



Improved sine cosine algorithm with simulated annealing and singer chaotic map for Hadith classification

Mohammad Tubishat¹ · Salinah Ja'afar² · Norisma Idris³ · Mohammed Azmi Al-Betar^{4,5} · Mohammed Alswaitti⁶ · Hazim Jarrah⁷ · Maizatul Akmar Ismail³ · Mardian Shah Omar²

Received: 16 January 2021 / Accepted: 17 August 2021

© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2021

Abstract

Feature selection (FS) represents an important task in classification. Hadith represents an example in which we can apply FS on it. Hadiths are the second major source of Islam after the Quran. Thousands of Hadiths are available in Islam, and these Hadiths are grouped into a number of classes. In the literature, there are many studies conducted for Hadiths classification. Sine Cosine Algorithm (SCA) is a new metaheuristic optimization algorithm. SCA algorithm is mainly based on exploring the search space using sine and cosine mathematical formulas to find the optimal solution. However, SCA, like other Optimization Algorithm (OA), suffers from the problem of local optima and solution diversity. In this paper, to overcome SCA problems and use it for the FS problem, two major improvements were introduced to the standard SCA algorithm. The first improvement includes the use of singer chaotic map within SCA to improve solutions diversity. The second improvement includes the use of the Simulated Annealing (SA) algorithm as a local search operator within SCA to improve its exploitation. In addition, the Gini Index (GI) is used to filter the resulted selected features to reduce the number of features to be explored by SCA. Furthermore, three new Hadith datasets were created. To evaluate the proposed Improved SCA (ISCA), the new three Hadiths datasets were used in our experiments. Furthermore, to confirm the generality of ISCA, we also applied it on 14 benchmark datasets from the UCI repository. The ISCA results were compared with the original SCA and the state-of-the-art algorithms such as Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Grasshopper Optimization Algorithm (GOA), and the most recent optimization algorithm, Harris Hawks Optimizer (HHO). The obtained results confirm the clear outperformance of ISCA in comparison with other optimization algorithms and Hadith classification baseline works. From the obtained results, it is inferred that ISCA can simultaneously improve the classification accuracy while it selects the most informative features.

Keywords Sine cosine algorithm · Simulated annealing · Hadith text classification · Chaotic maps

1 Introduction

Text classification (text categorization) means to classify a given text (document) into one of the given classes (categories) based on the text contents. The classified document will be grouped according to the class it belongs to. For example, a document that discusses sports topics will be classified as a sports documents, whereas a document that mainly discusses political subjects will be classified as a political document. Text classification is an important task for many applications such as social media, sentiment analysis, data mining, and medical applications. There are

many methods in the literature that were used for text classification such as Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Naive Bayes, or deep learning [9].

The importance of text classification comes from the need to classify different types of documents, for example, the growing volume in web contents which require to be classified for different purposes. Text classification was applied to different types of applications such as sentiment analysis. However, few works were conducted on the classification of Hadith (Prophet Mohammed (Peace and blessings of Allah be upon him (PBUH)) sayings) [52]. Most of the studies on text classification were conducted on languages such as English and Chinese with few works on

Extended author information available on the last page of the article

the Arabic language. Hadiths were written using the Arabic language. Hadiths are considered as the second main resource of Islam after the Quran. Hadith in Islam means a report of sayings, deeds, and teachings of the Prophet Mohammed (PBUH). There are thousands of written Hadiths which are organized into different classes. Automatic classification of Hadiths is considered as an important phase in data mining and Natural Language Processing (NLP) tasks [52]. However, few works were directed toward Hadiths classification. Furthermore, to the best of our knowledge, there is no study in the literature that used an optimization algorithm for FS on Hadith classification. The importance of the optimization algorithm comes from its ability to select the most informative features and discard irrelevant set. Thus, it can be used as a preprocessing step to improve the classification performance. FS has become an imperative to deal with the high number of irrelevant features in many applications. Hence, researchers consider developing FS techniques and applying them on various fields, including computer vision, pattern recognition, and classification, to name but a few. Examples of such contributions can be found in [64] and [63]. These studies and their significant contributions motivated us to apply optimization algorithm (OA) for FS on Hadith classification.

There are thousands of available Hadiths. The first step of classifying these Hadiths is to carry some preprocessing steps such as tokenization, stemming, and stop words removal. The next step is to build the term frequency-inverse document frequency (TF-IDF) matrix on the left terms. However, not all left terms are relevant. Thus, before using the machine learning classifier on the whole set of features, FS is required to select the most informative features. There are two main FS approaches, one is called the filter-based FS approach, while the other is wrapper-based FS. Examples of the filter-based FS are Gini Index (GI), Chi-square (CHI), and Information Gain (IG). However, these FS techniques are unable to make direct contact with the used machine learning classifier. On the other hand, wrapper-based FS approaches can directly contact with both the features and the classifier. The wrapper-based FS idea is based on using an OA to select a subset of features from the full set of features to train the classifier [39].

In this work, the Sine Cosine Algorithm (SCA) as a recent OA is proposed to select the most informative features in wrapper-based FS mode for Hadiths classification. SCA represents one of the recently developed meta-heuristic algorithms by Mirjalili [43]. The main idea of SCA is based on the use of sine and cosine mathematical functions to improve solutions' positions. The SCA algorithm begins with generating a number of random solutions (i.e., search agents) in the solution space. Iteratively, it will

update the positions of these candidate solutions toward or outward the destination solution (i.e., best solution) using sine and cosine functions [43]. SCA is simple in terms of concepts, derivative-free parameters, easy-to-use, and sound and complete. Therefore, SCA algorithm has been utilized for different benchmark functions and has achieved promising results in comparison with other famous optimization algorithms [43]. Nowadays, SCA is able to tackle a wide range of optimization problems due to its successes characteristics. The impressive merits of using SCA is that it has almost no parameters compared to other famous optimization algorithms [43].

As aforementioned, SCA has been adapted to solve several optimization problems. For example, M. A. [17, 18] improved SCA by using the opposition learning technique and applied it on engineering and global optimization problems. M. E. A. Elaziz, Ewees, Oliva, Duan, & Xiong [18] improved SCA by using Differential Evolution (DE) operators for FS problem. Zhao, Zou, & Chen [71] applied SCA for community detection problem. Belazzoug, Touahria, Nouioua, & Brahimi [14] improved SCA exploration by using a new equation for updating the solutions' positions for FS problem. Ramteke, Gurjar, & Deshmukh [50] applied SCA for FS problem. [57] improved SCA by using local search and heuristic crossover and applied it on traveling salesman problem. Lan, Fan, Liu, & Yang [36] improved SCA by using variable neighborhood search (VNS), which is a local search algorithm and applied it on scheduling problems. [25] improved SCA by using crossover and greedy selection mechanism for global optimization problem. [28] hybridized SCA with Ant Lion Optimizer (ALO) for FS problem. [22] improved SCA by using quasi-opposition learning strategy, random weighting agent, and adaptive mutation strategy for global optimization problem. Guo, Wang, Dai, & Xu [24] improved SCA by using optimal neighborhood and quadratic interpolation strategy for global optimization problem. Further, [26] hybridized SCA with artificial bee colony (ABC) for global optimization and image segmentation problems.

However, according to No Free Lunch (NFL) theorem [66], there is no single optimization algorithm that is superior to all other optimization algorithms in solving all types of problems. Hence, this motivates the usage of the SCA algorithm for FS on Hadith classification in this work. However, the SCA algorithm as other optimization algorithms tends to fall in local optima and has a problem in solutions diversity. This gives room for improvement over the basic version of SCA to avoid the mentioned problems. One successful way to improve the convergence behavior of an OA is utilizing chaotic map concepts.

In chaotic map theory, chaos has a unique characteristic, which includes its capability of generating numbers that

satisfy the following: regularity, ergodicity, unpredictable, and non-repetitive numbers [65, 70]. The chaotic theory has been incorporated with many metaheuristic algorithms in the literature and proved its capability to improve these optimization algorithms. Examples of algorithms, which were improved by using chaotic theory, are as follows. [70] used the chaotic maps with a bean optimization algorithm (BOA) to improve its population diversity and global search. [65] used the chaotic maps with moth-flame optimization (MFO) to improve the balance between the exploitation and exploration, enhance the MFO convergence speed, and avoid being stuck at local optima. [47] used the chaotic maps with whale optimization algorithm (WOA) to avoid the local optima and to solve random distribution problems for WOA internal parameters. [21] used the chaotic maps to improve the global optimization of monarch butterfly optimization (MBO). [44] used the chaotic maps to improve convergence speed for Gravitational Search Algorithms (GSA) and to avoid falling into local optima. Sharma, Kaur, Sharma, Sharma, & Bansal [55] improved the stochastic nature of the Spider Monkey Optimization (SMO) algorithm by employing chaotic theory. [53, 54] improved the convergence speed of the Dragonfly Algorithm (DA) by using the chaotic theory. Similarly, [34] improved the convergence speed and achieved the balance between the exploitation and exploration processes of the firefly algorithm (FA) using the chaotic theory. Also, [53, 54] improved the convergence speed of the Crow Search Algorithm (CSA) and the solutions being stuck into local optima by employing chaos theory.

As outlined, many studies in the literature employed chaotic maps and that proves its efficiency in improving the used OA. Thus, this advantage of using chaotic maps motivates the current work to utilize it with the SCA algorithm. The first contribution includes the integration of the standard SCA with a chaotic Singer map. The Singer map will be used to improve the balance between SCA exploration and exploitation, which in turn improves its solutions diversity.

The second contribution to the SCA includes the hybridization of simulated annealing (SA) [35] at the end of each chaotic SCA iteration to improve its exploitation (local search) by improving the current best solution. SA algorithm has been used by many researchers in the literature to improve other optimization algorithms and proved its ability to improve the performance of these algorithms. Examples of algorithms, which were improved by using SA algorithm, are as follows. [40] hybridized WOA with SA to improve WOA's exploitation. Azmi, Pishgoo, Norozi, Koohzadi, & Baesi [12] hybridized GA with SA to exploit the advantages of both algorithms. [58] hybridized FA with SA to improve FA's exploitation. [32] hybridized

PSO with SA to improve PSO's local search. Afshar-Nadjafi, Yazdani, & Majlesi [3] hybridized SA with tabu search (TS) to take the advantages of both algorithms. Potthuri, Shankar, & Rajesh [48] hybridized differential evolution (DE) algorithm with SA for global optimization. [51] improved the exploitation ability of ACO by using SA as a local search operator. Furthermore, [68] improved the exploitation of Coral reefs optimization (CRO) by using SA as a local operator at the end of each CRO iteration.

From our findings, there are very few studies conducted on Hadith classification [52]. Adding to that, there is no study in the literature that applied OA for FS on Hadith classification. Thus, an improved SCA (ISCA) is proposed in this study to be used for Hadith classification. Furthermore, ISCA is used in this work to improve Hadith classification accuracy and to reduce the number of selected features.

In the proposed work, the main objectives can be summarized as follows:

- The GI as a filter-based approach and ISCA algorithm is embedded to complement their advantages and solve their shortcomings and thus classify the Hadith text more accurately.
- The ISCA is proposed using Chaos theory and SA algorithm as follows:
 - a. Singer chaotic map: by incorporating singer chaotic map within SCA to improve the diversity of its solutions.
 - b. SA algorithm: the embedding of the SA algorithm as a local search operator at the end of each SCA iteration to improve its exploitation and to avoid the local optima problem.
- The proposed ISCA was tested on three Hadiths datasets. The experiments conducted showed the superiority of the ISCA compared to the five other comparative methods (i.e., SCA, PSO, GA, GOA, and HHO) and other Hadith classification baseline works.
- The collection of three different types of Hadiths datasets including D1, D2, and D3 datasets, where D2 represents a new type of dataset in Hadith research. In addition, for the D3 dataset, it represents the English translated version of Sahih Al-Bukhari in which no research in the literature has applied classification tasks for English Hadiths translated version. In addition, the application of the ISCA to the English dataset confirms the ability of the proposed algorithm to work on different languages.
- The proposed ISCA was tested on 14 benchmark datasets from the UCI repository to confirm the

generality of the ISCA. Again, the conducted experiments proved the superiority of the ISCA against other comparative algorithms including (SCA, PSO, GA, GOA, and HHO)

The rest of this paper is organized as follows: Sect. 2 includes the related works. Section 3 presents the details of the proposed ISCA algorithm. Section 4 describes the conducted experiments and provides in-depth analysis for the results obtained. Section 5 presents discussion about ISCA algorithm. Finally, the conclusion of the study is presented in Sect. 6.

2 Related works

Several studies were conducted for Arabic text classification. For example, [4] applied a first-order Markov model for the classification of the hierarchical Arabic text. Bahassine, Madani, Al-Sarem, & Kissi [13] improved Chi-square filter feature selection and used it with the SVM classifier for Arabic documents classification. [1] used linear discriminant analysis (LDA) for the classification of Arabic documents. Another examples such as in [42] they used classification for verifying the clustering results. Moreover, [41] proposed a technique for classifying Indic documents based on using k-means clustering, latent semantic analysis, and Gaussian clustering.

The following represents examples of studies that were conducted for Hadith classification. For example, [46] conducted a comparison between three machine learning classifiers for classifying Malay translated Hadith based on Sanad. The experiment was conducted on 100 Hadiths that were labeled manually. They reported that the SVM classifier outperformed other classifiers with 82% accuracy. Al Faraby, Jasin, & Kusumaningrum [8] applied SVM classifier with the kernel to classify Hadiths of Al-Bukhari book into three categories. They collected 1650 Hadiths from the Bahasa translated version of the Al-Bukhari book. In addition, they labeled each Hadith manually into one of three classes as a negative suggestion, positive suggestion, or information. They reported that the best-achieved result was 88% using the F1-score. Mohammed Naji [6] used a dataset from Sahih Al-Bukhari book containing 200 Hadiths which are grouped into eight books (classes). In the developed system, they determined the Term frequency-inverse document frequency (TF-IDF) matrix of the terms and then, the system ranks the classified Hadiths by its subject. They reported that the best achieved accuracy was 83.2%.

In a study by [7], four different types of classifiers were compared including the Rocchio algorithm, Naive Bayes (NB), SVM, and KNN. The experiment was conducted on

1500 Hadiths which collected from Sahih Al-Bukhari and categorized into eight books (classes). TF-IDF was used to calculate the frequency of the terms. They reported that the highest precision was 67.11% using the Rocchio classifier. [45] used the associative rule mining classification for classifying Hadiths into either Da'ief (rejected) Hadith or Sahih (accepted). However, no results were reported in the study. [2] used TF-IDF with the Random Forest classifier to classify Hadiths. They applied the experiment to 1650 Hadiths which were collected from the translated version of Sahih Al-Bukhari into Bahasa. They reported that the best-achieved result was 90% in terms of the F1-score. Mohammed N Al-Kabi, Wahsheh, & Alsmadi [7] compared three different classification algorithms including NB, LogiBoost, and Bagging for Hadiths classification. They applied the experiment on 227 Hadiths which were collected from the Sahih Al-Bukhari book. They reported that the precision and recall results of the NB classifier were the best with 59.9% and 60.4%, respectively. [5] compared the use of NB, Euclidean, cosine, Jaccard, inner product, and Dice for Hadith classification. They applied the experiment on a dataset that was collected from the Sahih Al-Bukhari book. They reported that the NB classifier achieved the best results with F1-score 85%. Harrag, El-Qawasmah, & Al-Salman [30] compared using three different stemming techniques with two classifiers (Artificial Neural Networks (ANN) and SVM for Hadith classification. These examined stemming techniques such as light stemming, dictionary-lookup stemming, and root stemming. They applied the experiment on a dataset that was collected from the Prophetic encyclopedia with a total of 453 Hadiths. They reported that the use of dictionary-lookup stemming with the ANN classifier achieved the best results with an F1-score value of 50%. [29] used ANN classifier with singular value decomposition (SVD) to classify Hadiths. They applied the experiment on a dataset that was collected from the Prophetic encyclopedia with a total of 453 Hadiths. They reported that the use of ANN with SVD achieved the best result with F-score is 88%.

Table 1 presents a summary of the discussed works on Hadiths classification, which highlighted the adopted method for each study, the dataset used, and the results obtained from the conducted experiments. Based on Table 1, it is observed that all the previous Hadith classification works used classifiers without using an optimization algorithm for minimizing the number of features. Therefore, an OA can be used to solve this issue and improve the classification performance.

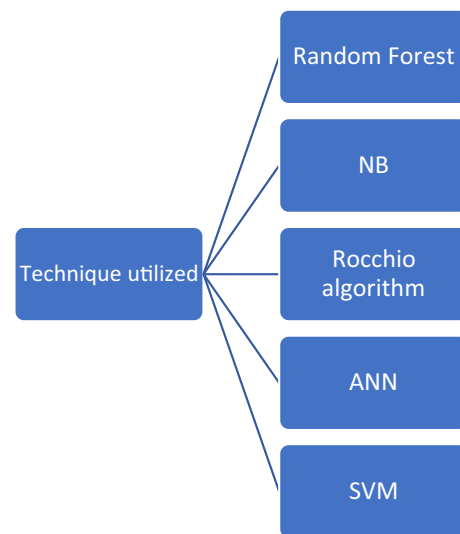
The reason for selecting an OA algorithm is to select the most informative features and ignore irrelevant ones. In the literature, several studies were conducted using different optimization algorithms for the FS problem, but none of these studies were applied on Hadith datasets for Hadith

Table 1 Comparative summary of previous works in Hadiths classification

Ref	Method used	Dataset used	Best performance achieved
[6]	TF-IDF	Sahih Al-Bukhari contains 200 Hadiths falls into eight books (classes)	Accuracy: 83.2%
[7]	TF-IDF+Rocchio algorithm	Sahih Al-Bukhari contains 1500 Hadiths falls into eight books (classes)	Precision: 67.11%
[7]	NB+removal of Sanad and removal of Arabic Diacritics from Matn	Sahih Al-Bukhari contains 227 Hadiths falls into four books (classes)	Accuracy: 60.4%
[5]	NB+TF-IDF	80 Hadiths distributed across 12 books (classes) from Sahih Al-Bukhari book	F- score: 85%
[30]	ANN+dictionary-lookup stemming	From the Prophetic encyclopedia with a total of 453 Hadiths distributed among 14 classes	F-score: 50%
[29]	ANN+SVD	From the Prophetic encyclopedia with a total of 453 Hadiths distributed over 14 classes	F-score: 88%
[46]	SVM	100 Hadiths which were labeled manually	Accuracy: 82%
[8]	SVM classifier with kernel	1650 Hadiths from Bahasa translated version of Al-Bukhari book and label each Hadith manually	F1-score: 88%
[2]	TF-IDF with Random Forest	1650 Hadiths from Bahasa Indonesia translated version of Al-Bukhari book	F1-score: 90%

classification. Examples of algorithms that were utilized for the FS problem are as follows. For example, [67] improved the GSA and applied it to the FS problem. [38] improved the PSO algorithm and applied it to the FS problem. [16] applied the Cuttlefish Algorithm (CFA) for the FS problem. Emary, Zawbaa, Grosan, & Hassenian [20] applied the GWO algorithm for the FS problem. Zawbaa, Emary, Parv, & Sharawi [69] applied the MFO for the FS problem. [23] applied the competitive swarm optimizer (CSO) for the FS problem. [10] applied the GOA for FS and parameter optimization of the SVM classifier. [39] improved the WOA algorithm using crossover and mutation operators for the FS problem. [11] applied the Butterfly Optimization Algorithm (BOA) for the FS problem. Tubishat, Abushariah, Idris, & Aljarah [59] improved the WOA and applied it to the FS problem. Tubishat, Idris, Shuib, Abushariah, & Mirjalili [61]; Tubishat, Ja'afar, et al. [62] improved Salp Swarm Algorithm (SSA) for the FS problem. [60–62] improved BOA for FS problem. [19, 56] improved HHO and used it for FS problem. Hammouri, Mafarja, Al-Betar, Awadallah, & Abu-Doush [27] improved DA for FS problem. Further, Hu, Pan, & Chu [31] improved Gray Wolf Optimizer (GWO) for FS problem.

Although several studies were conducted in the literature for the FS problem, none of these algorithms can outperform all other OA on all types of datasets based on the NFL theorem. Therefore, based on the above-mentioned reasons, we propose to improve the SCA and apply it for FS on Hadiths classification.

**Fig. 1** Hadith Classification Techniques used by previous studies

To show the association between these techniques, Fig. 1 shows the techniques used by the previous studies.

As shown in Fig. 1 and based on Table 1, these previous studies just used simple preprocessing approach without selecting the most informative features. Therefore, no FS was conducted in these studies.

3 The proposed Improved Sine Cosine Algorithm (ISCA)

The main improvements proposed for the SCA algorithm include the use of singer chaotic map and the hybridization of SA algorithm. These two improvements are used to solve the weaknesses of the original SCA algorithm. Before the details of the proposed ISCA are discussed, the descriptions of the standard SCA, SA algorithm, and chaotic map are presented.

3.1 The Sine Cosine Algorithm (SCA)

The SCA is considered as one of the most recent optimization algorithms. The SCA mainly based on using cosine and sine mathematical operators to update the search agents' positions in the search space. The SCA, as other optimization algorithms, starts with creating a number of random solutions. Afterward, these solutions will be evaluated based on the adopted fitness function, and then, the best one will be assigned as the destination solution. Moreover, the search agents' positions will be updated to new positions based on sine and cosine equations. Finally, the SCA optimization process will be stopped when the maximum number of iterations is reached. The cosine and sine operators will drive the search agents toward the optimal solution in the search space. The main equation which is used by the SCA is shown as in Eq. (1) [43]

$$X_i^{t+1} = \begin{cases} X_i^t + r_1 \times \sin(r_2) \times |r_3 P_i^t - X_i^t| & \text{if } r_4 < 0.5 \\ X_i^t + r_1 \times \cos(r_2) \times |r_3 P_i^t - X_i^t| & \text{if } r_4 \geq 0.5 \end{cases} \quad (1)$$

where r_1 , r_2 , r_3 , and r_4 represent random values. The first parameter r_1 is responsible for directing the search agents' positions of the SCA either inside or outside the space area between the current search agent X and the destination P . The second parameter r_2 is responsible for specifying the movement length either outwards from the destination or toward the destination. The third parameter r_3 is responsible for assigning the destination by random weights. The last parameter r_4 is used for switching between cosine and sine equations. Other terms such as X_i^t represent the search agent's current position. X_i^{t+1} represents the new position of the current search agent based on sine and cosine equations. P_i^t represents the current obtained best solution (destination). Figure 2 displays the pseudocode of the SCA algorithm [43].

In the standard SCA, r_1 is used for changing the search direction either to exploitation or explorations based on Eq. (2)

$$r_1 = a - t \frac{a}{T} \quad (2)$$

where a represents constant value, t is the current iteration, and T is the maximum number of iterations [43].

3.2 The Simulated Annealing (SA)

The SA algorithm was originally proposed by [35], where the main idea of the SA was inspired by the annealing theory. The SA mimics the simulating of the cooling process which exists in molten materials. The SA algorithm like other optimization algorithms needs some parameter configurations. Firstly, the SA starts with an initial temperature value and an initial solution. Then, at each iteration, the SA finds the neighbor of the current solution based on the used neighbor structure and finds the objective function value of the neighbor solution. In the next step, the difference between the neighbor solution fitness and the best solution fitness will be calculated. If the fitness difference between the current solution and neighboring solution is less than 0, then the new neighbor solution will be considered as the new best solution; otherwise, the new solution will be accepted as a current solution if the random value is less than the Boltzmann probability, wherein this case the SA accepts the current worse neighbor solution. Finally, if no condition met, the best solution is left unchanged. At the end of each iteration, the SA updates the temperature value. Figure 3 displays the pseudocode of the SA algorithm.

3.3 Chaotic map

The convergence of optimization algorithms is highly affected by the used random number generators. However, one main problem of the normal random generators is its incapability to produce random values over a particular distribution. Recently, many studies used chaos instead of normal random number generators, and it proved its ability to improve these algorithms. Chaos has the characteristics of a nonlinear system and shows chaotic behavior, which can be defined mathematically as the generation of randomness using deterministic systems. The chaos has a special property which includes its ability to generate regularity, ergodicity, unpredictable, and non-repetitive numbers [65, 70]. Therefore, it can give better results than the normal stochastic search techniques which generate their randomness based on probabilities. This important feature of chaos will control the diversity of generated solutions.

Chaos theory was embedded in many optimization algorithms to improve their performance and to solve their problems of premature convergence. One of the most important features of chaos is its ergodicity, which can improve the diversity among search agents and prevent the algorithm from being stuck in local optima. There are

Fig. 2 SCA algorithm [43]

```

1. Initialize the search agents' positions (solutions) randomly ( $X$  values)
2. Do
3.     Evaluate the fitness value for each search agent.
4.      $P =$  the best search agent based on fitness value ( $P = X^*$ )
5.     Update the values of  $r_1, r_2, r_3,$  and  $r_4$ 
6.     If  $r_4 < 0.5$ 
7.          $X_i^{t+1} = X_i^t + r_1 \times \sin(r_2) \times |r_3 P_i^t - X_i^t|$ 
8.     Else
9.          $X_i^{t+1} = X_i^t + r_1 \times \cos(r_2) \times |r_3 P_i^t - X_i^t|$ 
10.    while ( $t <$  maximum number of iterations)
11.    Return the best solution obtained as the global optimum solution

```

Fig. 3 The SA algorithm [35]

```

1. Let  $f$  be the fitness function
2. Let  $T$  be the temperature
3.  $Z =$  Current Solution
4. Let  $Z^*$  the neighbor solution
5. Best Solution =  $Z$ 
6. Best Cost =  $f(Z)$ 
7. Initialize  $T$ 
8. Current Sol =  $Z$ 
9. Current Cost =  $f(Z)$ 
10. While termination condition not met
11.     $Z^* =$  Generate new neighbor Solution of Current Solution
12.    New_Cost =  $f(Z^*)$ 
13.     $th =$  New_Cost - Current Cost
14.    If ( $th < 0$ ) then
15.        Best Solution =  $Z^*$ 
16.        Best Cost = New_Cost
17.        Current Solution =  $Z^*$ 
18.        Current Cost = New_Cost
19.    Else
20.        If ( $\text{rand}(0,1) \leq e^{-\frac{th}{T}}$ ) then
21.            Current Sol =  $Z^*$ 
22.            Current Cost = New_Cost
23.        EndIf
24.    EndIf
25.     $T = \alpha T$ 
26. End While

```

several chaotic maps functions, and one of these functions is the singer map where the equation of the chaotic singer function is as in Eq. (3) [33]. The singer map equation is as follows:

$$y_{k+1} = \mu(7.86y_k - 23.31y_k^2 + 28.75y_k^3 - 13.302875y_k^4), \quad (3)$$

where μ can be any value in the range $[0.9, 1.08]$ and $y_0 \in (0, 1)$ and $y_k \in (0, 1)$.

3.4 Gini Index (GI)

The GI measures the relevance weight of each feature toward the problem class labels [49]. Based on GI weights, these features can be ranked. Therefore, this weight can be used to select the relevant features and ignore the irrelevant ones. The equation for finding GI for a given feature is shown in Eq. (4) [15].

$$GI(f) = 1 - \sum_{j=1}^n P_j^2, \quad (4)$$

where f is the feature to find the GI for it, n is the number of class labels, and P_j represents the class probability for each feature.

3.5 Improved Sine Cosine Algorithm (ISCA)

This section presents the details of the proposed ISCA algorithm. Figure 4 presents the pseudocode of the proposed ISCA. In the pseudocode, the first improvement is indicated by lines 2 and 7 which include the use of singer chaotic map. The second improvement includes the use of the SA (line 14) to improve the SCA local search and avoid sticking at local optima. The proposed classification algorithm is composed of three main phases including preprocessing phase, filter phase, and wrapper phase.

3.5.1 Preprocessing phase

The preprocessing steps are carried out on the used datasets to remove the 'Sanad' from each Hadith where only the 'Matn' is considered for further preprocessing. It is noteworthy that Sanad in Hadith is the chain of narrators whereas Matn is the text of Hadith [52]. In addition, tokenization, stop words removal, stemming, and term weighting are conducted for the Hadith Matn. In tokenization, each Hadith Matn from the used datasets is divided into a number of individual tokens (words). In the stop words removal step, all words that are considered as stop words in English or Arabic are removed. In stemming step, the stem for each remaining token after the stop words removal will be determined. Finally, each Hadith left words will be represented as a vector where the TF-IDF matrix will be calculated from these vectors according to their terms' weights. Therefore, the input features within the TF-IDF matrix which will be used for further

processing in the experiments are Hadith words that are left from Hadith 'Matn' after applying preprocessing steps such as the stop words removal and stemming.

3.5.2 Filter phase

In this phase, the GI feature weighting will be used to find the weight of each feature. In other words, each feature will give a GI weight. This step is very important to minimize the number of features explored by the ISCA. In addition, using GI will filter the features by removing the irrelevant features and save only the relevant features to be further used by the ISCA.

3.5.3 Wrapper phase (Improved SCA based on SA and chaos)

The main improvements on the standard SCA include using singer chaotic map with the SCA. Besides that, the SA algorithm is embedded with the SCA by calling the SA at the end of each SCA's iteration to search for better solutions than the current best solution. These improvements are significant to ensure the suitable tradeoff between the exploration and exploitation of the standard SCA. As presented in Fig. 4, the first improvement to the standard SCA includes the use of a chaotic Singer map instead of a random value generated for the r_4 variable. In the improved ISCA algorithm, the value of r_4 is now received by the chaotic Singer map to improve the diversity of search agents' solutions. Moreover, the incorporation of a chaotic Singer map to generate r_4 value allows the ISCA to control the switching between sine and cosine equations which are used to update the search agents' positions. Thus, this improves solutions diversity and prevent the ISCA algorithm from being stuck at local optima. Firstly,

Fig. 4 The improved SCA algorithm

```

1 Initialize the search agents' positions (solutions) randomly ( $X$  values)
2  $SMV$  = Initialize single map vector using Eq. (3) with vector size equals to number of ISCA iterations
3 Evaluate the fitness value for each search agent.
4  $P$  = the best search agent based on fitness value ( $P=X^*$ )
5 Do
6     Update the values of  $r_1$ ,  $r_2$ , and  $r_3$ 
7      $r_4 = SMV(t)$ 
8     If  $r_4 < 0.5$ 
9          $X_i^{t+1} = X_i^t + r_1 \times \sin(r_2) \times |r_3 P_i^t - X_i^t|$ 
10    Else
11         $X_i^{t+1} = X_i^t + r_1 \times \cos(r_2) \times |r_3 P_i^t - X_i^t|$ 
12    Evaluate the fitness value for each search agent.
13     $P$  = the best search agent based on fitness value ( $P=X^*$ )
14    Call SA on  $P$  as initial solution
15    Update the best solution  $P$ 
16     $t = t + 1$ 
17 while ( $t < \text{maximum number of iterations}$ )
18 Return the best solution obtained as the global optimum solution

```


the ISCA generates a vector of values from a chaotic Singer map. Then, at each iteration of ISCA, the r_4 value from the chaotic Singer map vector will control the selection of either sine or cosine equations to update the position of the search agent.

The second improvement includes the hybridizing of the SA as a local search algorithm to improve the exploitation capability of the standard SCA. At the end of each iteration of the standard SCA, the SA algorithm will be invoked to search for a better solution than the current best solution received from SCA. The best solution generated from SCA at the end of each iteration will be considered as the initial solution to SA. The SA will search for a possible better solution than the current SCA best solution. If SA finds a better solution, it will replace the current best solution with the new one. Thus, the hybridization of SA with SCA will improve the exploitation capabilities of standard SCA and find better solutions.

The main steps of the proposed ISCA+GI are the following steps which are presented in Fig. 5:

- (1) **Preprocessing phase:** In this phase tokenization, stemming, removal of Stop words from each used Hadith, and building TF-IDF matrix will be carried out.
- (2) **Filter phase:** In this phase, to reduce the search space which will be explored by ISCA, GI will rank all features by GI weights. The top-ranked features based on GI weights will be selected as input to the ISCA. This step is crucial to avoid ISCA from searching for irrelevant space areas.
- (3) **Wrapper phase:** The usage of the ISCA to select the most informative features to learn the KNN classifier. This phase is composed of the following steps as shown in flowchart given in Fig. 5:
 - a. **ISCA initialization:** In this step, the ISCA takes input features selected from the filter phase and generates numbers of search agents' solutions randomly from these features. From these initial solutions, the ISCA finds the best solution where each solution contains a number of features selected randomly from a set of input features that are obtained from the filter phase.
 - b. **SMV initialization:** The SMV vector values will be initialized using a singer map equation as in Eq. (3). This singer map value represents the first improvement to the algorithm. The use of a chaotic singer map will improve the diversity of the solutions in the ISCA.
 - c. **Initial solutions evaluation:** In this step, the ISCA will evaluate the fitness value of each solution and set \mathbf{P} as the best solution so far.
 - d. **Search agents' positions update:** In this step, the positions of search agent at each ISCA iteration will be updated using either sine or cosine formula from Eq. (1). The selection of using either sine or cosine formula based on the value of r_4 assigned from chaotic Singer map using Eq. (3). Moreover, if the value of r_4 is less than 0.5, the sine equation will be used to update the current search agent position; otherwise, the cosine equation will be used to update its position. Finally, at the end of the current iteration, the ISCA will find the best search agent according to its fitness and update the current best solution. In this step, the improvement includes the use of r_4 value from the SMV vector to select a sine or cosine equation. Thus, the use of chaotic values will make ISCA balance between exploitation and exploration. In addition, it will improve solutions' diversity.
 - e. **Apply SA local search:** The SA takes the best solution resulted from step *d* as the initial solution, then tries to find a better solution than the current best solution by searching its neighbors. If the SA has found a better solution, then it will replace the best solution. This step represents the second improvement to the ISCA, which represents an important step to avoid ISCA from falling into local optima
 - f. **ISCA termination:** ISCA repeats steps *d* and *e* t times, and at the end of each iteration, it will update the best solution value if there is a better solution.

4 Experimental results and analysis

For the evaluation of the proposed algorithm, three Hadiths datasets were used in our experiments which were implemented using MATLAB and RapidMiner. In these experiments, tenfold cross-validation was used, where each dataset is divided into 10-folds, one is for testing and the other ninefolds are for training. This cross-validation process was repeated 10 times in each experiment. At the end of 10 runs of each experiment, the average accuracy, number of selected features, and fitness values were reported. In the conducted experiments, the number of search agents was 10, and the number of iterations was 40 for all algorithms. In addition, in all experiments, the major metric used for evaluation is the classification accuracy,

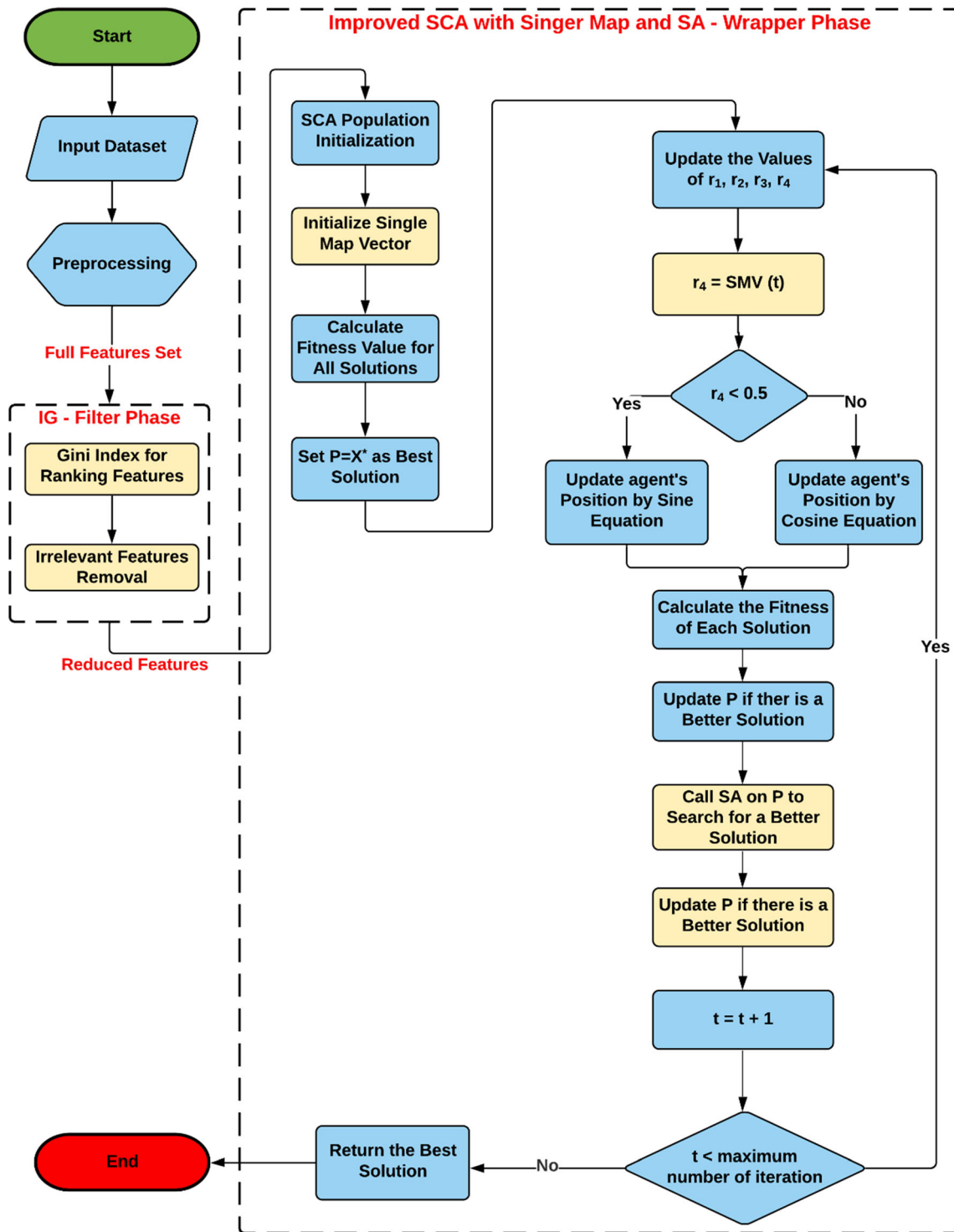


Fig. 5 The ISCA+GI algorithm

which is determined as the ratio of the correct classified Hadiths according to the actual correct classes. The parameter settings of all optimization algorithms are presented in Table 2.

4.1 Datasets

As aforementioned, three categories of datasets are used to study the effect of the proposed method on Hadith classifications. The characteristics of these datasets are provided in the following subsections, and Tables 3 summarizes the

Table 2 Parameters setting of the used optimization algorithms

Algorithm	Parameter
PSO	Acceleration constants
	$C_1 = 1.5$
	$C_2 = 2$
	Inertia Weight
	$W_1 = 1$
GA	Crossover ratio=0.9
	Mutation ratio=0.1
GOA	cMax=1
	cMin=0.00004
HHO	$B=1.5$ J is a random value over [0,2]
SCA	r_1 its value over [0, a]
	r_2 is a random value over [0, 2π]
	r_3 is a random value over [0,2]
	r_4 is a random value over [0,1]
ISCA	r_1 is a random value over [0,1]
	r_2 is a random value over [0, 2π]
	r_3 is a random value over [0,2]
	r_4 is a random value obtained from singer chaotic map. The number of SA iterations is 10

statistics of each dataset. For comparative evaluation, the UCI dataset is also used and their characteristics are given in Table 4.

4.1.1 D1 dataset

The Sahih Al-Bukhari collection of Prophet Mohammed (PBUH) Hadiths which was classified by the book author into a number of 97 books (classes) based on Hadith's subject. The term book in Sahih Al-Bukhari means the class of the given Hadith. For the D1 dataset, it is composed of 430 Hadiths that were collected from the Arabic version of Sahih Al-Bukhari book (<https://sunnah.com/bukhari#>) and distributed over 8 classes.

4.1.2 D2 dataset

The D2 dataset is a collection of 481 Hadiths collected from Mohammad Al Albani Targheeb and Tarheeb's book (<http://islampost.com>). In this book, Hadiths are classified into two classes such as encouragement (Targheeb) and warning (Tarheeb). None of the research works in the literature has applied this type of datasets, which mimic the idea of sentiment analysis with two classes, positive and negative. In this dataset, Hadiths were originally classified

based on either it is encouragement (Targheeb) Hadith or warning (Tarheeb) Hadith.

4.1.3 D3 dataset

The D3 dataset is a collection of Hadiths from the English translated version of Sahih Al-Bukhari from (<https://sunnah.com/bukhari#>), which was translated from Arabic to English. For the D3 dataset, 433 Hadiths were collected which are distributed over 8 classes. Moreover, there is no study in the literature that has previously applied classification for the English version of Hadiths, where most studies were conducted for the Arabic version.

4.2 Experiments results and evaluation

The viability and performance of the proposed method have been measured using five different experimental scenarios. These experimental scenarios are discussed in the following subsections.

4.2.1 Experiment 1

Results using KNN, Random Forest, and SVR-RBF classifiers only.

The first experiment was conducted using the KNN, Random Forest, or SVR-RBF only. Each classifier was applied to each dataset using the full feature set within each dataset without using the GI feature reduction or optimization algorithm. Table 5 shows that the best Hadiths classification accuracy achieved using KNN was 0.65 on the D1 dataset, 0.77 on the D2 dataset, and 0.75 on the D3 dataset. Moreover, the results achieved by KNN is superior to other classifiers as shown in Table 5. However, these results demonstrate the importance of applying the FS and optimization algorithm to improve the performance and reduce the number of the used features.

4.2.2 Experiment 2

Hadiths classification results using KNN classifier with GI feature selection.

In this experiment, GI feature reduction was applied together with the KNN classifier. GI was used to rank all features, and the top-ranked features from these features were selected as shown in Table 6 (The Number of input features based on GI ratio represents the number of features selected from the whole set of features shown in Table 6 based on GI ranking). In this experiment, the GI ranked features were selected based on 10%, 20%, and 30% ratios of the top features from GI weighted features. Based on the results presented in Table 6, the highest classification accuracy using the D1 dataset was 0.66, which was

Table 3 Statistics of D1, D2, and D3 datasets

D1		
English Book (Class) Name	Arabic Book (Class) Name	Number of Hadiths
Charity	الزكاة	57
Prayers	الصلاة	45
Medicine	الطب	52
Pilgrimage	الحج	54
Fasting	الصيام	58
Dress	اللباس	59
Dreams	الاحلام	53
Funerals	الجنائز	52
D2		
Hadith Class	Number of Hadiths	
Encouragement (Targheeb)	243	
Warning (Tarheeb)	238	
D3		
English Book (Class) Name	Arabic Book (Class) Name	Number of Hadiths
Prayers	الصلاة	53
Pilgrimage	الحج	59
Charity	الزكاة	56
Medicine	الطب	50
Funerals	الجنائز	52
Food	الطعام	51
Fasting	الصيام	59
Dress	اللباس	53

obtained when either top 20% or 30% of the GI weighted features were selected and used by the KNN. The accuracy of GI with KNN on D1 outperformed the KNN only without feature reduction. For the D2 dataset, the highest accuracy achieved on D2 was 0.82 which was obtained when 10% of GI weighted features were used and selected. Thus, the accuracy achieved on D2 using GI with KNN is better than using the KNN classifier only without any FS technique. Moreover, the highest accuracy achieved on the D3 dataset was 0.73, which obtained when 10% of GI ranked features were used and selected. However, the accuracy received on D3 using GI with KNN is less than using KNN only, because the GI filter FS has no direct contact with the machine learning classifier. Therefore, it is important to select an optimal feature combination from these features using an optimization algorithm to train the KNN classifier and to achieve better performance.

4.2.3 Experiment 3

ISCA comparisons with other algorithms.

In this experiment, we compared the performance of the ISCA algorithm with the performance of other famous and most recent optimization algorithms including SCA, PSO, GA, GOA, and HHO. In particular, we implemented these baseline optimization algorithms and applied them to Hadith classification and compared them with our ISCA algorithm. These SCA, PSO, GA, GOA, and HHO optimization algorithms were implemented for Hadith classification because they have not been applied to Hadith classification. Each optimization algorithm was applied to the full feature set without using GI filter FS. The results obtained from applying these algorithms are presented in Table 7.

Based on the results displayed in Table 7, it is proved that the ISCA algorithm outperforms all other baselines algorithms over all used datasets by accuracy and fitness values (bolded results). Furthermore, the achieved accuracy resulted from the application of ISCA on the full feature set of each dataset is superior to using KNN only or GI with KNN classifier according to the comparison of the results from Table 5 and Table 6. This outperformance is based on ISCA's ability to select the most informative

Table 4 Statistics of the used UCI datasets

#	Dataset	Number of features
1	Heart	13
2	Credit	20
3	Australian	14
4	Exactly	13
5	Spect	22
6	QSAR	41
7	exactly2	13
8	Vehicle	18
9	tic-tac-toe	9
10	Vote	16
11	ionosphere	34
12	Vowel	13

Table 5 Comparison of Hadiths classification accuracy using only KNN, Random Forest, and SVR-RBF classifiers

Dataset	No. of Features	Percentage of Features Used (%)	KNN	Random Forest	SVR-RBF
D1	1277	100	0.65	0.60	0.62
D2	1149	100	0.77	0.70	0.75
D3	1768	100	0.75	0.55	0.67

Table 6 Hadiths classification accuracy applying GI feature reduction with KNN classifier

Dataset	Number of input features based on GI ratio	GI ratio (%)	Average accuracy
D1	129	10	0.65
D1	256	20	0.66
D1	384	30	0.66
D2	116	10	0.82
D2	231	20	0.81
D2	345	30	0.79
D3	178	10	0.73
D3	354	20	0.70
D3	531	30	0.67

The values in bold indicates the best results

features and discard irrelevant features. In addition, it is clearly noticed from Table 7 that ISCA outperforms SCA in terms of feature reduction across all used datasets D1, D2, and D3.

As shown in Fig. 6, ISCA is outperforming all other optimization algorithms in terms of accuracy. This out-performance resulted based on the introduced improvements into the native SCA.

4.2.4 Experiment 4

Comparisons of ISCA used with GI filter feature selection to other baselines algorithms used with GI.

To evaluate the performance of the proposed ISCA hybridized with GI, each one of the tested algorithms was used together with GI FS. The results of applying all algorithms together with GI FS are presented in Tables 8, 9, 10. At first, the features from each dataset were ranked by GI weight. The top GI weighted features were then selected and used by all algorithms based on 10%, 20%, and 30% ratios of top GI weighted features.

From Table 8, it is obviously observed that the ISCA outperforms all other algorithms in terms of classification accuracy over all three datasets as indicated in bold. These results proved the superiority of ISCA in comparison with other algorithms. Furthermore, the resulted accuracy from the hybridization of optimization algorithm with GI feature selection outperforms the results of the accuracy presented in Tables 6 and 7 when using either ISCA or GI only without hybridization.

In addition, based on the results from Table 9, it is clearly observed, the superiority of ISCA over other algorithms in feature reduction as indicated in bold. It is also found that HHO is outperforming ISCA in one case. Moreover, ISCA outperforms the standard SCA in feature reduction as it selects a fewer number of features in all datasets.

Furthermore, as shown in Table 10, the ISCA algorithm outperforms all other algorithms in terms of fitness value over all three datasets. These results are achieved as the ISCA algorithm always obtained the lowest classification accuracy error in comparison with other algorithms based on the used objective function. These ISCA fitness results were obtained due to ISCA's ability to select the best features combinations from the set of features and ignore other irrelevant features.

As presented in Figs. 7, 8, 9, ISCA+GI is outperforming all other optimization algorithms in terms of accuracy. This outperformance over other optimization algorithms resulted from ISCA's ability to balance between exploitation and exploration. In addition, it has ability to avoid local optima problem.

Based on the results from the conducted experiments, it is clearly observed that applying GI reduction before applying the optimization algorithm can improve the performance by minimizing the number of features and selecting the top relevant ones. It is also found that using GI before applying ISCA can minimize the number of explored features by ISCA by avoiding ISCA to search irrelevant areas. In addition, the proposed ISCA algorithm combined with GI not only reduces the number of features but also improve the Hadiths classification accuracy.

Table 7 The results obtained in Experiment 3, including accuracy, selected features, and fitness values

	D1			D2			D3		
	Accuracy	#Feature	Fitness	Accuracy	#Feature	Fitness	Accuracy	#Feature	Fitness
SCA	0.61	702	0.39	0.80	643	0.20	0.72	1363	0.28
ISCA	0.83	610	0.17	0.96	548	0.04	0.90	909	0.10
PSO	0.75	631	0.25	0.93	563	0.07	0.78	876	0.22
GA	0.73	605	0.27	0.91	544	0.09	0.81	831	0.19
GOA	0.63	642	0.37	0.80	570	0.20	0.65	880	0.35
HHO	0.67	582	0.33	0.83	549	0.17	0.75	1038	0.25

The values in bold indicates the best results

4.2.5 Experiment 5

Comparisons of ISCA used with GI feature selection to baselines Hadith Classification works.

The presented baseline works used different types of Hadiths datasets in their experiments which make a direct comparison unreliable. In addition to that, they used different evaluation metrics. Therefore, to compare these baseline works with the proposed work, the accuracy measure was used. Furthermore, all these baseline works were implemented and evaluated on the three datasets including D1, D2, and D3 for a fair comparison.

In this experiment, we implemented and evaluated the following baseline works as it represents the best work for Hadith classification as indicated in Table 1. These baseline methods were implemented and evaluated using D1, D2, and D3 datasets. ISCA with GI was compared to Hadith classification works including (Mohammed Najji [5] which used NB+TF-IDF for Hadith classification, [29] which used ANN+SVD, [2] work which used TF-IDF with Random Forest, and [8] which used SVM classifier with kernel. The results of the comparison are presented in Table 11.

As shown in Table 11, ISCA+GI outperformed all other Hadiths classification baseline works in terms of accuracy. This ISCA+GI outperformance resulted based on the ability of ISCA+GI to select the relevant features and discard irrelevant features to train the KNN classifier. In addition, increasing its ability to escape from local optima

Table 8 Comparison of average classification accuracy between the ISCA and other algorithms over 10 runs with using GI feature reduction

Dataset	GI ratio	Algorithms average accuracy					
		SCA	ISCA	PSO	GA	GOA	HHO
D1	10	0.67	0.81	0.69	0.74	0.64	0.73
D1	20	0.70	0.87	0.78	0.78	0.67	0.71
D1	30	0.69	0.90	0.81	0.81	0.66	0.72
D2	10	0.86	0.97	0.90	0.92	0.83	0.89
D2	20	0.86	0.98	0.91	0.94	0.84	0.90
D2	30	0.84	0.98	0.93	0.94	0.84	0.85
D3	10	0.77	0.92	0.85	0.86	0.74	0.82
D3	20	0.75	0.95	0.85	0.87	0.74	0.78
D3	30	0.73	0.94	0.87	0.87	0.72	0.78

The values in bold indicates the best results

and improve population diversity. Figure 10 presents a comparison between ISCA+GI and other Hadith classification baseline works. Clearly, the ISCA+GI shows superiority over all other baseline works.

4.3 Experiment 6

ISCA comparison with other optimization algorithms using UCI datasets.

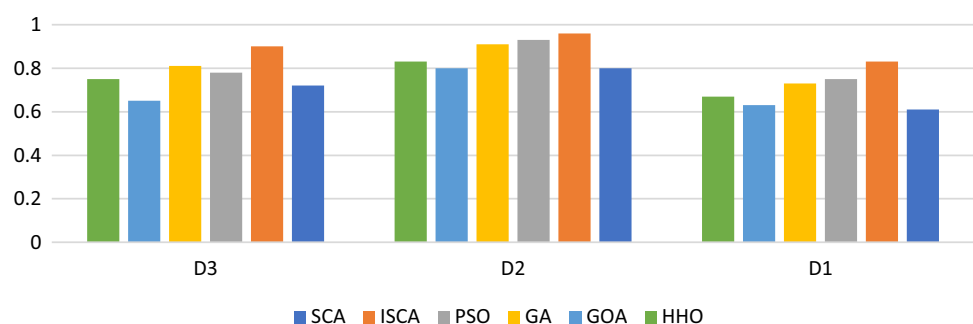
Fig. 6 ISCA accuracy compared with other optimization algorithms

Table 9 Comparison of the average number of selected features between ISCA and other algorithms over 10 runs using GI feature reduction

Dataset	Number of features using GI ratio	GI ratio	Average number of selected features					
			SCA	ISCA	PSO	GA	GOA	HHO
D1	129	10	102	52	63	54	65	60
D1	256	20	190	109	126	130	110	130
D1	384	30	246	168	189	169	191	160
D2	116	10	79	39	55	49	58	55
D2	231	20	162	87	112	101	115	107
D2	345	30	221	140	168	156	171	146
D3	178	10	127	71	88	77	90	94
D3	354	20	226	151	173	155	177	168
D3	531	30	316	232	260	242	262	242

The values in bold indicates the best results

Table 10 Comparison of average fitness between ISCA and other algorithms over 10 runs using GI feature reduction

Dataset	GI ratio	Algorithm average fitness					
		SCA	ISCA	PSO	GA	GOA	HHO
D1	10	0.33	0.19	0.31	0.26	0.36	0.27
D1	20	0.30	0.13	0.22	0.22	0.33	0.29
D1	30	0.31	0.10	0.19	0.19	0.34	0.28
D2	10	0.14	0.03	0.10	0.08	0.17	0.11
D2	20	0.14	0.02	0.09	0.06	0.16	0.10
D2	30	0.16	0.02	0.07	0.06	0.16	0.15
D3	10	0.23	0.08	0.15	0.14	0.26	0.18
D3	20	0.25	0.05	0.15	0.13	0.26	0.22
D3	30	0.27	0.06	0.13	0.13	0.28	0.22

The values in bold indicates the best results

In this experiment, to confirm the generality of the proposed ISCA algorithm, we applied ISCA on 14 different dataset types from the University of California at Irvine (UCI) repository [37].

In this experiment, the performance of the ISCA algorithm in comparison with other optimization algorithms

including (PSO, GA, GOA, SCA, and HHO) is reported. As shown in Table 12 and marked with bold fonts, the ISCA algorithm is superior to all other optimization algorithms based on the classification accuracy of the 14 UCI datasets. The outperformance of the ISCA is achieved because of its ability to improve the diversity of the solutions using the singer chaotic map and the avoidance of local optima using the SA algorithm. By evaluating the obtained accuracy results of the ISCA in Table 12, the outperformance of ISCA is confirmed with a clear significance over all used datasets. These obtained results by the ISCA confirm its stability to balance between exploration and exploitation while it is selecting the most informative features.

Now, based on the average number of selected features as shown in Table 13, the ISCA is superior to the standard SCA over all datasets. In addition, the ISCA is superior to other algorithms including PSO, GA, GOA, and HHO in 8 out of 14 datasets. The HHO outperformed other algorithms over 5 out of 14 datasets, where the outperformance of PSO, GA, and GOA occurs only over one dataset. The outperformance of the ISCA is justifiable because it always

Fig. 7 ISCA+GI accuracy comparison with other optimization algorithms over D1 dataset

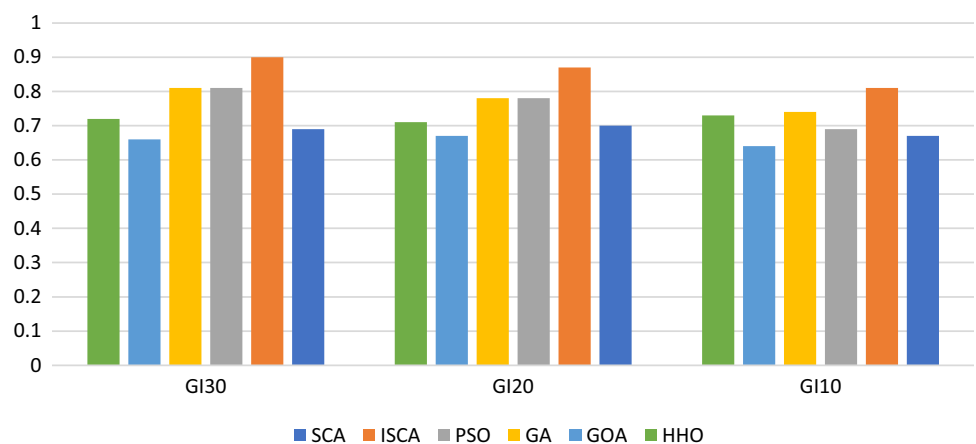


Fig. 8 ISCA+GI accuracy comparison with other optimization algorithms over D2 dataset

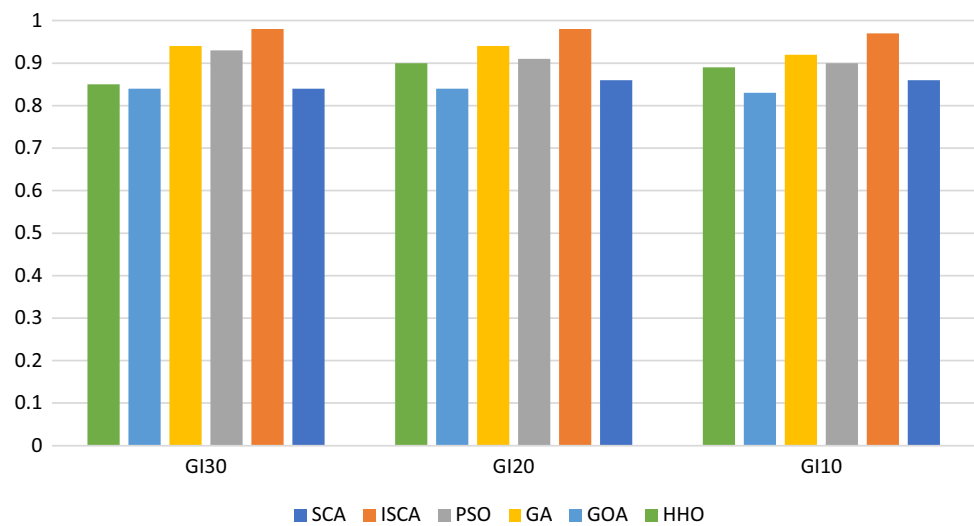
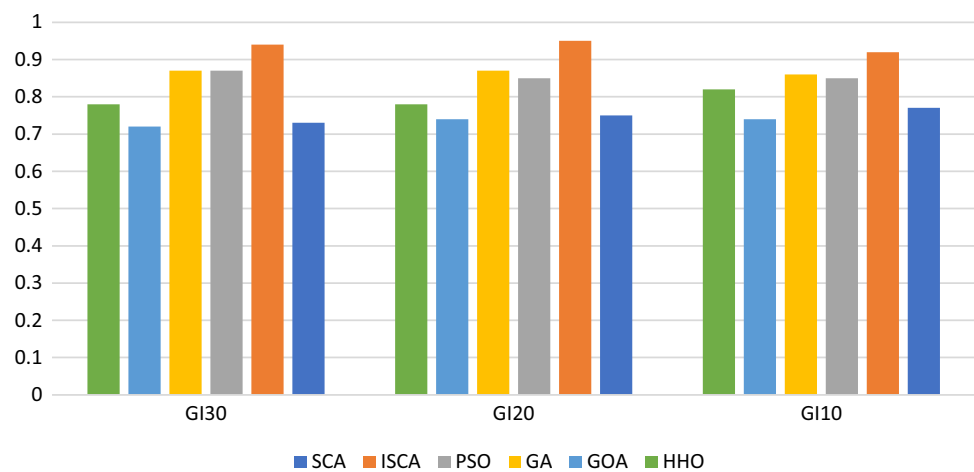


Fig. 9 ISCA+GI accuracy comparison with other optimization algorithms over D3 dataset



selects the most informative features which consequently improve the classification accuracy as shown in Table 12.

Furthermore, based on fitness results from Table 14, it is clearly observed that the ISCA is also superior to all other optimization algorithms. This is justifiable because the ISCA always targets the most informative features while it has the minimum classification error over all used datasets compared to other optimization algorithms.

To further confirm the superiority of ISCA algorithm in comparison with other optimization algorithms, paired T-test was calculated based on the accuracy results of ISCA and other optimization algorithms which is presented in Table 8, and the results of the statistical test are shown in Table 15.

As shown from Table 15, the T-test results of ISCA vs SCA, ISCA vs PSO, ISCA vs GA, ISCA vs GOA, and ISCA vs HHO rejects the null hypothesis at a significance level of 0.05. Therefore, this confirm that ISCA is

significantly different from other optimization algorithms over all Hadiths datasets.

Based on the achieved results by the ISCA in comparison with other well-known optimization algorithms and previous studies on Hadith classification, the ISCA algorithm achieved superior performance. As shown in Table 7, ISCA outperformed all other optimization algorithms overall Hadith datasets. Also, to improve the performance further, we combined GI with the ISCA, and one can notice that clearly in Table 8. Based on Tables 7 and 8, the accuracy of the ISCA and ISCA with GI is superior to other optimization algorithms as indicated in bold fonts in these tables. Moreover, as shown in Table 11, ISCA+GI outperformed other baseline works of Hadith classification over all datasets. For example, ISCA+GI outperformed [8] on D1 by 28%, D2 by 23%, and D3 by 28%. ISCA+GI outperformed (Mohammed Naji [5] on D1 by 29%, D2 by 22%, and D3 by 21%. ISCA+GI outperformed [29] on D1 by 21%, D2 by 19%, and D3 by 17%. ISCA+GI

Table 11 Comparison of accuracy between ISCA+GI and other baselines Hadith classification works

Dataset	Classifier	Accuracy
D1	ISCA+GI	0.90
D2	ISCA+GI	0.98
D3	ISCA+GI	0.95
D1	[8]	0.62
D2	[8]	0.75
D3	[8]	0.67
D1	[5]	0.61
D2	[5]	0.76
D3	[5]	0.74
D1	[29]	0.69
D2	[29]	0.79
D3	[29]	0.78
D1	[2]	0.60
D2	[2]	0.70
D3	[2]	0.55

outperformed [2] on D1 by 30%, D2 by 28%, and D3 by 40%. Furthermore, to confirm the applicability of the ISCA algorithm on other datasets with different types and dimensionalities, we applied it on 14 datasets from the UCI repository as shown in Table 12. This ISCA+GI

outperformance is imputed to a number of merits including 1) the combination of feature selection levels by using GI with ISCA, which resulted in selecting the most informative features by GI to avoid exploring irrelevant features by ISCA. Furthermore, the selection of most informative features by ISCA from the set of features that are selected by GI. These improvements which are introduced to the ISCA make it capable to work efficiently over datasets with different types and dimensionalities, where these datasets are ranging from low, mid, to high. The overall performance of the ISCA algorithm over all metrics is superior to other optimization algorithms and previous studies on Hadith classification. This outperformance is imputed to ISCA’s ability to avoid being stuck in local optima (the improvement into exploitation based on using SA algorithm to improve the current best solution) and also, ISCA’s ability to improve other solutions in the population based on using singer chaotic map. These improvements to ISCA improved its ability to explore hidden or unseen search space areas by other normal Hadith classification methods and other optimization algorithms. In other words, ISCA has a number of merits, such as the ability to select the most informative features and ignore irrelevant features, ISCA ability to improve the best solution in comparison with other algorithms by avoiding local optima

Fig. 10 ISCA+GI accuracy compared with other baselines Hadith classification

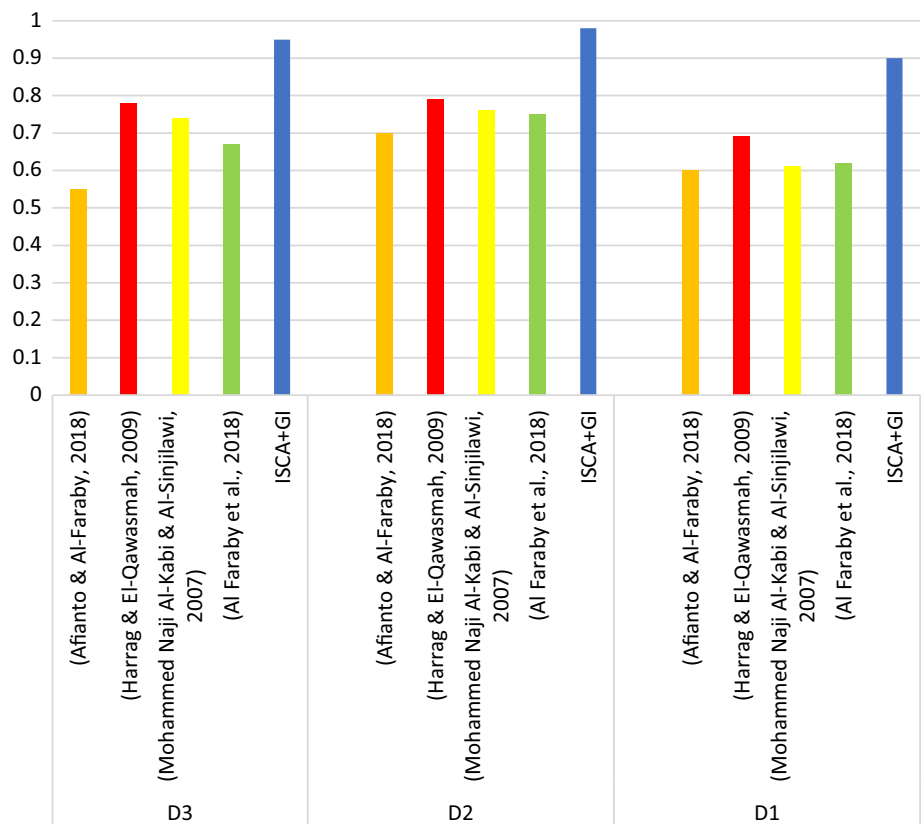


Table 12 Comparison of average classification accuracy between ISCA and other algorithms over 10 runs using UCI datasets

Dataset	Algorithms average accuracy					
	SCA	ISCA	PSO	GA	GOA	HHO
Heart	0.833	0.940	0.881	0.855	0.822	0.911
credit	0.748	0.829	0.800	0.766	0.740	0.793
Australian	0.771	0.915	0.884	0.849	0.830	0.888
exactly	0.802	1	0.876	0.756	0.734	0.939
Spect	0.768	0.914	0.854	0.821	0.761	0.847
QSAR	0.863	0.926	0.896	0.877	0.869	0.902
exactly2	0.773	0.817	0.796	0.793	0.762	0.786
Vehicle	0.754	0.826	0.768	0.765	0.720	0.786
tic-tac-toe	0.831	0.854	0.845	0.808	0.786	0.846
Vote	0.966	0.990	0.983	0.980	0.963	0.980
ionosphere	0.897	0.977	0.943	0.934	0.891	0.957
Vowel	0.952	0.980	0.969	0.958	0.927	0.971
Sonar	0.899	0.990	0.957	0.937	0.899	0.966
spambase	0.900	0.942	0.932	0.925	0.898	0.923

The values in bold indicates the best results

Table 13 Comparison of the average number of selected features between ISCA and other algorithms over 10 runs using UCI datasets

Dataset	Average number of selected features					
	SCA	ISCA	PSO	GA	GOA	HHO
Heart	7.5	4.4	6.04	5	6	4.6
credit	11.6	7.7	9.5	9.7	10.2	8.1
Australian	6.8	4.6	6.0	5.9	6.5	4.9
exactly	10.5	6	6.7	7.9	6.5	6.8
Spect	10.4	7.8	10.4	10.3	10.9	7.8
QSAR	26.7	16.4	20.5	18.9	20.5	18.2
exactly2	8.4	7	6.2	6.6	6.0	5.8
Vehicle	12.5	8.2	8.5	8.2	8.5	8.4
tic-tac-toe	7.2	6.1	4.9	5.7	5.0	6.5
Vote	8	3.1	6.7	3.8	8.5	4.1
ionosphere	15.1	7.4	14.8	13.2	16.0	7.2
vowel	9.2	7.4	7	8	6.9	7.8
sonar	32.6	17	27.6	28.2	30.1	16.3
spambase	43	30.2	28.6	29.7	28.6	26.9

The values in bold indicates the best results

problem, and its ability to work on datasets from different fields with different dimensionalities.

It is noteworthy that we have conducted an ablation study over different phases. First, we considered taking all the features without removing any of them. The results of applying KNN over all features without using feature selection are shown in Table 5. Moreover, we ranked all

Table 14 Comparison of fitness between ISCA and other algorithms over 10 runs using UCI datasets

Dataset	Algorithms average fitness					
	SCA	ISCA	PSO	GA	GOA	HHO
Heart	0.170	0.069	0.121	0.147	0.180	0.091
credit	0.255	0.171	0.202	0.237	0.263	0.209
Australian	0.231	0.086	0.118	0.153	0.172	0.114
exactly	0.204	0.001	0.128	0.248	0.268	0.066
Spect	0.234	0.087	0.148	0.182	0.241	0.154
QSAR	0.141	0.075	0.107	0.126	0.134	0.101
exactly2	0.231	0.184	0.207	0.210	0.240	0.216
Vehicle	0.249	0.175	0.233	0.237	0.281	0.216
tic-tac-toe	0.174	0.147	0.160	0.196	0.218	0.160
Vote	0.038	0.002	0.019	0.022	0.041	0.022
ionosphere	0.105	0.024	0.059	0.069	0.112	0.044
vowel	0.054	0.018	0.036	0.047	0.078	0.035
sonar	0.104	0.003	0.046	0.066	0.104	0.035
spambase	0.106	0.061	0.072	0.079	0.105	0.080

The values in bold indicates the best results

the features using GI, then we removed the features with the lowest weights. Next, we applied the ISCA on the top-weighted features according to the GI ranking, whereas in another experiment we applied the ISCA without using the GI. From these experiments, based on Tables 5, 7, and 8, one can notice clearly the outperformance of the models that employed the FS in comparison with applying the KNN on the whole set of features. For example, the accuracy resulted from using the KNN on the whole set of features in D1 is 65%, whereas the accuracy resulted from using the ISCA for FS is 83%. Furthermore, the accuracy resulted from using ISCA+GI for FS is 90%. For D2, the accuracy using the KNN on the whole set of features led to a classification accuracy of 77%, whereas using the ISCA and ISCA+GI for FS generated 96% and 98% classification accuracy, respectively. For D3, the classification accuracy of using KNN on the whole set of features was 75%, whereas the accuracy raised to 90% and 95% as a result of using the ISCA and ISCA+GI, respectively. Clearly, these results highlight the impact of applying FS on the overall performance. In particular, it is evident that considering all the features for classification purpose lead to poor performance in terms of accuracy. On the other hand, selecting relevant features improves the classification accuracy significantly.

In conclusion, taking all the ISCA obtained results in contrast to other Hadith classification baseline works and optimization algorithms, the proposed ISCA proved its stability in terms of feature reduction and outperformed the competent algorithms in terms of classification accuracy.

Table 15 Comparison of classification accuracy between ISCA and other optimization algorithms by using paired T-test

Dataset	GI ratio	SCA		PSO		GA		GOA		HHO	
		<i>t</i>	<i>P</i>	<i>t</i>	<i>P</i>	<i>t</i>	<i>P</i>	<i>t</i>	<i>P</i>	<i>t</i>	<i>P</i>
D1	10	-4.73	1.66E-04	-4.63	2.09E-04	-2.35	3.06E-02	-5.73	1.99E-05	-2.63	1.69E-02
D1	20	-8.81	6.08E-08	-5.14	6.89E-05	-4.28	4.55E-04	-10.14	7.26E-09	-6.52	4.00E-06
D1	30	-7.20	1.06E-06	-3.38	3.35E-03	-3.50	2.56E-03	-10.06	8.21E-09	-7.25	9.76E-07
D2	10	-5.60	2.61E-05	-6.33	5.73E-06	-2.89	9.84E-03	-5.05	8.27E-05	-6.15	8.37E-06
D2	20	-6.31	6.00E-06	-4.69	1.80E-04	-3.01	7.50E-03	-6.80	2.30E-06	-5.86	1.50E-05
D2	30	-10.49	4.26E-09	-4.01	8.13E-04	-2.58	1.91E-02	-11.43	1.09E-09	-9.43	2.20E-08
D3	10	-6.98	1.62E-06	-4.04	7.68E-04	-3.31	3.87E-03	-9.85	1.12E-08	-5.32	4.72E-05
D3	20	-8.62	8.30E-08	-4.36	3.79E-04	-4.41	3.35E-04	-8.11	2.01E-07	-9.91	1.02E-08
D3	30	-11.10	1.75E-09	-5.20	6.05E-05	-4.35	3.90E-04	-12.98	1.41E-10	-6.67	2.94E-06

Therefore, we can conclude that the proposed modifications to the SCA improved its ability to solve different types of datasets as confirmed from results, which in turn prove its generality and applicability to different domains.

5 Discussions

As discussed above, the current study has several contributions including 1) the development of hybrid FS approach based on filter approach combined with wrapper approach, 2) the improvement introduced to the SCA algorithm solves its weakness and makes it fit for FS on Hadith classification, 3) the collection of the three new Hadith datasets, and 4) to confirm the superiority of the proposed algorithm, we tested it on 14 datasets from the UCI repository. Furthermore, we implemented some other optimization algorithms for FS on Hadith classification, and we compared the ISCA results with these algorithms' results. From the obtained results, the ISCA outperformed these well-known optimization algorithms. Also, to confirm the usefulness of the proposed approach, we compared the ISCA results with other existing works on Hadith classification as shown in Table 11, and from these results, it is confirmed the outperformance of the ISCA over these works. The advantage of the proposed approach over previous Hadith works resulted from its ability to select the most informative features and discard irrelevant features, whereas in the previous Hadith works no FS was conducted using an optimization algorithm. Accordingly, they got poor classification performance because they included irrelevant features in the classification process. Besides, another advantage of the proposed approach over these is the reduction in features dimensionality. Therefore, one can clearly notice how the ISCA algorithm excels in Hadith classification. Thus, this can motivate other researchers to apply it to other new datasets. For future work, the proposed hybrid feature selection method can be applied to

other problems such as sentiment analysis and classification problems. Besides, the ISCA can be applied to engineering problems, global optimization problems, and other problems.

In a nutshell, the proposed work can improve the behavior of SCA algorithm to be more suitable for FS problems in Hadith classification. The main contributions of this work can be summarized as follows. We, firstly, collect a number of newly hadiths datasets. Thereafter, the KNN classifiers is applied with and without using feature selection process with favor to the rear one. Thus, it is clearly noticed that the FS will improve the classification performance for Hadith datasets. The research, then, considers improving SCA by utilizing the chaotic theory and hybridizing with SA as a local exploiter. These improvements help in overcoming the weaknesses of SCA, and thus, the proposed algorithm performs better, solving the FS problem. In addition, we hybridized ISCA with GI to take the advantages of both and discard their disadvantages. To check and confirm the superiority of our work, we compared the ISCA with SCA and other optimization algorithms. Moreover, we compared the ISCA achieved results with other Hadith baseline works. Furthermore, we applied ISCA on another benchmarks datasets from UCI repository to further confirm the outperformance and superiority of the proposed algorithm. The proposed framework is able to powerfully classify the Hadith datasets with high accurate results, and thus, it can be generalized for other classification problems with more complex features.

6 Conclusion

This research work mainly focused on improving the SCA optimization algorithm and applying it to the FS problem for Hadith classifications. Thus, in this research, two main improvements were proposed to the standard SCA. The

first improvement included updating the r_4 parameter value in standard SCA with a singer chaotic map to improve its solution diversity. The second improvement included hybridization of the SA algorithm inside the improvement loop of SCA as a local search algorithm to improve SCA exploitation. In addition, the GI filter feature selection was used to avoid ISCA from exploring irrelevant search space areas. GI is used to rank all features, and the top-ranked GI features were selected based on specified ratios. The improved ISCA algorithm was applied to GI selected features in wrapper mode. ISCA selects the optimal features subset from whole features to train the KNN classifier and improving Hadiths classification accuracy.

To evaluate the improved ISCA algorithm, it was compared with other well-known optimization algorithms and the most recent algorithms including the standard SCA, GA, PSO, GOA, and HHO. In addition, ISCA+GI was compared with Hadiths classification baseline works. Three types of Hadith datasets are used to evaluate the proposed methods. Three types of measurements are used to measure the viability of the proposed method which are: classification accuracy, number of selected features, and the fitness function values.

The achieved results from all experiments prove that the ISCA algorithm achieved better classification accuracy with a reduced number of selected features. Specifically, the attained accuracy and fitness values by the ISCA outperformed all other algorithms over all experiments. It also outperformed other algorithms in the majority of the experiments in terms of features reduction. Thus, the proposed improved ISCA can improve Hadiths classification accuracy and achieve comparable results to related Hadiths classification baseline works. Furthermore, to confirm the generality of the proposed ISCA algorithm, it was applied to 14 benchmark datasets from the UCI repository. From the obtained results, it is also confirmed that ISCA outperformed other optimization algorithms over these used UCI datasets in terms of classification accuracy. This outperformance proves the ability of the ISCA to be applied to different types of problems. More investigation on the proposed algorithm could be made in the future by applying the ISCA to other domains such as Malay Hadiths translated versions or other languages, sentiment analysis, and parameters optimization.

Acknowledgements This study was funded by the University of Malaya (Grant No: UMRG RP043C-17HNE).

Declarations

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

References

1. AbuZeina D, Al-Anzi FS (2018) Employing fisher discriminant analysis for Arabic text classification. *Comput Electr Eng* 66:474–486
2. Afianto MF, Al-Faraby S et al. (2018) Text categorization on hadith Sahih Al-Bukhari using randomforest. In: *Journal of physics: conference series*, vol 971. IOP Publishing, 012037
3. Afshar-Nadjafi B, Yazdani M, Majlesi M (2017) A hybrid of Tabu search and simulated annealing algorithms for preemptive project scheduling problem. In: Benferhat S, Tabia K, Ali M (eds) *Advances in artificial intelligence: from theory to practice*. IEA/AIE, vol 10350. *Lecture Notes in Computer Science*. Springer, Cham
4. Al-Anzi FS, AbuZeina D (2018) Beyond vector space model for hierarchical Arabic text classification: a Markov chain approach. *Inf Process Manage* 54(1):105–115
5. Al-Kabi MN, Al-Sinjalawi SI (2007) A comparative study of the efficiency of different measures to classify Arabic text. *Univ Sharjah J Pure Appl Sci* 4(2):13–26
6. Al-Kabi MN, Kanaan G, Al-Shalabi R, Al-Sinjalawi SI, Al-Mustafa RS (2005) Al-Hadith text classifier. *J Appl Sci* 5(3):584–587
7. Alkhatib M (2010) Classification of Al-Hadith Al-Shareef using data mining algorithm. In: *European, mediterranean and middle eastern conference on information systems*. EMCIS2010, Abu Dhabi, UAE, pp 1–23
8. Al Faraby S, Riviera E, Jasin R (2018) Classification of hadith into positive suggestion, negative suggestion, and information Classification of hadith into positive suggestion, negative suggestion, and information. *J Phys Conf Ser* 971(012046):1–8
9. El-Halees AM (2008) A comparative study on Arabic text classification. *Egypt Comput Sci J* 30(2)
10. Aljarah I, Al-Zoubi AM, Faris H et al (2018) Simultaneous feature selection and support vector machine optimization using the grasshopper optimization algorithm. *Cogn Comput* 10:478–495
11. Arora S, Anand P (2019) Binary butterfly optimization approaches for feature selection. *Expert Syst Appl* 116:147–160
12. Azmi R, Pishgoo B, Norozi N, Koochzadi M, Baesi F (2010) A hybrid GA and SA algorithms for feature selection in recognition of hand-printed Farsi characters. In: *2010 IEEE international conference on intelligent computing and intelligent systems*, pp 384–387
13. Bahassine S, Madani A, Al-Sarem M, Kissi M (2018) Feature selection using an improved Chi-square for Arabic text classification. *J King Saud Univ Comput Inf Sci* 32:225–231
14. Belazzoug M, Touahria M, Nouioua F, Brahimi M (2019) An improved sine cosine algorithm to select features for text categorization. *J King Saud Univ-Comput Inf Sci* 32:454–464
15. Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) *Classification and Regression Trees*, The Wadsworth Statistics and Probability Series, Wadsworth International Group, Belmont California (pp. 356)
16. Eesa AS, Orman Z, Brifcani AMA (2015) A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems. *Expert Syst Appl* 42(5):2670–2679
17. Elaziz MA, Oliva D, Xiong S (2017) An improved opposition-based sine cosine algorithm for global optimization. *Expert Syst Appl* 90:484–500
18. Elaziz Mohamed EA, Ewees AA, Oliva D, Duan P, Xiong S (2017) A hybrid method of sine cosine algorithm and differential evolution for feature selection. In: *International conference on neural information processing*, Springer, Berlin, pp 145–155

19. Elgamel ZM, Yasin NBM, Tubishat M, Alswaitti M, Mirjalili S (2020) An improved harris hawks optimization algorithm with simulated annealing for feature selection in the medical field. *IEEE Access* 8:186638–186652
20. Emary E, Zawbaa HM, Grosan C, Hassenian AE (2015) Feature subset selection approach by gray-wolf optimization. In: *Afro-European conference for industrial advancement*. Springer, Cham, pp 1–13
21. Feng Y, Yang J, Wu C, Lu M, Zhao X-J (2018) Solving 0–1 knapsack problems by chaotic monarch butterfly optimization algorithm with Gaussian mutation. *Memetic Comput* 10(2):135–150
22. Feng Z-K, Liu S, Niu W-J, Li B-J, Wang W-C, Luo B, Miao S-M (2020) A modified sine cosine algorithm for accurate global optimization of numerical functions and multiple hydropower reservoirs operation. *Knowl Based Syst* 208:106461
23. Gu S, Cheng R, Jin Y (2018) Feature selection for high-dimensional classification using a competitive swarm optimizer. *Soft Comput* 22(3):811–822
24. Guo W-Y, Wang Y, Dai F, Xu P (2020) Improved sine cosine algorithm combined with optimal neighborhood and quadratic interpolation strategy. *Eng Appl Artif Intell* 94:103779
25. Gupta S, Deep K (2019) Improved sine cosine algorithm with crossover scheme for global optimization. *Knowl-Based Syst* 165:374–406
26. Gupta S, Deep K (2020) Hybrid sine cosine artificial bee colony algorithm for global optimization and image segmentation. *Neural Comput Appl* 32(13):9521–9543
27. Hammouri AI, Mafarja M, Al-Betar MA, Awadallah MA, Abudoush I (2020) An improved Dragonfly Algorithm for feature selection. *Knowl Based Syst* 203:106131
28. Hans R, Kaur H (2020) Hybrid binary Sine Cosine Algorithm and Ant Lion Optimization (SCALO) approaches for feature selection problem. *Int J Comput Mater Sci Eng* 9(01):1950021
29. Harrag F, El-Qawasmah E (2009) Neural network for Arabic text classification. In: *Second international conference on the applications of digital information and web technologies*. IEEE, London, UK, pp 778–783
30. Harrag F, El-Qawasmah E, Al-Salman AMS (2011) Stemming as a feature reduction technique for Arabic Text Categorization. In: *2011 10th international symposium on programming and systems*, pp 128–133
31. Hu P, Pan J-S, Chu S-C (2020) Improved binary grey wolf optimizer and its application for feature selection. *Knowl Based Syst* 195:105746
32. Javidrad F, Nazari M (2017) A new hybrid particle swarm and simulated annealing stochastic optimization method. *Appl Soft Comput* 60:634–654
33. Jurgens H, Peitgen H-O, Saupe D (1992) *Chaos and fractals: new frontiers of science*. New Springer-Verlag, New York
34. Kaveh A, Javadi S (2019) *Chaos-based firefly algorithms for optimization of cyclically large-size braced steel domes with multiple frequency constraints*. *Comput Struct* 214:28–39
35. Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. *Science* 220(4598):671–680
36. Lan S, Fan W, Liu T, Yang S (2019) A hybrid SCA-VNS metaheuristic based on Iterated Hungarian algorithm for physicians and medical staff scheduling problem in outpatient department of large hospitals with multiple branches. *Appl Soft Comput* 85:105813
37. Dua D, Graff C (2017) UCI machine learning repository. Retrieved from <http://archive.ics.uci.edu/ml>
38. Lu Y, Liang M, Ye Z, Cao L (2015) Improved particle swarm optimization algorithm and its application in text feature selection. *Appl Soft Comput* 35:629–636
39. Mafarja M, Mirjalili S (2018) Whale optimization approaches for wrapper feature selection. *Appl Soft Comput* 62:441–453
40. Mafarja MM, Mirjalili S (2017) Hybrid whale optimization algorithm with simulated annealing for feature selection. *Neurocomputing* 260:302–312
41. Meedeniya D, Perera A (2009) Evaluation of partition-based text clustering techniques to categorize Indic language documents. In: *2009 IEEE international advance computing conference*, pp 1497–1500. IEEE
42. Meedeniya DA, Perera AS (2008) A comparative study on data representation to categorize text documents. In: *Proceedings of the 20th international conference on software engineering and knowledge engineering (SEKE'08)*, pp 371–374
43. Mirjalili S (2016) SCA: a sine cosine algorithm for solving optimization problems. *Knowl-Based Syst* 96:120–133
44. Mirjalili S, Gandomi AH (2017) Chaotic gravitational constants for the gravitational search algorithm. *Appl Soft Comput* 53:407–419
45. Najeeb MM (2014) Towards innovative system for Hadith Isnad processing. *Int J Comput Trends Technol* 18(6):257–259
46. Najib SRM, Rahman NA, Ismail NK, Nor ZM, Alias MN, Alias N (2017) Comparative study of machine learning approach on Malay translated Hadith text classification based on Sanad. In: *8th international conference on mechanical and manufacturing engineering 2017 (ICME'17) MATEC Web Conf*, vol 135, pp 1–9
47. Oliva D, El Aziz MA, Hassanien AE (2017) Parameter estimation of photovoltaic cells using an improved chaotic whale optimization algorithm. *Appl Energy* 200:141–154
48. Potthuri S, Shankar T, Rajesh A (2016) Lifetime improvement in wireless sensor networks using hybrid differential evolution and simulated annealing (DESA). *Ain Shams Eng J* 9:655–663
49. Raileanu LE, Stoffel K (2004) Theoretical comparison between the gini index and information gain criteria. *Ann Math Artif Intell* 41(1):77–93
50. Ramteke SP, Gurjar AA, Deshmukh DS (2019) A novel weighted SVM classifier based on SCA for handwritten marathi character recognition. *IETE J Res*. <https://doi.org/10.1080/03772063.2019.1623093>
51. Riahi V, Kazemi M (2018) A new hybrid ant colony algorithm for scheduling of no-wait flowshop. *Oper Res Int Journal* 18(1):55–74
52. Saloot MA, Idris N, Mahmud R, Ja'afar R, Thorleuchter D, Gani A (2016) Hadith data mining and classification: a comparative analysis. *Artif Intell Rev* 46(1):113–128
53. Sayed GI, Hassanien AE, Azar AT (2019) Feature selection via a novel chaotic crow search algorithm. *Neural Comput Appl* 31(1):171–188
54. Sayed GI, Tharwat A, Hassanien AE (2019) Chaotic dragonfly algorithm: an improved metaheuristic algorithm for feature selection. *Appl Intell* 49(1):188–205
55. Sharma N, Kaur A, Sharma H, Sharma A, Bansal JC (2019) *Chaotic Spider Monkey Optimization Algorithm with Enhanced Learning Soft Computing for Problem Solving* (pp. 149–161): Springer.
56. Sihwail R, Omar K, Ariffin KAZ, Tubishat M (2020) Improved harris hawks optimization using elite opposition-based learning and novel search mechanism for feature selection. *IEEE Access* 8:121127–121145
57. Tawhid MA, Savsani P (2019) Discrete sine-cosine algorithm (DSCA) with local search for solving traveling salesman problem. *Arab J Sci Eng* 44(4):3669–3679
58. Tayal A, Singh SP (2018) Integrating big data analytic and hybrid firefly-chaotic simulated annealing approach for facility layout problem. *Ann Oper Res* 270(1–2):489–514

59. Tubishat M, Abushariah MAM, Idris N et al (2019) Improved whale optimization algorithm for feature selection in Arabic sentiment analysis. *Appl Intell* 49:1688–1707. <https://doi.org/10.1007/s10489-018-1334-8>
60. Tubishat M, Alswaitti M, Mirjalili S, Al-Garadi MA, Rana TA (2020) Dynamic butterfly optimization algorithm for feature selection. *IEEE Access* 8:194303–194314
61. Tubishat M, Idris N, Shuib L, Abushariah MA, Mirjalili S (2020) Improved Salp Swarm Algorithm based on opposition based learning and novel local search algorithm for feature selection. *Expert Syst Appl* 145:113122
62. Tubishat M, Ja'afar S, Alswaitti M, Mirjalili S, Idris N, Ismail MA, Omar MS (2020) Dynamic Salp swarm algorithm for feature selection. *Expert Syst Appl* 164:113873
63. Wan M, Chen X, Zhan T, Xu C, Yang G, Zhou H (2021) Sparse fuzzy two-dimensional discriminant local preserving projection (SF2DDLPP) for robust image feature extraction. *Inf Sci* 563:1–15
64. Wan M, Yang G, Sun C, Liu M (2019) Sparse two-dimensional discriminant locality-preserving projection (S2DDLPP) for feature extraction. *Soft Comput* 23(14):5511–5518
65. Wang M, Chen H, Yang B, Zhao X, Hu L, Cai Z, Huang H, Tong C (2017) Toward an optimal kernel extreme learning machine using a chaotic moth-flame optimization strategy with applications in medical diagnoses. *Neurocomputing* 267:69–84
66. Wolpert DH, Macready WG (1997) No free lunch theorems for optimization. *IEEE Trans Evol Comput* 1(1):67–82
67. Xiang J, Han X, Duan F, Qiang Y, Xiong X, Lan Y, Chai H (2015) A novel hybrid system for feature selection based on an improved gravitational search algorithm and k-NN method. *Appl Soft Comput* 31:293–307
68. Yan C, Ma J, Luo H, Patel A (2019) Hybrid binary coral reefs optimization algorithm with simulated annealing for feature selection in high-dimensional biomedical datasets. *Chemom Intell Lab Syst* 184:102–111
69. Zawbaa HM, Emary E, Parv B, Sharawi M (2016) Feature selection approach based on moth-flame optimization algorithm. In: 2016 IEEE congress on evolutionary computation (CEC), pp 4612–4617. <https://doi.org/10.1109/CEC.2016.7744378>
70. Zhang X, Feng T (2018) Chaotic bean optimization algorithm. *Soft Comput* 22(1):67–77
71. Zhao Y, Zou F, Chen D (2019) A discrete sine cosine algorithm for community detection. In: International conference on intelligent computing. Springer, pp 35–44

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

Authors and Affiliations

Mohammad Tubishat¹ · Salinah Ja'afar² · Norisma Idris³ · Mohammed Azmi Al-Betar^{4,5} · Mohammed Alswaitti⁶ · Hazim Jarrah⁷ · Maizatul Akmar Ismail³ · Mardian Shah Omar²

✉ Salinah Ja'afar
b1salina@um.edu.my

Mohammad Tubishat
mohammad.tubishat@zu.ac.ae

Norisma Idris
norisma@um.edu.my

Mohammed Azmi Al-Betar
m.albetar@ajman.ac.ae

Mohammed Alswaitti
alswaitti.mohammed@xmu.edu.my

Hazim Jarrah
hazimjarrah@gmail.com

Maizatul Akmar Ismail
maizatul@um.edu.my

Mardian Shah Omar
mardianso@um.edu.my

² Department of Linguistics, Academy of Malay Studies
University of Malaya, 50603 Kuala Lumpur, Malaysia

³ Department of Artificial Intelligence, Faculty of Computer
Science and Information Technology, University of Malaya,
50603 Kuala Lumpur, Malaysia

⁴ Artificial Intelligence Research Center (AIRC), Artificial
Intelligence Research Center (AIRC), Department of
Information Technology, MSAI College of Engineering and
Information Technology, Ajman University, Ajman, United
Arab Emirates

⁵ Department of Information Technology, Al-Huson University
College, Al-Balqa Applied University, Al-Huson, P.
O. Box 50, Irbid, Jordan

⁶ School of Electrical and Computer Engineering (ICT),
Xiamen University Malaysia, Bandar Sunsuria,
43900 Sepang, Selangor Darul Ehsan, Malaysia

⁷ School of Information Technology, Skyline University
College, Sharjah, United Arab Emirates

¹ College of Technological Innovation, Zayed University,
Abu Dhabi, United Arab Emirates

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval, sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com