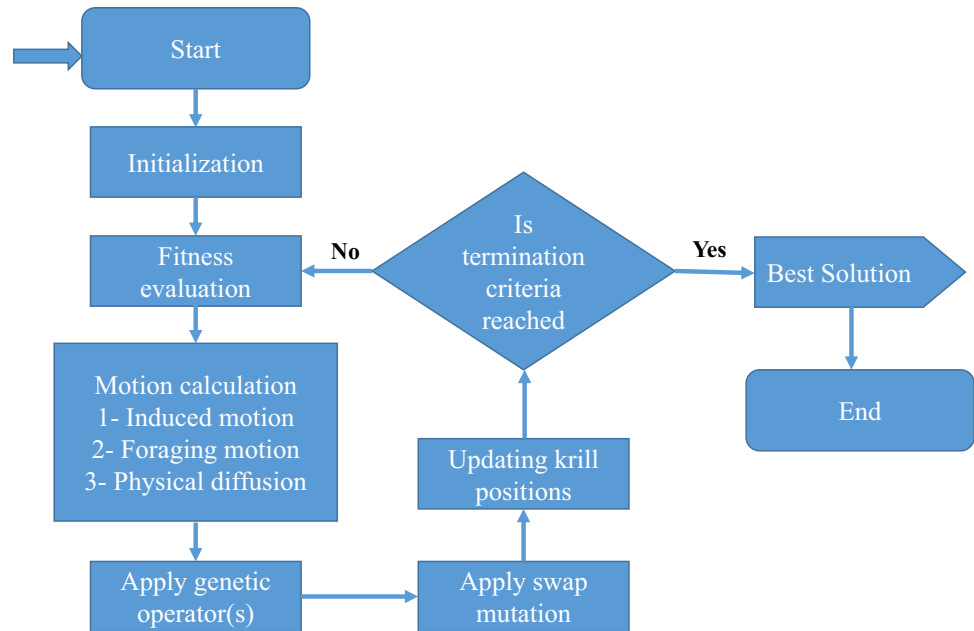
This procedure gives the global optimal solution from the covering the elementary membranes, over the given $m - 1$

Merging rule: if the best solution has not developed over a specific number of continuous iterations (set to 10 throughout the paper), the diversity of population will be calculated in each membrane and the least two diversified membranes will be merged using the following merging rule.

$$c^k, []_n []_k \rightarrow []_l, [c^j \rightarrow c^{j+1}]_m, j \in 0, 1, \dots k \quad (11)$$

Rule as in Eq. (11) is invoked to merge the least two diversified membranes into a single membrane, where all of their

Fig. 2 Illustration of the general structure of the proposed MHKHA



objects will be placed inside the merged membrane. n, k, l are arbitrary membrane labels. Rule (11) is utilized to renew the given counter (i.e., c) in each repetition.

Diversity calculation: though various proposed methods for maintaining the population diversity, one of the most generally utilized genotypic-based diversity analysis is used in this paper, namely the pair-wise Hamming distance measure [29]. The Hamming distance Hm value of N strings with a defined length L is defined as:

$$Hm = \sum_{j=1}^{j=N-1} \sum_{j'=j+1}^N \left(\sum_{i=1}^{i=L} |Y_{ij} - Y_{ij'}| \right) \quad (12)$$

The more high-priced Hm is the more distinct population. o_{out} is the production group wherever the best reached solution will be placed.

Mutation Rule: once the two membranes have been merged, mutation operator as illustrated in Eq. 13) will be performed.

$$[]_l \rightarrow []_n []_k \quad (13)$$

Separation Rule: are included in P systems with effective membranes. For each separation rule, a different subset U of objects is defined to activate the layer to separate, and U objects are inserted into the two new membranes.

Candidate solution distribution: randomly group of individuals N of the whole initialized solutions in the population into $m - 1$ simple layers.

3.3 Hybrid krill herd algorithm

The general structure of the introduced hybrid krill herd with swap mutation algorithm (MHKHA) is shown in Fig. 2.

3.3.1 Mathematical notations

The introduced krill herd algorithm performed in three chief actions (see Algorithm 2) to renew the position of each individual: (1) the computation process (i.e., KH change calculation), (2) the differential evaluations (i.e., crossover operator and mutation operator), and (3) swap mutation strategy. The motion estimation steps are as below:

- Migration influenced by the behavior of other krills: the movement of individual krill is affected by the individual next-door-neighbors in the visible field.
- Foraging activity: each krill individual search for an eating root.
- Stochastic diffusion: exclusive movement of each individual according to the frequency zones.

Algorithm 2 Hybrid krill herd and swap mutation algorithm

- 1: **Input:** A set of features with (t) dimension.
- 2: **Output:** A new subset of features with (m) dimension.
- 3: **Parameter Setting:** Setting the KHA parameters as ($KHM, N_{max}, D_{max}, K, etc.$).
- 4: **Population:** Generate the KH memory.
- 5: **Repeat:**
- 6: **Motion calculations:**
 - $N_i \leftarrow$ Motion induced by the other krill individuals by Eq. (17).
 - $F_i \leftarrow$ Foraging motion by Eq. (18).
 - $D_i \leftarrow$ Physical diffusion by Eq. (19).
- 7: **Fitness function:** Evaluate the produced solutions.
- 8: **Genetic operators:** Applying the genetic mechanisms.
 - Crossover operator**
 - Mutation operator**
- 9: **Fitness function:** Evaluation the new obtained solutions.
- 10: **Swap mutation:** Applying the swap mutation strategy.
- 11: **Fitness function:** Evaluation the new obtained solutions.
- 12: **Replacement:** Replace the worst solution by the best solution.
- 13: **Termination criteria:** Stop the algorithm running when reached the max number of iterations (I_{max}).

The updating equation of the i_{th} value is as follows:

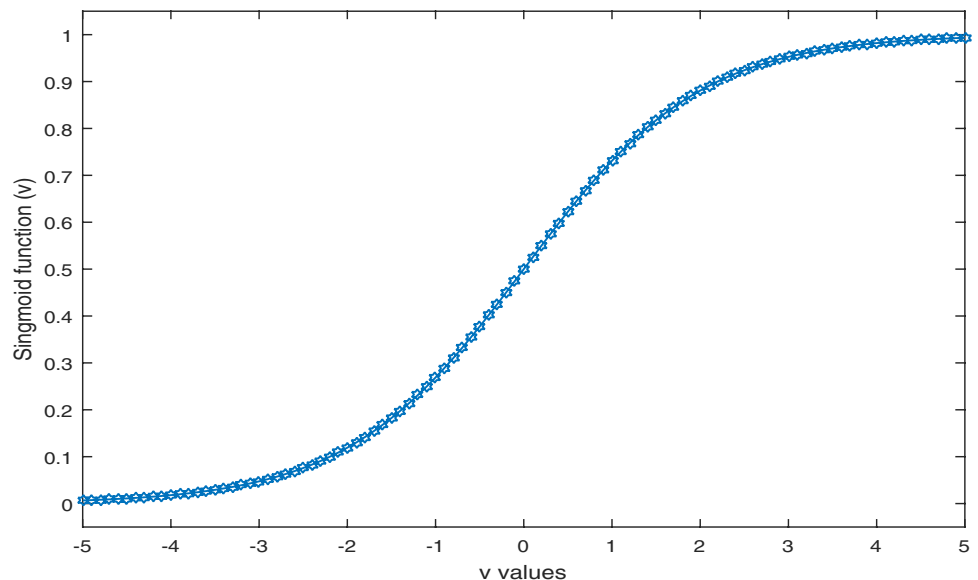
$$\frac{dx_i}{dt} = N_i + F_i + D_i, \quad (14)$$

where, N_i refer to the migration influenced by the behavior of other individuals (More details in Sect. 3.3.2). F_i refer to the food root for the i_{th} krill (More details in Sect. 3.3.3). D_i is

the net movement for the i_{th} krill (More details in Sect. 3.3.4) [30].

The proposed hybrid krill herd and swap mutation algorithm deal with binary optimization problem (i.e., feature selection problem) [31]. Hence, this algorithm is adjusted to renew the solution's points by binary value for each feature. Equation (15) demonstrates the Sigmoid equation that

Fig. 3 Illustration of the Sigmoid function values



is utilized for defining the likelihood of the i_{th} value, and Eq. (16) is utilized to change the current status value. The Sigmoid equation rates of the changing positions values are shown in Fig. 3.

$$x_{i,j} = \begin{cases} 1 & \text{if } randvalue < \frac{1}{1+exp^{-v_{ij}}} \\ 0 & \text{Else} \end{cases} \quad (15)$$

where, *rand* value is a stochastic method to generate a value which taken randomly among [0, 1], $x_{i,j}$ denotes to the value of the j_{th} value in solution number j , and $-v_{ij}$ presents the speed of solution number i at the value j , $j = 1, 2, 3, \dots, t - 1, t$.

$$x_{i,j} = \begin{cases} 1, & randvalue < s_{i,j} \\ 0, & \text{Else} \end{cases} \quad (16)$$

3.3.2 Movement influenced by the behavior of other individuals

The path of individual movement of the i_{th} is described as follows:

$$N_i^{new} = N_{max}\alpha_i + \omega_n N_i^{old}, \quad (17)$$

where N_{max} is a measurable factor for attuning the changing physical location caused by last individuals. This factor is experimentally determined as in the Sect. 4.4. α_i is calculated from the local group frequency, where ω_n is the inertia weight among [0, 1], and N_i^{old} is the current migration done by the i_{th} krill [23]. For more details refer [32].

3.3.3 Foraging motion

The foraging change of krill individuals is determined by two impacts: the new food and the old food locations [32]. The foraging change of the i_{th} individual is determined by Eq. (18).

$$F_i = V_f \beta_i + \omega_f F_i^{old}, \quad (18)$$

where V_f is the parameter used to perform a tuning for the foraging rate and it is found by experiments (i.e., 0.030),

β_i is the center point of the i_{th} krill individual, ω_f is a value in range [0, 1] which represents the inertia weight of the foraging speed, and F_i^{old} is the value of the current foraging motion of the i_{th} individual. For more details refer [32].

3.3.4 Physical diffusion

Physical diffusion strategy is used to attract each krill from high to low density region [33, 34]. Physical diffusion is found by Eq. (19).

$$D_i = D_{max} \left(1 - \frac{I}{I_{max}} \right) \delta, \quad (19)$$

where D_{max} is the factor utilized to perform a tuning for the diffusion speed concluded by experiment (i.e., 0.030), and δ indicates to an array contains random numbers between [-1, 1]. For more details refer [32].

3.3.5 The genetic operators

The genetic operators (GOs) are combined into the hybrid krill herd algorithm to increase its effectiveness in solving the underlying optimization problems [32, 34].

3.3.6 Crossover

The crossover is an effective operator for finding the global best solutions. This operator walks by a likelihood value (Cr) by creating a uniformly distributed value between [0, 1]. The u_{th} component of x_i is calculated by Eq. (20).

$$x_{i,u} = \begin{cases} x_{p,u}, & \text{if } rand_{i,u} < Cr \\ x_{i,u}, & \text{else} \end{cases} \quad (20)$$

3.3.7 Mutation

The mutation is an effective operator for finding the global best solution. This operator walks by a probability value (Mu). The probability value of the mutation operator is calculated by Eq. (21).

$$x_{i,u} = \begin{cases} x_{gbest,u} + \mu(x_{p,u} - x_{q,u}), & \text{if } rand_{i,u} < Mu \\ x_{i,u}, & \text{else} \end{cases} \quad (21)$$

p and q are to random positions, μ is a number between 0 and 1.

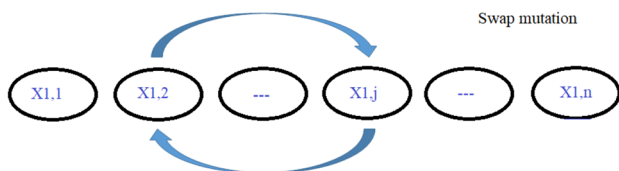


Fig. 4 Illustration of the swap mutation procedures

Table 2 Text document datasets characteristics

Datasets	Source	No. of Documents	No. of Terms	No. of Clusters
DS1	CSTR	100	1260	2
DS2	20Newsgroup	200	6518	4
DS3	Domz-Businness	250	1156	6
DS4	Domz-Computer	300	1247	8
DS5	Reuters-21578	350	3258	10
DS6	ACM	350	24457	10
DS7	WebAce	1560	8880	20

3.4 Swap mutation strategy

Swap mutation is a powerful local search strategy used to improve the solution diversity, in the proposed algorithm, by avoiding the falling in local optimum during the search process [33]. In this strategy, two random positions in the selected random solution in the search solution are selected. The job in those two random positions is swapped to each other. The swap mutation procedure is described in Fig. 4. The swap operation is repeated by probability value ($\text{rand} < 0.2$), where is the number of jobs.

3.5 Updating the krill positions:

This development is produced in the enhanced algorithm to retain the most suitable agents so far and ensure that the genetic procedures will enhance the ultimate solutions or keep it. This changing employs GOs significantly. The motion calculation steps improve each position near the best value of the objective measure using Eq. (22).

$$x_i(I+1) = x_i(I) + \Delta t \frac{dx_i}{dt}, \quad (22)$$

where,

$$\Delta t = C_t \sum_{i=1}^n (UB_i - LB_i), \quad (23)$$

Δt is a sensible factor which is fixed the first author (i.e., 0.05), n is all krill individuals in each solution, LB_i is the lower and UB_i is the upper bounds value of the i_{th} krill ($i = 1, 2, 3, \dots, n$). C_t is a constant value taken between [0, 2].

4 Experimental results and discussions

In experiments of this paper, all the systems (i.e., harmony search algorithm, genetic algorithm, particle swarm optimization algorithm, Bat algorithm, and gray wolf optimizer is utilized for determining the optimal subset of text features for solving the problem of the text feature selection) have been programmed using a Matlab software package (version 7.10.1) in Windows 7 editions on computers with 4GB RAM.

In the follows subsections, all datasets details that have been used in this paper and its organization, the test cases of the analysis criteria, and experiments results with its discussion are shown.

4.1 Text datasets

Table 2 come up with the datasets that have been utilized for the experimental purposes, which are seven common benchmark datasets usually used for applying the proposed solution in solving the features selection problem. The given datasets are applied to examine the effectiveness of the proposed algorithm (i.e., Dynamic membrane-driven hybrid krill herd algorithm with the local search strategy) with other comparative algorithms.

Benchmark datasets are, available free at (http://sites.lab.icmc.usp.br/text_collections), shown in Table 2. The first dataset called (DS1), known as CSTR, includes one hundred irregular text documents from the outlines of several technical reports that go to two topics. The second dataset called (DS2), known as twenty Newsgroups, contains two hundred random documents from exclusively collected newsgroups that goes to four topics. The third dataset called (DS3), known as Domz-business, includes two hundred and fifty random documents from an online catalog of web resources that goes to six topics. The fourth dataset called (DS4), known as Domz-computer, includes three hundred random documents from exclusively different newsgroups that goes to ten topics. The fifth dataset called (DS5), known as Reuters-21578, includes three hundred and fifty random documents from the newswire in (1987) that go to ten topics. The sixth dataset called (DS6), called ACM, includes two hundred random documents from an association for computing appliance that goes to ten topics. The last dataset (the seventh dataset) called (DS7), known as WebAce, includes one thousand and five hundred and sixty random documents from an online tool for document categorization and investigation that goes to twenty topics.

4.2 Evaluation measures

In this paper, several visible evaluation criteria, Precision measure (P), Recall measure (R), F-measure (F), and Accuracy measure (Ac), were utilized for performing comparative evaluations. These measures are approved evaluation criteria adopted to carefully test the clusters qualities, its results, and analyze the various methods to find out which the best method that can solve the given problem accurately [15]. These results of the clustering technique reflect the outcomes of the proposed feature selection algorithm (MHKHA) in selecting the informative features to presents the contents of the documents precisely.

4.2.1 Precision

The precision measurement, it is one of the statistical variability, is utilized to describe the random errors.

$$P(i,j) = \frac{n_{ij}}{n_j}, \quad (24)$$

where, n_{ij} is the number of true documents of class i in cluster j and n_j is the number of all documents of cluster j .

4.2.2 Recall

$$R(i,j) = \frac{n_{ij}}{n_i}, \quad (25)$$

where n_{ij} is the number of true documents (objects) of class i in cluster j , n_j is the number of all documents of cluster j , and n_i is the number of all documents of class i .

4.2.3 F-measure

F-measure is a common visible measurement employed especially on the text clustering area. F-measure determines

Table 4 Convergence scenarios of the basic krill herd algorithm (BKHA)

Scenario	S	V_f	D_{max}	N_{max}	The obtained best results
Standard	↓	0.020	0.002	0.010	
1	10	0.020	0.002	0.010	06
2	20	0.020	0.002	0.010	08
3	30	0.020	0.002	0.010	05
4	40	0.020	0.002	0.010	05
5	50	0.020	0.002	0.010	04
6	20	0.005	0.002	0.010	04
7	20	0.010	0.002	0.010	03
8	20	0.030	0.002	0.010	05
9	20	0.040	0.002	0.010	09
10	20	0.050	0.002	0.010	07
11	20	0.040	0.008	0.010	03
12	20	0.040	0.030	0.010	03
13	20	0.040	0.060	0.010	05
14	20	0.040	0.090	0.010	12
15	20	0.040	0.100	0.010	05
16	20	0.040	0.090	0.020	02
17	20	0.040	0.090	0.040	04
18	20	0.040	0.090	0.060	16
19	20	0.040	0.090	0.080	05
20	20	0.040	0.090	0.100	01

the percentage of the suited clusters depending on Precision and Recall measurements by Eq. (27) [15].

$$F(i,j) = \frac{2P(i,j)R(i,j)}{P(i,j) + R(i,j)}, \quad (26)$$

where $P(i,j)$ is the precision of the true documents of class i in cluster j , $R(i,j)$ is the recall of the class i in cluster j , and F-measure for all clusters is calculated by Eq. (27) according to the number of all documents n .

Table 3 Summary of experimental model

Model	Name	Discription
1	(GA)	Pure genetic algorithm to solve the feature selection problem
2	(HS)	Pure harmony search algorithm to solve the feature selection problem
3	(PSO)	Pure particle swarm optimization algorithm to solve the feature selection problem
4	(GWO)	Pure gray wolf optimizer algorithm to solve the feature selection problem
5	(KHA)	Pure krill herd algorithm to solve the feature selection problem
6	(HGA)	Hybrid genetic algorithm with swap mutation strategy to solve the feature selection problem
7	(HHS)	Hybrid harmony search algorithm with swap mutation strategy to solve the feature selection problem
8	(HPSO)	Hybrid particle swarm optimization algorithm with swap mutation strategy to solve the feature selection problem
9	(HGWO)	Hybrid gray wolf optimizer algorithm with swap mutation strategy to solve the feature selection problem
10	(MHKHA)	Dynamic membrane-driven hybrid krill herd algorithm with swap mutation algorithm

Table 5 The optimal parameters values of the krill herd algorithm

KHA	
Parameter	Value
Population	20
F min	0.3
F max	1.0
Factor(A)	0.5
Factor(<i>r</i>)	0.5
Factor(α)	0.9
Factor(γ)	0.9
# of iterations	1000

$$F = \sum_j \frac{n_j}{n} \max_i \{n(i, j)\}, \quad (27)$$

4.2.4 Accuracy evaluation

The accuracy is one of the most accurate visible measurements that is utilized to precisely measure the percentage of true selected text documents to each group by using Eq. (28) [35].

Table 6 Accuracy performance based on clusters quality

Dataset	GA	HS	PSO	GWO	KHA	HGA	HHS	HPSO	HGWO	MHKHA
DS1	0.5432	0.5547	0.5656	0.5697	0.5694	0.6258	0.6354	0.6651	0.6614	0.6687
DS2	0.3325	0.3399	0.3405	0.3357	0.3454	0.4493	0.44658	0.4587	0.4580	0.4570
DS3	0.3154	0.3245	0.3249	0.3270	0.3301	0.3954	0.3931	0.4054	0.3851	0.4102
DS4	0.2658	0.2633	0.2745	0.2597	0.2654	0.3521	0.3254	0.3452	0.3452	0.3652
DS5	0.4455	0.4493	0.4572	0.4490	0.4589	0.5325	0.5452	0.5298	0.5398	0.5441
DS6	0.3769	0.3852	0.3809	0.3781	0.3811	.4521	0.4759	0.4653	0.4655	0.4777
DS7	0.4578	0.4658	0.4701	0.4658	0.4755	0.5458	0.5421	0.5525	0.5741	0.5814

Table 7 Precision performance based on clusters quality

Dataset	GA	HS	PSO	GWO	KHA	HGA	HHS	HPSO	HGWO	MHKHA
DS1	0.5254	0.5243	0.5397	0.5314	0.5401	0.6014	0.6210	0.6148	0.6339	0.6395
DS2	0.3214	0.3365	0.3398	0.3259	0.3354	0.4259	0.4258	0.4395	0.4458	0.4484
DS3	0.3074	0.3111	0.3221	0.3218	0.3222	0.3852	0.3901	0.3975	0.3395	0.4002
DS4	0.2458	0.2548	0.2654	0.2514	0.2545	0.3336	0.3213	0.3225	0.3321	0.3252
DS5	0.4214	0.4354	0.4545	0.4298	0.4325	0.5147	0.5215	0.5201	0.5321	0.5309
DS6	0.3565	0.3695	0.3521	0.3495	0.3580	0.4444	0.4498	0.4541	0.4551	0.4451
DS7	0.4441	0.4410	0.4564	0.4426	0.4555	0.5258	0.5325	0.5415	0.5451	0.5479

Table 8 Recall performance based on clusters quality

Dataset	GA	HS	PSO	GWO	KHA	HGA	HHS	HPSO	HGWO	MHKHA
DS1	0.5362	0.5320	0.5421	0.5399	0.5495	0.6145	0.6350	0.6295	0.6385	0.6420
DS2	0.3299	0.3341	0.3425	0.3324	0.3404	0.4365	0.4358	0.4436	0.4496	0.4494
DS3	0.3154	0.3159	0.3269	0.3300	0.3278	0.3965	0.3958	0.4028	0.4121	0.4122
DS4	0.2521	0.2589	0.2664	0.2625	0.2615	0.3395	0.3321	0.3325	0.3320	0.3302
DS5	0.4258	0.4420	0.4588	0.4420	0.4395	0.5147	0.5321	0.5258	0.5412	0.5449
DS6	0.3622	0.3741	0.3612	0.3521	0.3561	0.4541	0.4601	0.4621	0.4652	0.4666
DS7	0.4541	0.4499	0.4598	0.4652	0.4599	0.5325	0.5400	0.5445	0.5525	0.5559

Table 9 F-measure performance based on clusters quality

Dataset	GA	HS	PSO	GWO	KHA	HGA	HHS	HPSO	HGWO	MHKHA
DS1	0.5314	0.5289	0.5325	0.5300	0.5395	0.6100	0.6250	0.6214	0.6245	0.6321
DS2	0.3214	0.3300	0.3389	0.3289	0.3391	0.4310	0.4235	0.4325	0.4365	0.4399
DS3	0.3112	0.3112	0.3215	0.3258	0.3186	0.3875	0.3819	0.3956	0.4005	0.4022
DS4	0.2458	0.2511	0.2589	0.2600	0.2587	0.3301	0.3291	0.3256	0.3311	0.3264
DS5	0.4210	0.4369	0.4412	0.4336	0.4317	0.5014	0.5269	0.5142	0.5321	0.5360
DS6	0.3510	0.3654	0.3514	0.3412	0.3482	0.4441	0.4499	0.4583	0.4524	0.4590
DS7	0.4451	0.4321	0.4436	0.4528	0.4415	0.5252	0.5310	0.5395	0.5451	0.5466

$$AC = \frac{1}{n} \sum_{i=1}^K n_{i,i} \quad (28)$$

where, $n_{i,j}$ denotes to all documents in the class i as well it belong to cluster i , n denotes to all documents and K denotes to all clusters that are given in the dataset.

4.3 Experiments setup

We conducted experiments analysis using ten optimization algorithms namely, genetic algorithm (GA), harmony search algorithm (HS), particle swarm optimization algorithm (PSO), gray wolf optimizer algorithm (GWO), krill herd algorithm (KHA), hybrid genetic algorithm (HGA), hybrid harmony search algorithm (HHS), hybrid particle swarm optimization algorithm (HPSO), hybrid gray wolf optimizer algorithm (HGWO), and dynamic membrane-driven hybrid krill herd algorithm (MHKHA). These algorithms are tested to investigate the most efficient algorithm to address the problem of the feature selection. Table 3 presents the properties of the proposed algorithms.

4.4 Parameter settings for the proposed algorithms

Several different experiments are executed and compared several recent published algorithms such as harmony search (HS) algorithm [9], genetic algorithm (GA) [36], particle swarm optimization (PSO) algorithm [15], and gray wolf optimizer (GWO) [37] with the proposed dynamic membrane-driven hybrid krill herd algorithm with a local search strategy (MHKHA). These matching algorithms utilized various independent (adjustable) parameters. The parameter values for the matching optimization algorithms have been possessed from specialist published articles as scholars confirmed them based on their experimental investigation. The best selected parameters value are taken from the original papers.

The experimental study is conducted to evaluate the performance of the proposed (MHKHA) with attention to the text feature selection problem. Our experiments support

finding optimal parameter values by tuning the independent parameters of the (MHKHA), and thus the optimal parameters values of the proposed (MHKHA) can be selected. Table 4 presents a set of experiments that were carried out using twenty convergence scenarios, each of which has various system parameter settings in four categories. These experiments conclude the suitability of the proposed algorithm for tackling the text feature selection problem. The given experiments investigate four various parameters, namely, KHM volume (S), foraging activity (V_f), physical diffusion value (D_{max}), and movement induced value (N_{max}). Every parameter is analyzed by holding the other parameters fixed. The highest amount of iterations (I_{max}) is 1000, which is fixed for all experiments [24, 38, 39]. Each convergence scenario was experimented and analyzed on the all used datasets (i.e., seven datasets).

Table 4 presents the four groups of interchange scenarios. A parameter in each scenario is changed to determine the optimal value. For instance, in scenario (11) to scenario (15), 3 of the main parameters are constant (i.e., S , V_f , and N_{max}) and one parameter is changed (i.e., D_{max}). The first group checks the KH memory size (S) with 5 various values (i.e., $S=10-50$), which is the number of solutions. The second group checks the foraging activity value (V_f) (i.e., $V_f=0.005$, $V_f=0.010$, $V_f=0.030$, $V_f=0.040$, and $V_f=0.050$). The third group checks the physical diffusion value (D_{max}). Finally, the fourth group checks the movement induced value. The initial values of these parameters have been chosen based on the most significant similar works in the literature [18, 32] and feature selection problem [25]. The optimal parameters values, (i.e., $S=20$, $V_f=0.030$, $D_{max}=0.008$ and $N_{max}=0.100$), are obtained by scenario number (18) as shown in Table 5, which appropriate for the studied problem. This scenario obtained best results 16 out of 28 as shown in Table 4.

4.5 Results and discussion

This part illustrates the outcomes achieved throughout the research experiments to prove the effectiveness of the introduced methods in terms of clustering quality according to

Table 10 The selected number of features alongside the number of iterations and the final reduction ratio on DS1

No. of iteration	GA	HS	PSO	GWO	KHA	HGA	HHS	HPSO	HGWO	HKHA	MHKHA
0	1260	1260	1260	1260	1260	1260	1260	1260	1260	1260	1260
100	1210	1200	1190	1195	1200	1150	1140	1144	1135	1130	1120
300	1142	1130	1135	1129	1132	1080	1070	1077	1068	1065	1055
500	985	990	988	978	977	933	932	920	921	923	915
700	862	865	850	855	864	845	840	847	844	843	840
900	820	820	815	817	810	807	801	805	803	802	800
1000	801	802	804	801	798	790	785	779	782	788	773
Redaction rate	37%36%	36%	36%	36%	37%	37%	38%	38%	38%	38%	39%

Table 11 The selected number of features alongside the number of iterations and the final reduction ratio on DS2

No. of iteration	GA	HS	PSO	GWO	KHA	HGA	HHS	HPSO	HGWO	HKHA	MHKHA
0	6518	6518	6518	6518	6518	6518	6518	6518	6518	6518	6518
100	6012	6120	6210	6014	6045	6057	6050	6100	6080	6075	6001
300	5765	5847	5789	5847	5888	5911	5748	5740	5650	5710	5600
500	5100	5123	5210	5185	5145	5001	5045	5068	5078	5012	4970
700	4756	4875	4754	4745	4874	4540	4510	4600	4652	4620	4550
900	4251	4125	4321	4212	4212	3987	3897	3862	3895	3856	3785
1000	3750	3845	3762	3688	3649	3524	3590	3622	3710	3625	3485
Redaction rate	42%	41%	42%	44%	44%	46%	45%	45%	43%	44%	47%

Table 12 The selected number of features alongside the number of iterations and the final reduction ratio on DS3

No. of iteration	GA	HS	PSO	GWO	KHA	HGA	HHS	HPSO	HGWO	HKHA	MHKHA
0	1156	1156	1156	1156	1156	1156	1156	1156	1156	1156	1156
100	1025	1054	1032	1045	1056	1023	1022	1010	1005	1012	1002
300	950	960	966	945	956	925	923	915	912	917	915
500	875	866	900	845	856	827	824	820	819	822	828
700	800	772	834	745	756	729	725	725	726	727	741
900	725	678	768	645	656	631	626	630	633	632	654
1000	701	671	690	640	640	630	621	624	633	625	619
Redaction rate	39%	42%	40%	45%	45%	45%	46%	46%	45%	46%	47%

Table 13 The selected number of features alongside the number of iterations and the final reduction ratio on DS4

No. of iteration	GA	HS	PSO	GWO	KHA	HGA	HHS	HPSO	HGWO	HKHA	MHKHA
0	1247	1247	1247	1247	1247	1247	1247	1247	1247	1247	1247
100	1115	1133	1142	1123	1123	1085	1075	1086	1065	1067	1055
300	983	1019	1037	999	999	923	903	925	883	887	863
500	851	905	932	875	875	761	731	764	701	707	671
700	830	850	877	865	856	750	720	745	700	699	660
900	809	795	822	855	837	739	709	726	699	691	649
1000	788	740	767	845	818	728	698	707	698	683	638
Redaction rate	40%	41%	39	32%	35%	42%	44%	43%	44%	45%	49%

the performance, its ability, dimensional term, and computational time, statistical analysis and compare the obtained outcomes with the other comparative algorithms.

4.5.1 General performance

Tables 6, 7, 8, 9 present the results of the k-means clustering technique after applying the suggested feature selection method by applying the proposed optimization algorithms,

Table 14 The selected number of features alongside the number of iterations and the final reduction ratio on DS5

No. of iteration	GA	HS	PSO	GWO	KHA	HGA	HHS	HPSO	HGWO	HKHA	MHKHA
0	3258	3258	3258	3258	3258	3258	3258	3258	3258	3258	3258
100	3056	3068	3069	3075	3085	3023	3014	3010	2998	2975	2960
300	2854	2878	2880	2892	2912	2788	2770	2762	2738	2692	2662
500	2652	2688	2691	2709	2739	2553	2526	2514	2478	2409	2364
700	2450	2498	2502	2526	2566	2318	2282	2266	2218	2126	2100
900	2248	2308	2313	2343	2393	2083	2100	2018	2000	1900	1768
1000	2046	2118	2124	2160	2220	1848	1794	1770	1698	1560	1470
Redaction rate	37%	35%	35%	34%	32%	43%	45%	46%	48%	52%	55%

Table 15 The selected number of features alongside the number of iterations and the final reduction ratio on DS6

No. of iteration	GA	HS	PSO	GWO	KHA	HGA	HHS	HPSO	HGWO	HKHA	MHKHA
0	24457	24457	24457	24457	24457	24457	24457	24457	24457	24457	24457
100	21540	22014	21561	21900	22312	20451	20436	21365	20412	20123	20050
300	18623	19571	18665	19343	20167	16445	16415	18273	16367	15789	15643
500	15706	17128	15769	16786	19001	12439	12394	15181	12322	12455	11236
700	14750	16540	15248	15800	16420	12020	12147	14750	12100	11321	10854
900	13794	15952	14727	14814	14218	11601	11900	14319	12010	11187	10750
1000	12838	15364	14206	13828	13200	11182	11653	13888	11700	11053	10500
Redaction rate	47%	37%	42%	43%	46%	54%	52%	43%	52%	55%	57%

Table 16 The selected number of features alongside the number of iterations and the final reduction ratio on DS7

No. of iteration	GA	HS	PSO	GWO	KHA	HGA	HHS	HPSO	HGWO	HKHA	MHKHA
0	8880	8880	8880	8880	8880	8880	8880	8880	8880	8880	8880
100	8550	8560	8450	8625	8570	8325	8425	8265	8365	8456	8326
300	8220	8240	8020	8370	8260	7770	7970	7650	7850	8032	7772
500	7541	7456	7250	7652	7624	7365	7456	7265	7369	7498	7256
700	6985	6852	6784	6856	6921	6523	6698	6589	6475	6587	6420
900	6429	6248	6318	6060	6218	5681	5940	5913	5581	5676	5584
1000	6230	6102	6055	5987	5896	5321	5452	5321	5142	5218	5201
Redaction rate	30%	31%	32%	33%	40%	39%	39%	41%	42%	41%	41%

as shown in Table 3. Note, the k-means clustering technique, refer to for more details [28], is applied over the completion of the feature selection algorithm to test its effectiveness. The best-obtained results are emphasized in **bold** size. All the obtained results from the k-mean technique were tested in terms of precision, recall, F-measure, and accuracy evaluation measures. These measures are the standard evaluation criteria utilized in the text mining domain to validate the clustering results by finding the scoring rate of the correct clustering results. This procedure is to evaluate the distinct subset of features taken from the proposed feature selection algorithms [22, 24].

In Table 6, the proposed MHKHA has got the best-recorded results on five out of seven datasets according to the accuracy measure. This method is a suitable meta-heuristic algorithm used to enhance the text feature selection

problem. The second algorithm is HGWO, which got the best-recorded results on one out of seven datasets according to the accuracy measure. As well as, the third algorithm, HPSO, is obtained the best-recorded results on one out of seven datasets according to the accuracy measure.

In Table 7, the proposed MHKHA has got the best-recorded results on four out of seven datasets according to the precision measure. This method is a suitable meta-heuristic algorithm used to enhance the text feature selection problem. The second algorithm is HGWO, which got the best-recorded results on two out of seven datasets according to the precision measure. As well as, the third algorithm, HGA, is reached the best-recorded conclusions on one out of seven datasets according to the precision measure. Lastly, the proposed algorithm obtained all the best results almost in solving the feature selection problem.

Fig. 5 The selected number of features alongside the number of iterations on DS1

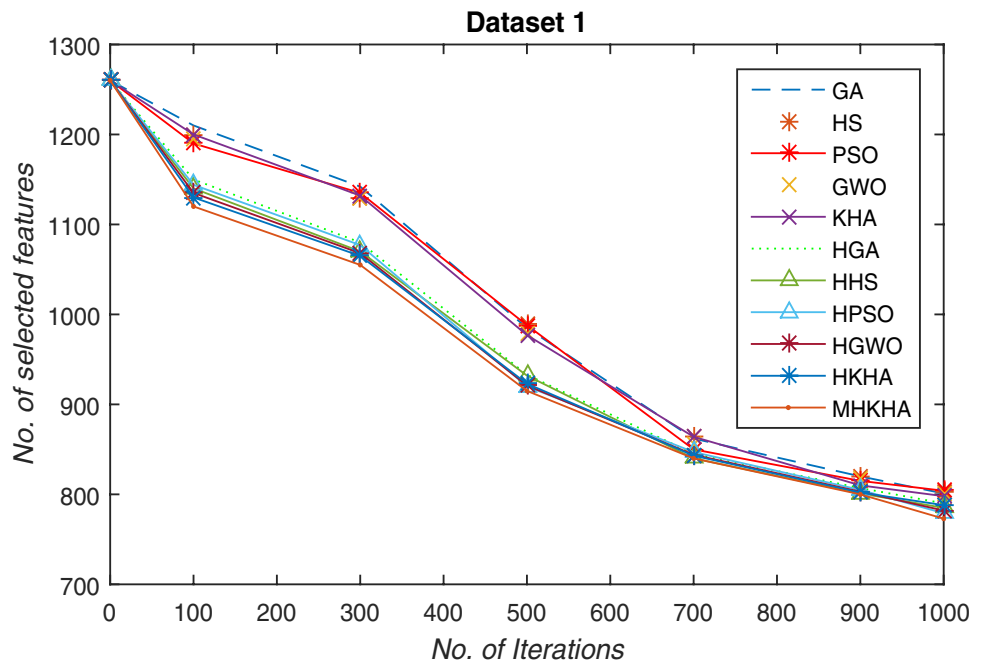
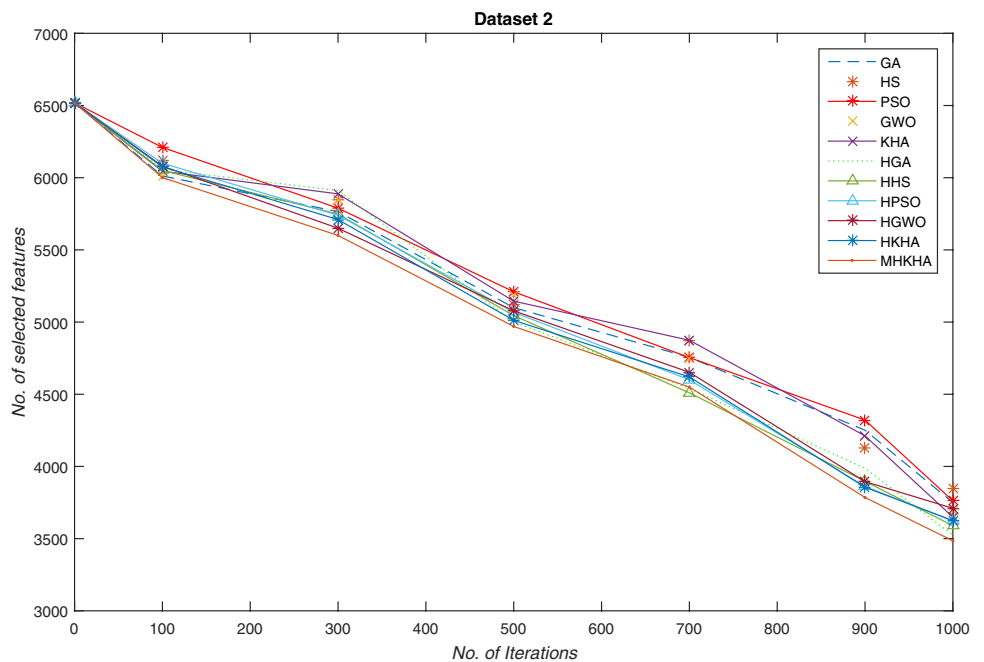


Fig. 6 The selected number of features alongside the number of iterations on DS2



In Table 8, the proposed MHKHA has got the best-recorded results on five out of seven datasets according to the recall measure. This method is a suitable meta-heuristic algorithm used to enhance the text feature selection problem. The second algorithm is HGWO, which got the best-recorded results on one out of seven datasets according to the recall measure. As well as, the third algorithm, HPSO, is got the best-recorded results on one out of seven datasets according to the recall measure.

In Table 9, the proposed MHKHA has got the best-recorded results on six out of seven datasets according to the F-measure. This method is a suitable meta-heuristic algorithm used to enhance the text feature selection problem. The second algorithm is HGWO, which got the best-recorded results on one out of seven datasets according to the F-measure. Lastly, the proposed algorithm achieved almost all the best results with the dynamic membrane-driven and

Fig. 7 The selected number of features alongside the number of iterations on DS3

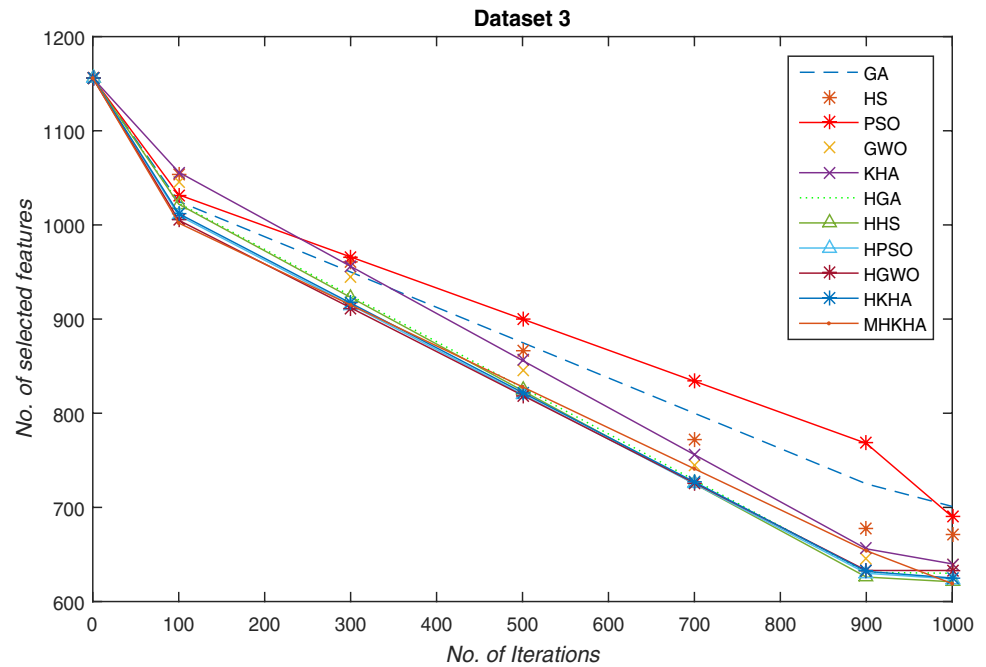
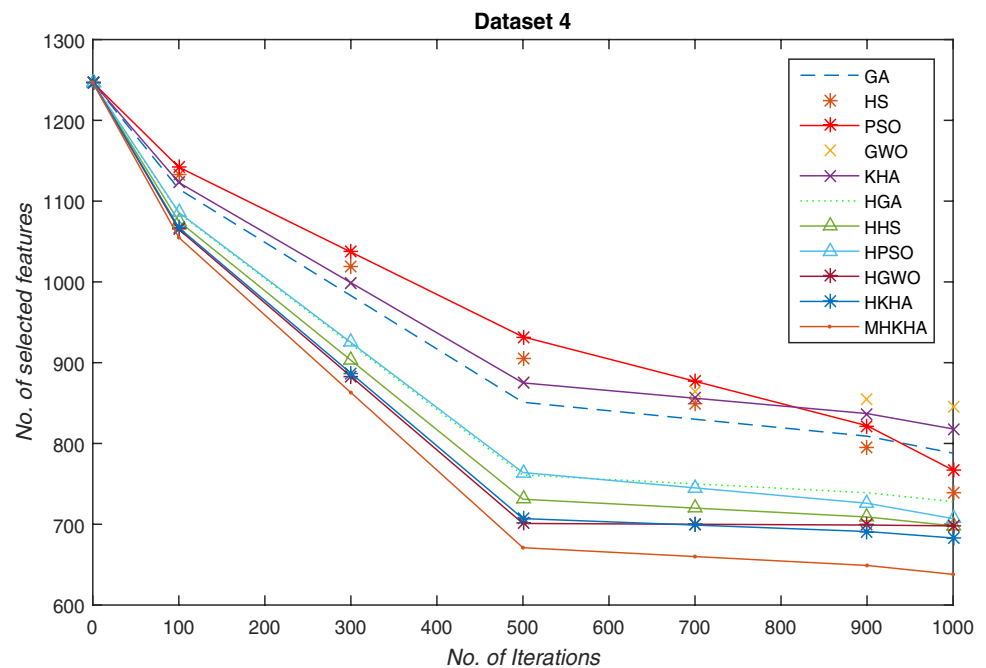


Fig. 8 The selected number of features alongside the number of iterations on DS4



hybrid krill herd algorithm with the swap mutation strategy for solving the feature selection problem.

MHKHA includes the dynamic membrane-driven and hybrid krill herd algorithm with the swap mutation strategy. This proposed method is the best method over the comparatives that improved the text clustering technique by a new robust subset of informative text features. In this method, MHKHA, the hybrid krill herd algorithm with the swap

mutation strategy under the framework of dynamic membrane computing, is applied. These additional elements (the dynamic membrane-driven and hybrid krill herd algorithm) in the proposed method give these features. Firstly, dynamic membrane-driven helps in maintaining the variety of population and adjusting the exploration and exploitation, in other words, that is a balance achieved between exploration and exploitation desirable but incompatible features. Second, the

Fig. 9 The selected number of features alongside the number of iterations on DS5

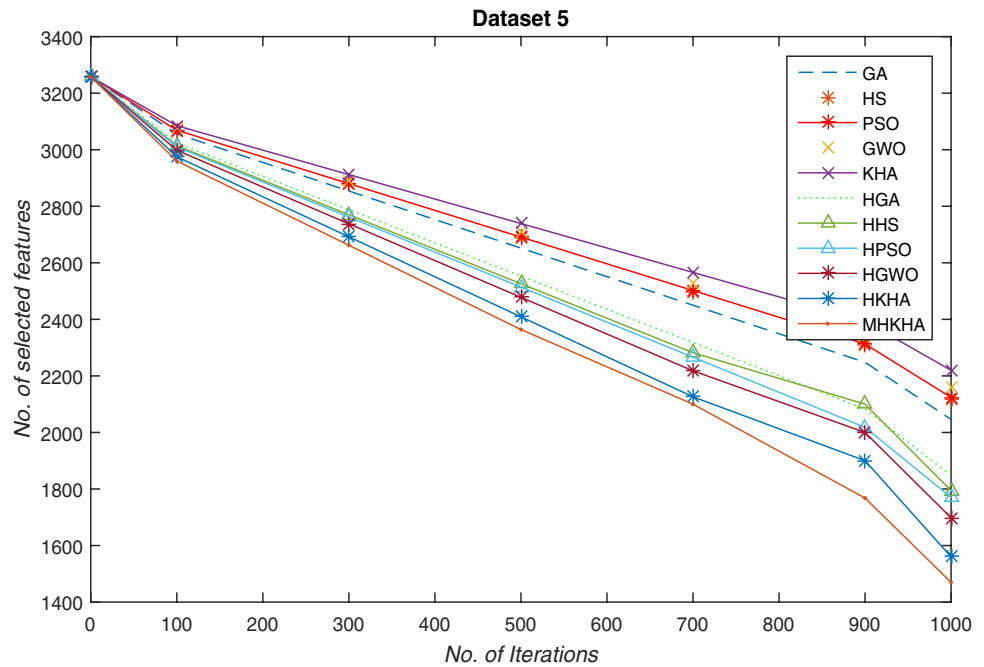
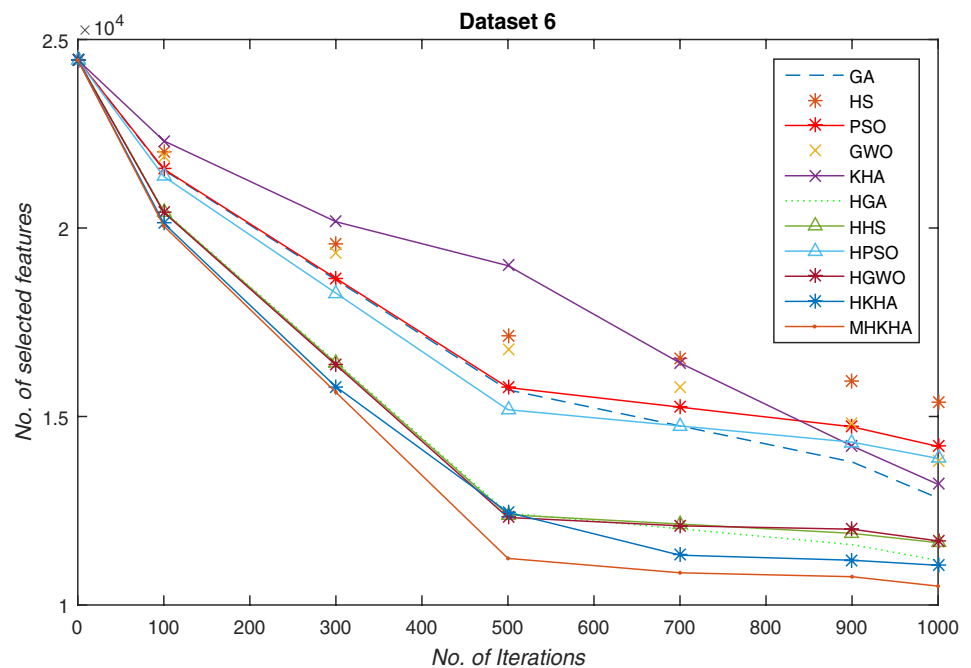


Fig. 10 The selected number of features alongside the number of iterations on DS6



KHA is simple to implement, the algorithm easily transacts with such a high-dimensional problem, and a few parameters are fixed. As well as, the local search method (swap mutation) is used to avoid the stuck in local search of the KHA. We conclude that by using the proposed method, the robust text clustering is reached by improving the performance and reduce its computation time, which means that the feature selection method is employed effectively according to its performance.

Tables 10, 11, 12, 13, 14, 15, 16 show the feature reduction ratio achieved by the feature selection technique after finishing the feature selection algorithms. The best feature selection algorithm is the proposed one (i.e., MHKHA), it got the smaller subsets of informative features six out of seven datasets. K-mean technique obtained the best results where the selected elements from 1260 to 773 in DS1 by the proposed MHKHA, which showed a decrease of 39%. The number of selected elements of DS2 reduced from 6518 to

Fig. 11 The selected number of features alongside the number of iterations on DS7

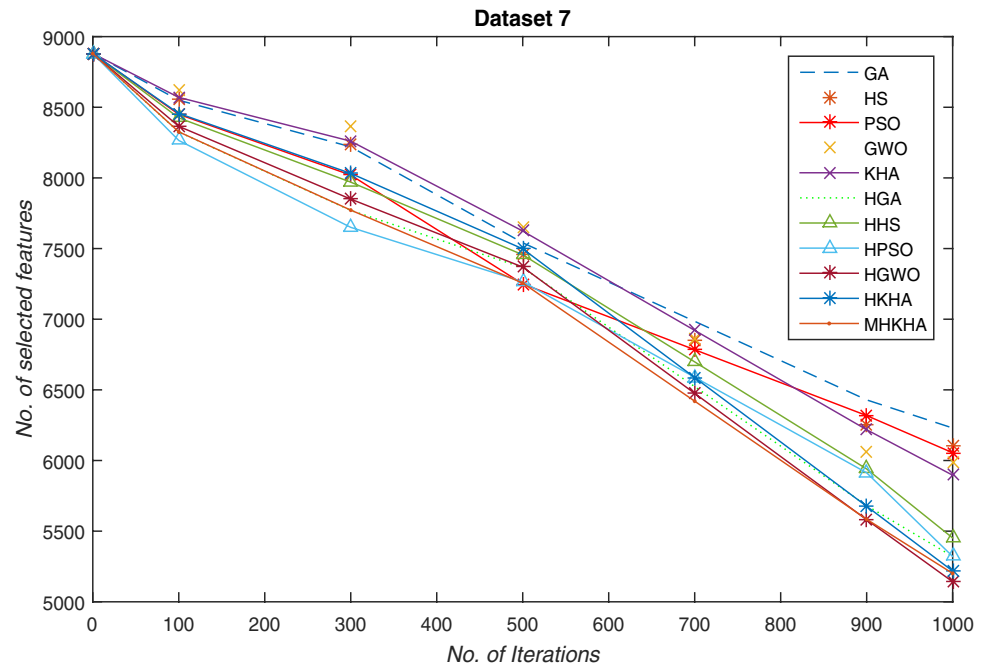


Table 17 The average ranking of the feature selection methods based on the F-measure. i.e., lower final ranking value is the best method

Method No.	Description	Dataset							Mean rank	Ranking
		1	2	3	4	5	6	7		
01	GA	08	10	09	10	10	08	07	08.85	10
02	HS	10	08	09	09	07	06	10	08.42	09
03	PSO	07	07	07	07	06	07	08	07.00	06
04	GWO	09	09	06	06	08	10	06	07.71	07
05	KHA	06	06	08	08	09	09	09	07.85	08
06	HGA	05	04	04	02	05	05	05	04.82	05
07	HHS	02	05	05	03	03	04	04	03.71	04
08	HPSO	04	03	03	05	04	02	03	03.42	03
09	HGWO	03	02	02	01	02	03	02	02.14	02
10	MHKHA	01	01	01	04	01	01	01	01.42	01

The lowest ranked algorithm is the best one

Table 18 Significance tests of the basic KHA and the proposed MHKHA using *t*-test with $\alpha < 0.05$. Highlight (bold) denote that result is significantly different

Dataset	KHA	MHKHA	<i>p</i> -value
DS1	0.5395	0.6321	4.856e-4
DS2	0.3391	0.4399	3.565e-3
DS3	0.3186	0.4022	3.565e-3
DS4	0.2587	0.3264	4.598e-3
DS5	0.4317	0.5360	3.496e-3
DS6	0.3482	0.4590	4.905e-4
DS7	0.4415	0.5466	3.223e-3

3485 by the proposed MHKHA, with a reduction of 47%. The number of selected features of DS3 reduced from 1156 to 619 by the proposed MHKHA, with a decrease of 47%. The number of selected elements of DS4 reduced from 1247 to 638 by the proposed MHKHA, with a reduction of 49%. The number of selected features of DS5 reduced from 3258 to 1470 by the proposed MHKHA, with a reduction of 55%. The number of selected elements of DS6 reduced from 24457 to 10500 by the proposed MHKHA, with a decrease of 57%. Finally, the number of selected features of DS7 reduced from 8880 to 5142 by HGWO, with a reduction

Table 19 Key to the comparator methods

No.	Method	Key	Year	Reference
01	Parallel Membrane Framework and Hybrid Krill Herd Algorithm	M-KHA	–	Our proposed
02	K-mean Text Document Clustering	TV-DF+K-mean	2015	[40]
03	An Improved Bee Colony Optimization Algorithm	IBCOCLUST	2015	[39]
04	Chaotic Gradient Artificial Bee Colony	ABC	2016	[41]
05	K-mean Text Document Clustering	BPSO+K-mean	2016	[25]
06	Hierarchical Clustering Algorithm	Hier	2016	[42]
07	Dimension reduction and feature section for K-mean	(LFW)PSO-DDR	2017	[28]
08	β -hill climbing	β HC	2017	[43]
09	K-mean clustering with PSO feature selection	H-FSPSOTC	2017	[44]
10	Link based particle swarm optimization	LBPSO	2017	[45]
11	A combination method of GSA and K-harmonic means	GSA-KHM	2017	[46]
12	Feature Selection And KH algorithm for TC	HKHA	2019	[20]

Table 20 Description of text document datasets that used by the comparative methods

Dataset	Source	# of documents	# of clusters	Reference
C-DS01	Reuters-21,578	1339	08	[25, 40, 41]
C-DS02	Classic4	2000	04	[25, 40, 41]
C-DS03	WebKB	2803	04	[25, 40, 41]
C-DS04	Reuters-21587	1000	05	[46]
C-DS05	TREC	414	09	[3, 20, 28, 43, 45]
C-DS06	TREC	313	08	[20, 28]
C-DS07	TREC	204	06	[3, 20, 28, 43]
C-DS08	20Newsgroups	500	20	[46]
C-DS09	Reuters-21587	180	06	[42]
C-DS10	TREC	873	08	[39]
C-DS11	DMOZ	697	14	[39]
C-DS12	20Newsgroups	9249	10	[39]
C-DS13	WebAce	1560	20	[39]
C-DS14	Technical reports	299	4	[3, 20, 28, 43]
C-DS15	Web pages	334	4	[3, 20, 28, 43]
C-DS16	TREC	878	10	[20, 28]
C-DS17	MEDLINE	913	10	[20, 28]
C-DS18	20newsgroup	18828	20	[28]

of 42%. While, the number of selected elements of DS7 reduced from 8880 to 5201 by the proposed MHKHA, with a decrease of 41%.

Figures 5, 6, 7, 8, 9, 10, 11 illustrate the number of selected features alongside with the number of algorithm iterations (1-1000 iterations) in each used dataset using the proposed method (i.e., MHKHA called MHKHA in the Figs. 5, 6, 7, 8, 9, 10, 11) and the other comparative ways (i.e., HS, PSO, GWO, KHA, HGA, HHS, HPSO, HGWO, and HKHA). We conclude that the MHKHA is more efficient to solve the text feature selection algorithm regarding the number of selected informative features. The number of selected informative features using MHKHA for all datasets is lower according to the number of elements chosen, as

shown in Figs. 5, 6, 7, 8, 9, 10, 11 and the chosen features are more informative based on the performance of the clustering results. The proposed MHKHA is the most powerful algorithm in decreasing the number of features. Thus, the subset of the MHKHA algorithm is chosen as a better method to apply in the clustering technique.

4.5.2 Statistical analysis

The statistical evaluation is conducted using the most standard evaluation measure in this domain; F-measure. This test is the primary evaluation criteria utilized to evaluate cluster accuracy to validate the introduced feature selection approach. The average rankings of the ten methods are being

Table 21 A comparison of the results obtained by MHKHA and best-published results

Dataset	Method	Precision	Recall	F-measure	Purity	Entropy
C-DS01	01	0.4510	0.3865	0.4910	–	–
	02	0.5392	0.3224	0.4022	–	–
	03	0.5841	0.3760	0.4550	–	–
	04	0.6752	0.3790	0.4855	–	–
C-DS02	01	0.7565	0.7251	0.7391	–	–
	02	0.8881	0.8254	0.8080	–	–
	03	0.6161	–	0.6716	–	–
	04	0.8798	0.8065	0.8416	–	–
C-DS03	01	0.4453	0.4150	0.4256	–	–
	02	0.4562	0.3503	0.3948	–	–
	03	0.3926	0.3668	0.3820	–	–
	04	0.4726	0.3340	0.3914	–	–
C-DS04	01	–	–	0.3548	–	0.4400
	18	–	–	0.3400	–	0.4600
C-DS05	01	0.5009	0.4871	0.4952	0.8665	0.4359
	07	0.4824	0.4649	0.4727	0.8406	0.4419
	08	0.3014	0.2951	0.3054	0.4510	1.2540
	09	0.3515	0.3321	0.3415	0.5410	0.9541
	10	–	–	–	0.7319	–
	12	0.4587	0.4418	0.4493	0.6612	0.4565
C-DS06	01	0.5221	0.6682	0.5252	0.7934	0.5097
	07	0.5176	0.5193	0.5143	0.7864	0.5277
	12	0.4614	0.6648	0.4626	0.7086	0.5831
C-DS07	01	0.4011	0.3941	0.4075	0.6751	0.5544
	07	0.3617	0.3587	0.4059	0.6601	0.4616
	08	0.2741	0.2961	0.2845	0.3451	1.1410
	09	0.3369	0.3514	0.3499	0.4951	0.7562
	12	0.3867	0.3950	0.3890	0.6627	0.4145
C-DS08	01	–	–	0.5325	–	0.3621
	11	–	–	0.5100	–	0.3400
C-DS09	01	–	–	0.5146	–	–
	05	–	–	0.2800	–	–
C-DS10	01	–	–	0.7514	–	–
	06	–	–	0.8826	–	–
C-DS11	01	–	–	0.8651	–	–
	06	–	–	0.8574	–	–
C-DS12	01	–	–	0.8056	–	–
	06	–	–	0.7902	–	–
C-DS13	01	–	–	0.7741	–	–
	06	–	–	0.7342	–	–
C-DS14	1	0.6141	0.5911	0.6055	0.8051	0.4154
	07	0.5127	0.5203	0.5648	0.6418	0.4088
	08	0.2541	0.3047	0.2854	0.4125	0.8954
	09	0.3515	0.3547	0.3524	0.5146	0.7841
	12	0.5166	0.5073	0.5070	0.7530	0.3209
C-DS15	1	0.8745	0.8468	0.8541	0.9954	0.2142
	07	0.7512	0.7721	0.7611	0.9902	0.3913
	08	0.5314	0.5433	0.5454	0.7051	0.6541
	09	0.5800	0.5984	0.5840	0.8100	0.4650
	12	0.4483	0.6405	0.6438	0.7981	0.3468

Table 21 (continued)

Dataset	Method	Precision	Recall	F-measure	Purity	Entropy
C-DS16	01	0.5001	0.5410	0.5241	0.7451	0.3335
	07	0.4883	0.5254	0.5054	0.6783	0.3313
	12	0.4255	0.4488	0.4362	0.7291	0.3793
C-DS17	01	0.5295	0.5258	0.5278	0.6745	0.5849
	07	0.4991	0.4837	0.4830	0.6386	0.6662
	12	0.4183	0.4196	0.4186	0.5832	0.6252
C-DS18	01	0.4512	0.4494	0.4609	0.5412	0.5665
	07	0.3956	0.4099	0.4011	0.4958	0.5983
Best Results	1	8/11	09/11	16/18	08/08	06/10

The best results are given in bold font

The highest results value is the best except the Entropy measure^{*}, the lowest results values is the best

used to enhance the k-means clustering technique reviewed in Table 17. The introduced (MHKHA) algorithm ranked the most powerful among all comparative methods, followed by HGWO, HPSO, HHS, HGA, PSO, GWO, KHA, HS, and GA in all seven datasets.

The proposed MHKHA is the best method to generate a new subset of robust, informative features to increase the effectiveness of the text k-mean clustering technique at the finish. The second (HGWO), third (HPSO), fourth (HHS), and fifth (HGA) methods are proven the effectiveness of the hybrid algorithms with the local search strategy. We acknowledge that the proposed swap mutation strategy with the used optimization algorithms achieved the optimal outcomes overall comparative methods.

The effectiveness of the introduced method (MHKHA) is further assessed utilizing the *t*-test on the seven datasets for twenty series, as displayed in Table 18. The significant differences between the obtained results are assessed utilizing the *t*-test with $\alpha < 0.05$. The *t*-test outcomes among the basic KHA and the introduced KHA (MHKHA) are reviewed in Table 18. The results showed that there are seven out of seven, in all cases, significant enhancements in performance between the basic version of the KHA and the enhanced version of the KH algorithm. According to this test, no evidence has no significant improvements. Generally, the empirical results proved that the introduced algorithm (MHKHA) achieved more approving outcomes in comparison with the other comparative algorithm significantly.

4.6 Comparison with previous works

Table 19 records the results of the introduced feature selection method (MMKHA) were analyzed and compared with most of the similar well-knowing researches utilizing their benchmark datasets. This table comprises twelve clustering methods after applying its proposed feature selection methods using eighteen datasets. Table 20 displays information

about the benchmark datasets that have been employed in this measurement to examine with the other previously published techniques.

Table 21 presents the obtained outcomes of the MMKHA matched with other recent selected methods based on various measures employed in the area of text mining. This table is classified into eighteen comparisons according to their benchmark text datasets. Each row in the table introduces the research outcomes along with its values of the evaluation tests. This table covers the results of the proposed method (i.e., its number is one as shown in Table reftbl:comrator1) and the other comparative purposes (i.e., its numbers are from 2 to 12 as shown in Table reftbl:comrator1). The symbol “-” in the table represents that the current method beside this symbol did not achieve any result according to the selected evaluation measure. The best-obtained results using all evaluation measures are the highest (most important); however, the essential result using the entropy measure is the lowest value.

Table 21 presents the final results of all comparative methods based on the used evaluation measures. The introduced method, MMKHA, produced the best outcomes in eight out of eleven cases according to the precision test, in nine out of ten cases according to the recall test, in sixteen out of eighteen cases according to the F-measure analysis, in eight out of eight cases according to the purity test, and in six out of ten cases according to the entropy test. Finally, it is clear that the results reached by the proposed MMKHA succeed almost overall the similar methods carried by several well-known scholars who have performed important work to achieve the best feature selection and clustering results.

4.7 Complexity analysis

In this section, the time complexity of the proposed MHKHA is analyzed. The complexity of the MHKHA is depended on the complexity of the KSA, Swap mutation (Sm), and

membrane framework (MF). Consequently, the complexity of the proposed MHKHA is presented as follows:

$$O(\text{MHKHA}) = X_s O(\text{KSA}) + O(\text{Sm}) + O(\text{MF})$$

where,

$$O(\text{KSA}) = O(t(\text{Dim} \times X + F \times X + X \log X))$$

$$O(\text{SM}) = O(t(\text{Dim} \times X + F \times X))$$

$$O(\text{MF}) = O(t \times X)$$

where, t denotes to the total number of iterations, Dim means to the number of given variables, F is the value of fitness function, and X indicates the population size. The time complexity of the comparative algorithms are divided into main classes: (1) the first class is the time complexity for the basic algorithms, which are $O(\text{GA}) = O(t(\text{Dim} \times X + F \times X + X \log X))$, $O(\text{HS}) = O(t(\text{Dim} \times X + F \times X + X \log X))$, $O(\text{PSO}) = O(t(\text{Dim} \times X + F \times X + X \log X))$, $O(\text{GWO}) = O(t(\text{Dim} \times X + F \times X + X \log X))$, and $O(\text{KHA}) = O(t(\text{Dim} \times X + F \times X + X \log X))$. (2) the second class is the time complexity for the hybrid algorithms, which are $O(\text{GA}) = X_s O(\text{GA}) + O(\text{Sm})$, $O(\text{HS}) = X_s O(\text{HS}) + O(\text{Sm})$, $O(\text{PSO}) = X_s O(\text{PSO}) + O(\text{Sm})$, $O(\text{GWO}) = X_s O(\text{GWO}) + O(\text{Sm})$, and $O(\text{KHA}) = X_s O(\text{KHA}) + O(\text{Sm})$.

5 Conclusion

In this paper, an innovative feature selection method is proposed for tackling the text feature selection problem in the application of text document clustering. The proposed method used a parallel membrane-inspired framework and hybrid KH algorithm with a swap mutation search strategy to enhance the performance of the KH algorithm (MHKHA). In which krill herd algorithm and swap mutation strategy are hybridized and incorporated within the parallel membrane computing framework. Seven standard benchmark text datasets were utilized to benchmark the proposed algorithm (MHKHA). The obtained results are compared and analyzed using ten other methods. The results confirmed the advantage of the proposed algorithm. Quantitative and qualitative outcomes also showed that the convergence velocity of the introduced algorithm is dominant, which cause in finding a subset of more informative features for the plurality of the datasets. Relatively, it is observed that the introduced MHKHA got excellent performance among the popular feature selection algorithms. Other related problems (that can be optimized) can be determined to solve after time to guarantee the ability of the introduced algorithm in different domains. Furthermore, other robust local search strategies can be combined to increase the exploitation search capability of the KH algorithm.

Compliance with ethical standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Al-Sai ZA, Abualigah LM (2017) Big data and e-government: A review. In: Information Technology (ICIT), 2017 8th International Conference on, IEEE, pp 580–587
2. Peng H, Wang C, Guan X (2010) Swarm intelligent optimization algorithm for text clustering. In: 2010 3rd International Conference on Computer Science and Information Technology, volume 5, IEEE, pp 200–203
3. Abualigah LM, Khader AT (2017) Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering. The Journal of Supercomputing 73:4773–4795
4. Janani R, Vijayarani S (2019) Text document clustering using spectral clustering algorithm with particle swarm optimization. Expert Systems with Applications 134:192–200
5. Sayed GI, Hassanien AE, Azar AT (2019) Feature selection via a novel chaotic crow search algorithm. Neural Computing and Applications 31:171–188
6. Abualigah LM, Hanandeh ES (2015) Applying genetic algorithms to information retrieval using vector space model. International Journal of Computer Science, Engineering and Applications 5:19
7. Zhang Y, Li H-G, Wang Q, Peng C (2019) A filter-based bare-bone particle swarm optimization algorithm for unsupervised feature selection. Applied Intelligence 1–10
8. Tubishat M, Abushariah MA, Idris N, Aljarah I (2019) Improved whale optimization algorithm for feature selection in arabic sentiment analysis. Applied Intelligence 49:1688–1707
9. Abualigah LM, Khader AT (2016) AI-Betar MA, Unsupervised feature selection technique based on harmony search. In: 2016 7th international conference on computer science and information technology (CSIT), IEEE
10. Hazir E, Erdinler ES, Koc KH (2018) Optimization of cnc cutting parameters using design of experiment (doe) and desirability function. Journal of forestry research 29:1423–1434
11. Luo M, Nie F, Chang X, Yang Y, Hauptmann AG, Zheng Q (2018) Adaptive unsupervised feature selection with structure regularization. IEEE transactions on neural networks and learning systems 29:944–956
12. Zhao M, Fu C, Ji L, Tang K, Zhou M (2011) Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes. Expert Systems with Applications 38:5197–5204
13. Wang C, Lin Y, Liu J (2019) Feature selection for multi-label learning with missing labels. Applied Intelligence 1–16
14. Hancer E, Xue B, Zhang M, Karaboga D, Akay B (2018) Pareto front feature selection based on artificial bee colony optimization. Information Sciences 422:462–479
15. Abualigah LM, Khader AT, Hanandeh ES (2018) A new feature selection method to improve the document clustering using

- particle swarm optimization algorithm. *Journal of Computational Science* 25:456–466
16. Amini S, Homayouni S, Safari A, Darvishsefat AA (2018) Object-based classification of hyperspectral data using random forest algorithm. *Geo-spatial Information Science* 21:127–138
 17. Abualigah L (2020) Multi-verse optimizer algorithm: a comprehensive survey of its results, variants, and applications. *Neural Computing and Applications* 1–21
 18. Bolaji AL, Al-Betar MA, Awadallah MA, Khader AT, Abualigah LM (2016) A comprehensive review: Krill herd algorithm (kh) and its applications. *Applied Soft Computing* 49:437–446
 19. Xu X, Liu Y (2017) Recent advances in intelligent robotic systems. *CAAI Transactions on Intelligence Technology* 2:141–141
 20. Abualigah LMQ (2019) Feature Selection and Enhanced Krill Herd Algorithm for Text Document Clustering, *Studies in Computational Intelligence book series*, Springer
 21. Abualigah LM, Khader AT, Al-Betar MA, Alyasseri ZAA, Alomari OA, Hanandeh ES (2017) Feature selection with β -hill climbing search for text clustering application. In: *Information and Communication Technology (PICICT), 2017 Palestinian International Conference on*, IEEE, pp 22–27
 22. Bharti KK, Singh PK (2014) A three-stage unsupervised dimension reduction method for text clustering. *Journal of Computational Science* 5:156–169
 23. Abualigah LM, Khader AT, Al-Betar MA, Awadallah MA (2016) A krill herd algorithm for efficient text documents clustering. In: *Computer Applications & Industrial Electronics (ISCAIE), 2016 IEEE Symposium on*, IEEE, pp 67–72
 24. Bharti KK, Singh P (2014) Chaotic artificial bee colony for text clustering. In: *2014 Fourth International Conference of Emerging Applications of Information Technology*, IEEE, 2014, pp 337–343
 25. Bharti KK, Singh PK (2016) Opposition chaotic fitness mutation based adaptive inertia weight bpsa for feature selection in text clustering. *Applied Soft Computing*
 26. Kushwaha N, Pant M (2018) Link based bpsa for feature selection in big data text clustering. *Future Generation Computer Systems* 82:190–199
 27. Abualigah LM, Khader AT, Hanandeh ES (2018) A novel weighting scheme applied to improve the text document clustering techniques. In: *Innovative Computing, Optimization and Its Applications*, Springer, pp 305–320
 28. Abualigah LM, Khader AT, Al-Betar MA, Alomari OA (2017) Text feature selection with a robust weight scheme and dynamic dimension reduction to text document clustering. *Expert Systems with Applications* 84:24–36
 29. Morrison RW, Jong KAD (2002) Measurement of population diversity. In: *Selected Papers from the 5th European Conference on Artificial Evolution*, Springer-Verlag, London, UK, UK, 2002, pp. 31–41. <http://dl.acm.org/citation.cfm?id=647456.727749>
 30. Chen G, Lu Z, Zhang Z (2018) Improved krill herd algorithm with novel constraint handling method for solving optimal power flow problems. *Energies* 11:76
 31. Babaoglu İ, Findik O, Ülker E (2010) A comparison of feature selection models utilizing binary particle swarm optimization and genetic algorithm in determining coronary artery disease using support vector machine. *Expert Systems with Applications* 37:3177–3183
 32. Gandomi AH, Alavi AH (2012) Krill herd: a new bio-inspired optimization algorithm. *Communications in Nonlinear Science and Numerical Simulation* 17:4831–4845
 33. Abdel-Basset M, Manogaran G, El-Shahat D, Mirjalili S (2018) A hybrid whale optimization algorithm based on local search strategy for the permutation flow shop scheduling problem. *Future Generation Computer Systems* 85:129–145
 34. Abualigah LM, Khader AT, Hanandeh ES (2019) Modified krill herd algorithm for global numerical optimization problems. In: *Advances in Nature-Inspired Computing and Applications*, Springer, pp 205–221
 35. Abualigah LM, Khader AT, Hanandeh ES (2018) A combination of objective functions and hybrid krill herd algorithm for text document clustering analysis. *Engineering Applications of Artificial Intelligence* 73:111–125
 36. Abualigah LM, Khader AT, Al-Betar MA (2016) Unsupervised feature selection technique based on genetic algorithm for improving the text clustering, in: *Computer Science and Information Technology (CSIT), 2016 7th International Conference on*, IEEE, pp 1–6
 37. Tu Q, Chen X, Liu X (2019) Multi-strategy ensemble grey wolf optimizer and its application to feature selection. *Applied Soft Computing* 76:16–30
 38. Abualigah L, Diabat A (2020) A novel hybrid antlion optimization algorithm for multi-objective task scheduling problems in cloud computing environments. *Cluster Computing* 1–19
 39. Forsati R, Keikha A, Shamsfard M (2015) An improved bee colony optimization algorithm with an application to document clustering. *Neurocomputing* 159:9–26
 40. Bharti KK, Singh PK (2015) Hybrid dimension reduction by integrating feature selection with feature extraction method for text clustering. *Expert Systems with Applications* 42:3105–3114
 41. Bharti KK, Singh PK (2016) Chaotic gradient artificial bee colony for text clustering. *Soft Computing* 20:1113–1126
 42. Rose JD (2016) An efficient association rule based hierarchical algorithm for text clustering, *Int J Adv Engg Tech/Vol. VII/Issue I/Jan.-March* 751 (2016) 753
 43. Abualigah LM, Sawaie AM, Khader AT, Rashaideh H, Al-Betar MA, Shehab M (2017a) β -hill climbing technique for the text document clustering. *New Trends in Information Technology* 60
 44. Abualigah LM, Khader AT, Al-Betar MA, Hanandeh ES (2017b) Unsupervised text feature selection technique based on particle swarm optimization algorithm for improving the text clustering. In: *Eai International Conference on Computer Science and Engineering*
 45. Kushwaha N, Pant M (2017) Link based bpsa for feature selection in big data text clustering. *Future Generation Computer Systems*
 46. Mirhosseini M (2017) A clustering approach using a combination of gravitational search algorithm and k-harmonic means and its application in text document clustering. *Turkish Journal of Electrical Engineering & Computer Sciences* 25:1251–1262

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.