



Late acceptance hill climbing aided chaotic harmony search for feature selection: An empirical analysis on medical data

Anurup Naskar^{a,*}, Rishav Pramanik^a, S.K. Sabbir Hossain^a, Seyedali Mirjalili^{b,c,d}, Ram Sarkar^a

^a Department of Computer Science and Engineering, Jadavpur University, Kolkata, 700032, India

^b Centre for Artificial Intelligence Research and Optimization, Torrens University, Australia, Fortitude Valley, Brisbane, QLD, 4006, Australia

^c Yonsei Frontier Lab, Yonsei University, Seoul, Republic of Korea

^d University Research and Innovation Center, Obuda University, Budapest, Hungary

ARTICLE INFO

Keywords:

Feature selection
Microarray data
Harmony search
Late Acceptance Hill Climbing
Metaheuristics
COVID-19 data
Optimization
Algorithm

ABSTRACT

In today's era of data-driven digital society, there is a huge demand for optimized solutions that essentially reduce the cost of operation, thereby aiming to increase productivity. Processing a huge amount of data, like the Microarray based gene expression data, using machine learning and data mining algorithms has certain limitations in terms of memory and time requirements. This would be more concerning, when a dataset comes with redundant and non-important information. For example, many report-based medical datasets have several non-informative attributes which mislead the classification algorithms. To this end, researchers have been developing several feature selection algorithms that try to discard the redundant information from the raw datasets before feeding them to machine learning algorithms. Metaheuristic based optimization algorithms provide an excellent option to solve feature selection problems. In this paper, we propose a music-inspired harmony search (HS) algorithm based wrapper feature selection method. At the beginning, we use a chaotic mapping to initialize the population of the HS algorithm in order to better coverage of the search space. Further to complement the inferior exploitation of the HS algorithm, we integrate it with the Late Acceptance Hill Climbing (LAHC) method. Thus the combination of these two algorithms provides a good balance between the exploration and exploitation of the HS algorithm. We evaluate the proposed feature selection method on 15 UCI datasets and the obtained results are found to be better than many state-of-the-art methods both in terms of the classification accuracy and the number of features selected. To evaluate the effectiveness of our algorithm, we utilize a combination of precision, recall, F1 score, fitness value, and execution time as performance indicators. These metrics enable us to obtain a comprehensive assessment of the algorithm's abilities and limitations. We also apply our method on 3 microarray based gene expression datasets used for prediction of cancer to ensure the scalability and robustness as a feature selection method in real-life scenarios. In addition to this, we test our approach using the COVID-19 dataset, and it performs better than several metaheuristic based optimization techniques.

1. Introduction

With the rise of advancement in Artificial Intelligence (AI) powered systems such as image processing, natural language processing, medical imaging and time-series prediction (Li & Liu, 2017) enormous data-points are created on daily basis. Technological advancements have a great impact on today's healthcare system (Malakar et al., 2022) where we can see a large number of computer aided supporting systems. Recently various governmental and non-governmental agencies are pushing hard to digitize medical health records for better treatment and future use of the same. Therefore, a lot of data points are being stored everyday. Application of AI in medical domain to process these data

points on a very large scale is rapidly boosting the global healthcare system (Yao et al., 2022). One such important medical domain is the analysis of Microarray based of gene expression datasets, which are very high-dimensional. However, in such datasets, there are exist only a few genes which are responsible for cancer (Kundu, Chattopadhyay, Cuevas, & Sarkar, 2022). In the sphere of machine learning (ML), the dimensionality reduction is often considered as one of the paramount steps before feeding the raw-data to the intended algorithm. Pre-processing steps such as dimensionality reduction and feature selection are important to reduce the need for computational resources and storage. In addition, feature selection lessens the search and time

* Corresponding author.

E-mail address: rup.anu2020@gmail.com (A. Naskar).

<https://doi.org/10.1016/j.eswa.2023.119745>

Received 31 October 2022; Received in revised form 9 February 2023; Accepted 21 February 2023

Available online 24 February 2023

0957-4174/Crown Copyright © 2023 Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

complexities of the algorithms used. Typically in the literature, we come across two type of feature selection methods: (1) Wrapper and (2) Filter. However, a few researchers also consider another type of feature selection technique, namely embedded method, which is a combination of the two approaches (Dokeroglu, Deniz, & Kiziloz, 2022).

Whereas using in wrapper based methods, we evaluate the usefulness of a subset of features using a learning algorithm, on the other hand, in the case of filter based methods, we select the most informative features using some intrinsic properties of the datasets, and these feature selection methods do not use any learning algorithm during the course of selection. Generally, metaheuristic based optimization algorithms are used in wrapper based approaches. Some popular wrapper based optimization algorithms include Particle Swarm Optimization (PSO), Equilibrium Optimizer (EO) and Memetic Algorithm (MA) among others. Some popular filter methods include the ReliefF, Chi-Square and Mutual Information. Mostly, wrapper based methods are costly in terms of computation, but are effective for producing good results (Guha, Ghosh, Mutsuddi, Sarkar and Mirjalili, 2020). Over the years metaheuristic based wrapper feature selection methods have been proven its worth whilst dealing with many state-of-the-art challenges like image contrast enhancement (Hossain, Mukhopadhyay, Ray, Ghosal, & Sarkar, 2022; Mukhopadhyay, Hossain, Malakar, Cuevas, & Sarkar, 2022), micro-structural classification (Sarkar et al., 2021), medical diagnosis (Pramanik, Sarkar and Sarkar, 2022), facial recognition (Saha et al., 2020) and many more. Recently, we have seen a rapid surge in AI-powered medical detection tools (Pramanik et al., 2022; Pramanik, Dey, Malakar, Mirjalili and Sarkar, 2022). From an AI perspective, the task is to extract certain features and classify based on the extracted features. The “the curse of dimensionality” is a long-lasting and prevalent issue here (Ghosh et al., 2021; Pramanik, Sarkar et al., 2022). Crucial medical diagnosis should be reliable for treatment of any present disease with the best possible medical advice. Thus from a social standpoint we should ensure the robustness, scalability and reliability of AI-powered medical diagnostics. Feature selection in this context plays an important role to rule out certain redundant features to help make AI-powered tools reliable for real life usage.

Finding the most optimal subset of features is considered as an NP-hard problem, and hence researchers have been proposing numerous solutions using optimization methods (Bhattacharyya et al., 2020; Ghosh, Singh, Hong, Geem and Sarkar, 2020). Such methods apply non-deterministic solutions to the problem. The reason behind this is, it is not possible for an algorithm to evaluate all the $2^N - 1$ feature subsets to evaluate the best feature subset. Considering the above fact that we come across numerous algorithms in the literature, it is quite obvious to ask this question: “Why do we need new algorithms to solve optimization problems like feature selection if there already exists a lot?”. To answer this question, we will resort to the “No Free Lunch” theorem (Wolpert & Macready, 1997), which advocates that no single optimization algorithm can take care of all sorts of optimization problems with same efficiency. This very fact motivates researchers to propose new algorithms in order to solve various complex optimization problems. Nature provides researchers with excellent opportunities to mimic itself for devising algorithms that give rise to optimized solutions.

One important aspect which researchers always try to focus on while developing an optimization algorithms is the trade-off between the exploration and exploitation capabilities of the same (Mirjalili, 2015; Mirjalili, Mirjalili, & Lewis, 2014). Although the literature of metaheuristic based optimization algorithms is vast, it does not explore much the use of the local optimization or to search for optimized solutions locally. Although there are limited local search methods found in the literature till date (Al-Betar, Aljarah, Awadallah, Farris, & Mirjalili, 2019; Burke & Bykov, 2017; Kirkpatrick, Gelatt, & Vecchi, 1983), they are often considered useful when used with any global search methods (Ahmed, Ghosh, Singh, Geem, & Sarkar, 2020; Guha, Ghosh, Mutsuddi et al., 2020).

In this article, we aim to improvise the searching capability of a very popular metaheuristic algorithm namely Harmony Search (HS) algorithm. Firstly, we apply a chaotic mapping to initialize the search space, which helps to diversify the search space and enhances the exploration capability of the algorithm. Secondly, we incorporate the Late Acceptance Hill Climbing (LAHC) method into the HS algorithm as a local search method to ensure proper exploitation of the search space. The overall framework has shown to dynamically balance between exploration and exploitation of the HS algorithm. Experimental validation on various UCI datasets like BreastCancer, HeartEW, Lymphography etc. along with Microarray based gene expression datasets (used for cancerous gene selection) shows the effectiveness of the proposed method. In a nutshell, the contributions of the proposed methods can be listed as:

- We propose a hybrid metaheuristic by combining HS and LAHC algorithms to solve feature selection problems.
- The method is aided with chaotic initialization for better exploration of the search space.
- The hybrid method uses LAHC algorithm for performing local search to overcome the inferior exploitation of the HS algorithm.
- The method is evaluated on 15 UCI datasets that include many medical datasets. Additionally, we test our method on Microarray based gene expression datasets and a COVID-19 dataset to ensure the usefulness of the same.

The rest of the paper is organized as follows: Section 2 details the related work in this domain. Section 3 provides the necessary details pertaining to the proposed method. Section 4 explains the relevant results and analyses the findings on UCI datasets. A brief comparative analysis of our proposed method with various other methods is elaborated in Section 5. We then discuss the behavior of our proposed method under Section 6 followed by the results on Microarray datasets in Section 7. We have also provided the results on a COVID-19 dataset in Section 8. Finally, we conclude our paper in Section 9.

2. Related work

The world of metaheuristic algorithms goes way back in 1975 when genetic algorithm (GA) was first devised based on Sir Charles Darwin's theory of “Survival of the fittest” (Holland, 1992). Subsequently, algorithms like Differential Evolution (DE) (Moscato et al., 1989) and Memetic Algorithm (Storn, 1996) came into picture. Slowly, starting from late nineties or early 2000s researchers started gaining interest to develop strategies to solve optimization based problems by mimicking the nature which led to formation of algorithms like HS (Geem, Kim, & Loganathan, 2001) and PSO (Eberhart & Kennedy, 1995). In the recent times mimicking nature to simulate optimal solutions have been one of the most lucrative field of studies. Researchers have adapted the principles of physical processes to model several optimization problems like EO (Faramarzi, Heidarinejad, Stephens, & Mirjalili, 2020), Optics inspired optimization (OIO) (Kashan, 2015), Flow regime optimizer (FRO) (Tahani & Babayan, 2019) etc.

Traditionally, any problem with a cost function can be devised through HS. The simplicity of implementation and a less number of hyperparameters make HS a preferred option to find an optimal solution to various complex optimization problems. As a result of that in the literature, we witness the use of HS to solve a variety of optimization problems. The work by Li, Zou, and Kong (2019) used HS to solve dynamic economic emission dispatch in power systems. Bavkar, Iyer, and Deosarkar (2021) devised HS for alcoholism screening using electroencephalogram (EEG) data. The method used amplitude-modulated band (BAM), and frequency modulated bandwidth (BFM) features of EEG signal and used KNN for classification. However the results obtained were less than the methods which enforced different mixtures of optimization techniques. Rajagopal et al. (2020) applied HS to classify scenes from unmanned aerial vehicles (UAVs). Tuncer, Dogan,

and Acharya (2021) introduced an electroencephalography (EEG) based feature selection model using local binary pattern (CLBP) and wavelet packet decomposition (WPD) techniques where CLBP was applied on decomposed signals and WPD was applied on EEG signals to extract the features. Huang and Chen (2020) modeled HS to detect fake news using features from sequential deep learners like long short-term memory and convolutional neural networks. Radman (2021) used HS for topology optimization of microstructures for materials. Tsakirakis, Marinaki, Marinakis, and Matsatsinis (2019) employed HS for team orienteering problem. The rest of this section discusses the application of basic HS and its variants for feature selection problems.

Inbarani, Bagyamathi, and Azar (2015) proposed hybrid feature selection method based on rough set and improved HS. This technique was tested on medical datasets and was found effective. It was also observed that the time taken for feature selection also decreased than the existing methods. An automated classification technique of remote sensing images was introduced by Karadal, Kaya, Tuncer, Dogan, and Acharya (2021) using multileveled MobileNetV2 and DWT techniques. An iterative neighborhood component analysis (INCA) was applied to reduce the feature set and select the best features which was then validated against two RSIC datasets. Guha, Ghosh, Bera, Sarkar, and Mirjalili (2023) proposed a binary adaptation of the Equilibrium Optimizer (EO) to solve binary optimization problems. Simulated annealing (SA) was used for local search and the method was evaluated against UCI datasets, microarray datasets and binary knapsack problems. Ahmed et al. (2020) proposed a hybrid feature selection algorithm of HS and Ring Theory based Evolutionary Algorithm (RTEA) and concluded that the proposed algorithm performs best with k -NN classifiers. RTEA was used to reduce dependence on the internal parameters in HS algorithm. Dash (2021) proposed an adaptive HS for feature selection using probability based gene selection. The selection was further refined with a bi-objective pareto based feature selection. Sarkar et al. (2021) proposed a wrapper-filter based feature selection technique. Here the authors define a new fitness function by using two very popular filter methods. This improvised algorithm was used to classify microstructural images.

Gholami, Pourpanah, and Wang (2020) proposed an improvised binary HS, where the authors claimed to obtain comparative results from the said improvisation. A genetic mutation strategy was proposed to obtain a better feature set. The algorithm used helped in reaching faster convergence towards optimal solution but involved tuning multiple parameters to achieve the optimal result. Shi et al. (2021) proposed a linear discriminant analysis based binary HS for feature selection. This method was applied to select channels for human computer interfacing, which belongs to a broader field of feature selection. Akbal, Barua, Dogan, Tuncer, and Acharya (2022) introduced a data encryption standard cipher model - DesPatNet25 which utilized a novel hand-modeled automated system and was evaluated on a public construction site monitoring (CSM) sound dataset. An accurate fire learning model was proposed by Dogan et al. (2022) where residual networks (ResNet) was used to create deep models. Two ensemble models were applied where the top 1000 features were chosen using a feature selector and support vector machine (SVM) classifier was used for classification in the first model. An iterative hard majority voting (IHMV) was applied to the generated results in the second model.

Pramanik, Sarkar et al. (2022) recently incorporated dynamic adaptive strategies for PSO by assuming the search time is directly dependent upon solution divergence. Further, they improvised PSO for feature selection using a data driven altruistic strategy. Awadallah, Al-Betar et al. (2022) introduced the S-shaped transfer function for Rat Swarm Optimizer. Further, the authors improvised the algorithms with insights from PSO and crossover operators devised to enhance the local search capability of the algorithms. Hammouri, Mafarja, Al-Betar, Awadallah, and Abu-Doush (2020) proposed a novel binarized version of the dragonfly algorithm devised specifically for the feature selection purpose. Awadallah, Hammouri, Al-Betar, Braik and Abd Elaziz

(2022) proposed a binarized version of Horse Optimizer for the feature selection purpose using multiple transfer functions and three crossover operators. A memory based fuzzy local search algorithm based on MA was proposed by Chatterjee, Ahmed, Bhattacharyya, and Sarkar (2022) which involved dynamic mutation and guided population initialization for solving feature selection and class imbalance problems. The method yielded good results and proved to be superior than many recent optimization algorithms. Cui, Li, Fan, Wang, and Zheng (2020) proposed an improvised version of Dragonfly Algorithm based on redundancy measures. The authors claim enhanced capabilities of the parent algorithm. The work by Bhattacharyya et al. (2020) hybridized Mayfly Algorithm (MA) with HS for dealing with feature selection problems. But there is a limitation to it. For datasets, the approach can occasionally be susceptible to premature convergence. Moayedikia, Ong, Boo, Yeoh, and Jensen (2017) introduced an improvised version of HS dubbed as SYMON which is based on symmetrical uncertainty and vector tuning operations to weigh the features based on their dependency to the class labels to get the best combination of features. This method was evaluated on the Microarray based gene expression datasets.

Guha, Ghosh, Chakrabarti, Sarkar and Mirjalili (2020) introduced chaotic clustering for binary gravitational search algorithm for classical feature selection algorithms. The authors claimed to enhance the exploration capability by introducing clustered initialization. Ghosh, Guha, Sarkar and Abraham (2020) proposed a wrapper-filter based ant colony optimization for feature selection. The authors evaluated the subsets using filter approach for their feature selection algorithm. Though the computational complexity is low because of the use of filter methods, the results obtained are less than the state-of-the-art methods which leveraged supervised learning model. Too, Mafarja, and Mirjalili (2021) proposed spatial bound Whale Optimization Algorithm (WOA) to deal with high dimensional feature selection problems. The authors used spatial bound properties to regulate the dimensions of each search agent.

Saha et al. (2020) proposed supervised filter HS algorithm (SFHSA) which is based on cosine similarity and minimal-redundancy maximal-relevance. This method was used for feature selection in the domain of facial emotion recognition. Bae, Kim, Lim, and Geem (2021) proposed a feature selection method to differentiate between colorectal cancer patients and normal individuals with the help of K-Means clustering and modified harmony memory operator which eliminates worst solution for feature selection from the clusters. Das, Singh, Bhowmik, Sarkar, and Nasipuri (2016) used HS for feature selection for holistic Bangla word recognition. Guha, Das et al. (2020) hybridized Naked Mole Rat Algorithm with HS for speech identification from audio signals. The authors found the method to converge faster than the original HS algorithm.

Ghosh et al. (2021) analysed the use of 10-filter based feature selection methods for microarray data analysis. The work by Kundu et al. (2022) proposes altruistic whale optimization algorithm for microarray data analysis. Ahmed, Sheikh, Mirjalili, and Sarkar (2022) proposes local search aided generalized normal distribution optimizer for COVID-19 feature selection. Similarly (Too & Mirjalili, 2021) proposed a hyper learning based dragonfly algorithm for feature selection on COVID-19 data. However, this approach has a drawback. In some circumstances, it might not be able to locate the global optima necessary to solve the issue. Tran, Xue, and Zhang (2018) proposed Variable-Length Particle Swarm Optimization (VLPSO) for feature selection on microarray data using the concepts of symmetrical uncertainty applied to PSO. A summary of some articles, where researchers have used HS algorithm in different domain is listed in Table 1 for the convenience of readers.

2.1. Literature gap

From the literature survey, it is clearly understood that the literature of HS is very well-developed. Even researchers all around the globe are continuously investing more and more time to come up with even

Table 1

Some applications of HS algorithm in various research fields.

| Work Ref. | Contribution | Classifier | Application area |
|--------------------------|---|------------|--|
| Li et al. (2019) | Improvising on harmony memory operator | N/A | Dynamic economic emission dispatch |
| Saha et al. (2020) | Wrapper-filter | SVM | Facial emotion recognition |
| Ahmed et al. (2020) | Hybridization with ring theory optimizer | k-NN | Feature selection |
| Bavkar et al. (2021) | Integration with statistical features | k-NN | EEG channels selection |
| Bae et al. (2021) | Hybrid with k-Means clustering | SVM | Colon cancer detection |
| Das et al. (2016) | Binary version | MLP | Handwritten word recognition |
| Moayedikia et al. (2017) | Feature weight process | SVM | Class imbalance |
| Dash (2021) | Real number encoding for harmony memory | SVM | Microarray gene classification |
| Gholami et al. (2020) | Integrating global search | k-NN | Feature selection |
| Sarkar et al. (2021) | Wrapper-filter | k-NN | Microstructural image classification |
| Inbarani et al. (2015) | Fusion of rough set with HS | Several | Feature selection |
| Tsakirakis et al. (2019) | Hybridization with similarity process | N/A | Team orienteering problem |
| Shi et al. (2021) | Binary version | Several | Brain-computer interface |
| Radman (2021) | Integrating bidirectional evolutionary structural optimizer | N/A | Topology optimization of microstructures |
| Huang and Chen (2020) | Modified pitch adjustment operator | N/A | Fake news detection |
| Rajagopal et al. (2020) | Modified pitch adjustment operator | SVM | Scene classification for UAV |

better performing algorithms. Hence, it is quite obvious to ask a very fundamental question: *what is the utility of new optimization algorithms?* The “No Free Lunch” theorem by Wolpert and Macready (1997) states “any two optimization algorithms are equivalent when their performance is averaged across all possible problems”. However, it has been frequently observed that most of the algorithms perform good on one problem while the same algorithm fails to perform optimally on another problem. This behavior is possibly observed due to the fact that each of these problems is NP-Hard in nature, and each of the problems presents with a unique computational challenge to counter with. Thus, it is very unfair to claim that any metaheuristic optimization algorithm is “better” in true sense than the other. This is one of the prime reasons which motivates the researchers to design algorithms over the times to solve specific optimization problems.

In this paper, we design a hybrid algorithm to solve the feature selection problem. From the existing literature, we observe that hybrid optimization algorithms are immensely popular in this domain that have been steadily performing well (Ahmed et al., 2020; Awadallah, Hammouri et al., 2022; Bhattacharyya et al., 2020). Specifically, we take the help from the theory of chaos to have chaotic maps included for guided search mechanisms for good exploration of the search phase in the initial phase followed by LAHC based local search for proper exploitation to have produce good solutions.

3. Methodology

This section discusses the relevant algorithms pertaining to the proposed feature selection framework. At first we discuss some preliminaries such as HS and LAHC. Further we discuss the proposed method in detail. The workflow of the proposed hybrid feature selection algorithm is shown in Fig. 1

3.1. Harmony search

HS (Geem et al., 2001), a metaheuristic optimization technique, inspires from musical performances where musicians try various combinations of music pitches to achieve a perfect state of harmony. Here, finding the best aesthetic state is considered as an optimization problem.

In HS, each musician is modeled onto a decision variable, the musical instruments correspond to a range of the decision variables, the solution vector at a certain iteration is the musical harmony. Last but not the least, just as the aesthetic of a fine music concert is based on the expert opinion from the audience, in HS the audience aesthetic is measured by the objective function itself. HS improvises onto a newer (better) harmony based on three principal operators: memory consideration, pitch adjustment, random consideration. Memory consideration is the state when the newer harmony is randomly inherited from one

of the previous harmonies. This can be thought of when an artist uses his/her state of memory to generate a better tune. Pitch adjustment is characterized by a pitch adjustment rate (PAR), which adjusts the pitch by a factor generated. This can be thought of as a scenario when an artist uses his/her state of absolute consciousness to generate a perfect state of melody. The random consideration is when a random solution pops which is required for better exploration. The motivation behind using random consideration arises when the artist struggles to find a state of perfect melody, and then the artist explores random tunes in hope to search for a better one. The steps of HS are presented in Algorithm 1

Algorithm 1 Harmony Search (HS)

```

1: Define HMCR, PAR
2: Define maximum number of iterations NI
3:  $HM \leftarrow \text{Generate Initial Population}()$ 
4: while iter  $\leq$  NI do
5:   if rand(0,1) < HMCR then ▷ HMCR Application
6:     for i in num_features do ▷ PAR Application
7:       Replace  $i^{th}$  feature for the worst agent from any randomly
       selected agent
8:     end for
9:   else ▷ Random Consideration
10:    for i in num_features do
11:      Randomly set the  $i^{th}$  feature for the worst agent
12:    end for
13:  end if
14:  Reevaluate the agents
15:  iter = iter + 1
16: end while
17: Return the best agent

```

3.2. Late acceptance Hill Climbing

LAHC (Burke & Bykov, 2017) incorporates the classic Hill Climbing (HC) heuristic with the basic difference being that the acceptance condition is based on a history of several previous iterations. In contrast to HC which selects a candidate solution based on previous iteration, LAHC takes in a control parameter. In other words, LAHC chooses a solution only if the value is better and stores the worse performing solution to help them converge into a better solution. In this work, we consider the value of length of history equal to 15. An algorithmic pseudo code is presented in Algorithm 2. To model the updated solution in the said algorithm, we adhere to the mutation strategy. For mutation, we randomly select 20% of the features and invert it (if it is initially selected it is rejected and vice-versa).

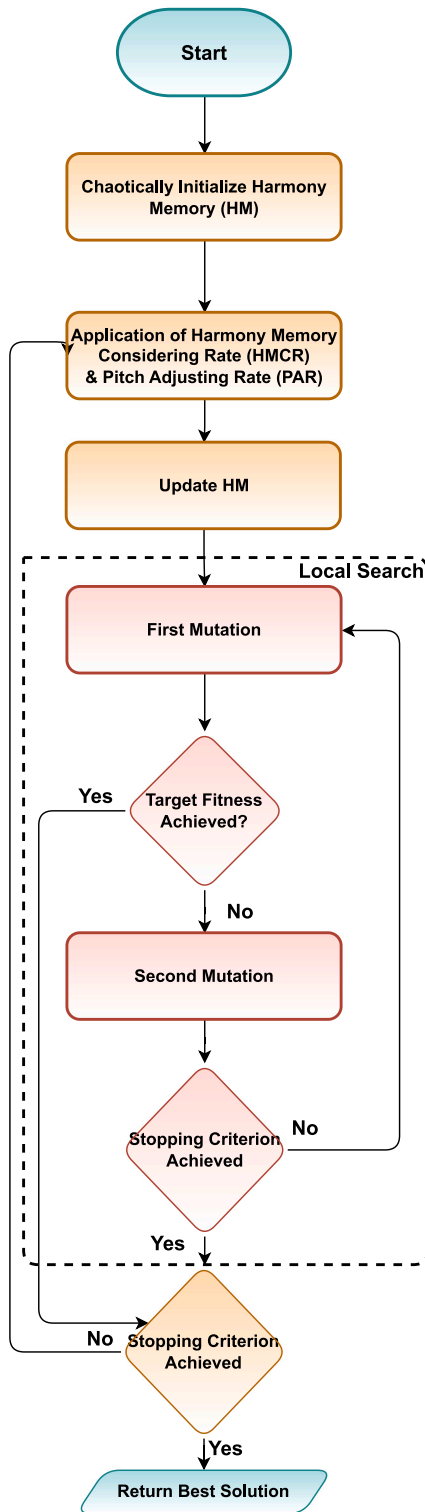


Fig. 1. Workflow of the proposed hybrid feature selection algorithm. The components of HS algorithm are marked in red, whereas the components of the LAHC based local search method are marked in yellow.

3.3. Chaotic initialization

The theory of chaos is widely used in group intelligence algorithms to enhance the diversity. This is useful for population initialization to improve the searching performance of various optimization algorithms because of its randomness and non-repetition. Compared with random

Algorithm 2 Late Acceptance Hill climbing

```

1: Calculate the initial cost  $c$ 
2: Define the solution set  $S$ 
3: Define Target Fitness  $f_t$  & length of history  $l_h$ 
4: for  $i$  in  $l_h$  do
5:    $S_{new} = Mutate(S)$  ▷ Mutation Stage-1
6:   if  $Fitness(S_{new}) \geq f_t$  then
7:     Accept and Return the new Solution
8:   else
9:     Update  $S$  with  $S_{new}$  if fitness increased else discard  $S_{new}$ 
10:    for  $j$  in  $l_h$  do
11:       $S_{new} = Mutate(S)$  ▷ Mutation Stage-2
12:      if  $Fitness(S_{new}) < Fitness(S)$  then
13:        Stop Stage-2 of Mutation ▷ Break the current loop
14:      else
15:        Update  $S$  with  $S_{new}$ 
16:      end if
17:    end for
18:  end if
19: end for
20: Return  $S$ 
  
```

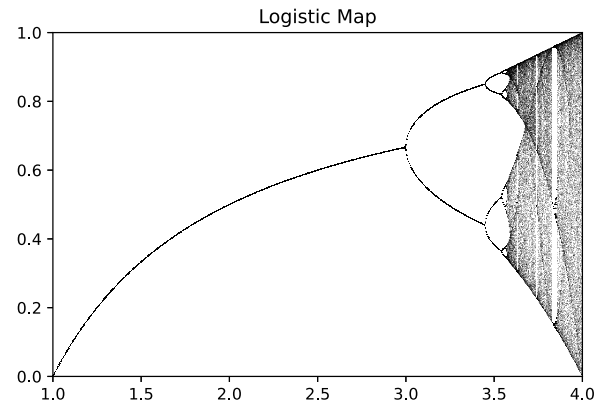


Fig. 2. Bifurcation diagram of logistic chaotic map used in this work.

search, chaos theory can explore the search space thoroughly. To sum up, in order to generate the initial population and to make it able to search the solution space as much as possible in the HS algorithm, we use chaotic maps. In this work, we use the logistic map to initialize the search space (Eq. (1)). A bifurcation diagram of Logistic map is provided under Fig. 2.

$$X_i = 4 \times X_{i-1} \times (1 - X_{i-1}) \quad (1)$$

3.4. Fitness value

Fitness function is a way to determine the quality of a candidate solution. Since the HA algorithm is a wrapper-based method, a learning algorithm needs to be used for the evaluation of its candidate solutions. For this work, we have used the K-Nearest Neighbors (k -NN) classifier to evaluate the classification accuracy of the solution vectors. Our fitness function has two components: classification accuracy and the number of features rejected. Both of these components support a higher value, i.e., higher the value for either of the two or both would be beneficial. This is because a lower error or a high accuracy value implies a high fitness value, so does a lower number of features. Eq. (2) shows the fitness function which evaluates a given feature subset, where $|F|$ is the number of features in the rejected feature subset, $|D|$ is the total number of features of the dataset, acc is the accuracy yielded by the KNN classifier with the feature subset, and α lies in $[0, 1]$, which

indicates the relative weight of the number of features rejected and the classification accuracy.

$$F = \alpha \times acc + (1 - \alpha) \times \frac{|F|}{|D|} \quad (2)$$

3.5. Computational complexity

When suggesting a new algorithm, we should ideally discuss the computational overhead required. In this section, we discuss the time complexity using the \mathcal{O} notation. Let us consider the number of iterations to be N , population size p and feature dimension d , and let the time required to evaluate each agent be f . Therefore, the computation complexity required to perform the local search is $\mathcal{O}(l_h^2 f)$ for each agent. Let us also consider the time required to swap values is constant (although it is technically not possible to be constant but for the sake of understanding we consider it to be constant.) Hence, the net time complexity to run the present algorithm is $\mathcal{O}(N \times (d + p \times f + p \times l_h^2 f))$. It can be noted that there is a slight increase in computation cost.

4. Results and analysis

In this section, we first discuss the datasets and subsequently discuss the experiments and analyse its findings. We have used the Python programming language to implement the proposed algorithm. The experiments have been performed on a Laptop having AMD Ryzen 5 3550H (2.10 GHz) processor with 8 GB of RAM and NVIDIA GeForce GTX 1650 Ti GPU with 4 GB of VRAM. The PC is powered by a Ubuntu LTS operating system.

4.1. Datasets

In this work, 15 different datasets from the UCI machine learning data repository have been chosen as described in the Table 2. These datasets are used to evaluate the efficiency of the proposed method. Mostly these datasets are used for biomedical applications. For example, the BreastCancer dataset is used to diagnose breast cancer, HeartEW is used for heart disease prediction and many more. As we discussed above, in the breast cancer dataset contains information about the radius, texture, smoothness, compactness, and concave points of each cell nuclei among many others. This information is vital in the context of predicting breast cancer. The goal of the HeartEW dataset is to provide whether a patient with a given set of features suffers from any heart disease. The available features are devised from the patient's age, sex, the number of cigarettes a person smokes per day, heart rate, blood pressure, cholesterol information amongst many. These vital parameters should be taken into consideration while designing an AI based automatic disease prediction model. In the case of lymphography dataset, the information about lymph nodes is made available. Thus, for designing an AI-powered tool for the said disease classification-based tasks considering the real-life datasets, we require appropriate techniques to process the available information. To be specific, there are certain features which are more important than others and the same can be useful for more accurate disease prediction. Hence, we have chosen these datasets for assessing the proposed hybrid metaheuristic based feature selection method.

4.2. Controlling parameters

Major controlling parameters for an optimization algorithm are its hyperparameters. These hyperparameters are usually set by experimenting with different values. In this work, we use tried-and-tested values for hyperparameters (Bhattacharyya et al., 2020). Here we consider GA, MA, WOA, Grey Wolf Optimizer (GWO), Bat Algorithm (BA), Equilibrium Optimizer (EO), HS, Red Deer Algorithm (RDA), and PSO for comparing the performance of the proposed feature selection algorithm. The used parameters are tabulated in Table 3.

4.3. Sensitivity analysis

To determine the sensitivity of the proposed algorithm to the various hyperparameters, we conduct a sensitivity analysis on three datasets having different sizes in terms of feature dimension—one small (BreastCancer), one medium (Sonar) and one large (Ionosphere). Fig. 3 shows the relevant results on the small-sized dataset. In the said figure, we observe that our model does not diverge much when we change the parameters. Mostly the accuracy score lies in the range of (96, 100) and the number of features gets selected in the range of (2, 5). Therefore, for both accuracy and the number of features selected, there is no huge divergence for any parameter. In Fig. 4, we see the same for a medium-sized dataset. Similar to our previous observations we see that the proposed algorithm does not diverge much with a change in parameter values. Similarly, in Fig. 5, we observe the sensitivity analysis for a large-sized dataset. It should be noted that the size is in terms of the number of feature attributes.

As we increase the number of attributes, we specifically note one thing—the range of divergence increases with the number of attributes. This behavior can be explained on the basis that, as we increase the number of features, the larger the pool of features and hence more possibility of redundant features. For some of the hyperparameters, more redundant features get included in the final feature set and hence we see an increased range. However, it should be kept in mind that for all datasets, the proposed algorithm does not diverge much from the optimal solution. Therefore, we can claim the proposed algorithm is not very much sensitive to hyperparameter, which can be attributed to the fact that chaos mapping helps to explore the search space better whereas the local search helps to perform better exploitation.

4.4. Experimental results

Table 4 presents a comparative view of the classification accuracies obtained by different optimization algorithms. The effectiveness of the proposed HS+LAHC method has been demonstrated in Table 5 using different evaluation metrics, such as precision, recall, F1 score, fitness value, and execution time(in minutes). Table 6 presents the comparative assessment based on the number of selected features. From Table 4 we observe the proposed method outperforms all other methods on 11 out of 15 datasets used here in terms of classification accuracy. On other datasets also, the results are found quite competitive. Whereas in terms of feature count the results are at par with their counterparts. This facts clearly show the applicability of the proposed framework on real life datasets. In addition to that the proposed framework also provides the required crucial trade-off between classification accuracy and the feature count which is required to design a feature selection algorithm.

4.4.1. Results in terms of classification accuracy

While inspecting Table 4, we observe competitive results for the proposed method. We note that for 11 out of 15 datasets, the proposed method outperforms the existing algorithms. As stated earlier, feature selection is a subset selection problem. Therefore the main objective of a feature selection method is to deduct redundancy from the said feature spaces. Algorithms like GWO, EO and WOA perform competitively well for some datasets but do not show a stable performance across the datasets considered here. This is reflected in Table 4, each of these algorithms works well for certain datasets and does not behave optimally for some other datasets. In this regard we can say that the proposed method is able to provide much stable results across the datasets. A noteworthy fact is that the parent algorithm of the proposed method i.e., HS, gives inferior results than the proposed hybrid method. As a consequence, we can claim the proposed method gives a perfect blend of exploratory capability and exploitative capability which ensures proper feature selection across the datasets.

Table 2
Description of UCI datasets used for experimentation.

| Dataset | No. of features | Sample size | No. of labels | Domain |
|--------------|-----------------|-------------|---------------|-----------------|
| BresatCancer | 9 | 699 | 2 | Biology |
| CongressEW | 16 | 435 | 2 | Politics |
| Exactly | 13 | 1000 | 2 | Biology |
| Exactly2 | 13 | 1000 | 2 | Biology |
| BreastEW | 30 | 569 | 2 | Biology |
| Lymphography | 18 | 148 | 2 | Biology |
| M-of-n | 13 | 1000 | 2 | Biology |
| PenglungEW | 325 | 73 | 2 | Biology |
| Sonar | 60 | 208 | 2 | Physical |
| Vote | 16 | 300 | 2 | Politics |
| HeartEW | 13 | 270 | 2 | Biology |
| Ionosphere | 34 | 351 | 2 | Electromagnetic |
| Wine | 13 | 178 | 3 | Chemistry |
| Zoo | 16 | 101 | 2 | Artificial |
| SpectEW | 22 | 267 | 2 | Biology |

Table 3
Description of hyperparameters used in different optimization algorithms.

| Optimization algorithm | Parameter | Value |
|------------------------|--|--|
| Generic parameters | Population | 30 |
| | Iterations | 60 |
| | Weight for accuracy (α) | $\alpha = 0.8$ |
| GA | Gene selection | Roulette wheel |
| | Crossover probability | 0.8 |
| | Mutation probability | 0.05 |
| EO | Pool size | 4 |
| | Constants ($a1, a2$) | $a1 = 1, a2 = 2$ |
| | Generation | 0.9 |
| MA | Attraction constant ($a1, a2$) | $a1 = 1, a2 = 1.5$ |
| | Initial nuptial dance coefficient | 0.1 |
| | Initial random walk coefficient | 0.1 |
| | Gravitational constant | 0.98 |
| | Visibility coefficient | 2 |
| PSO | Nuptial dance & Random walk updating factor (δ) | 0.9 |
| | Coefficients ($r1, r2$) | $r1$ and $r2$ lie in $[0, 1]$ |
| BA | minFrequency | 0 |
| | maxFrequency | 2 |
| | Loudness | 1 |
| | Pulse emission rate | 0.15 |
| GWO | Coefficients ($r1, r2$) (a, c) | $r1$ and $r2$ lie in $[0, 1]$ a and c lie in $[0, 2]$ |
| RDA | Lower bound | -5 |
| | Upper bound | 5 |
| | Gamma (γ) | 0.5 |
| | Alpha (α) | 0.2 |
| | Beta (β) | 0.1 |
| HS | HMCR | 0.9 |
| WOA | (A, C) | A and C lie in $[0, 2]$ |
| | l | l lies in $[0, 1]$ |
| | p | p lies in $[0, 1]$ |
| | Shape of spiral | 1 |

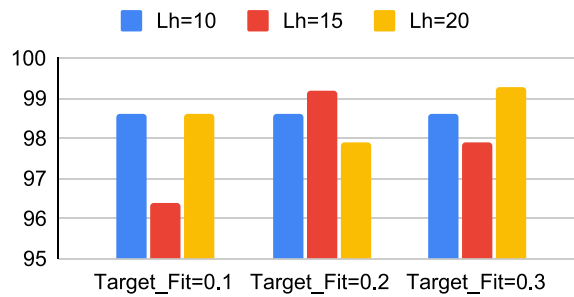
From Table 4 we can conclude that HS+LAHC produces extremely competitive results. In the case of M-of-n dataset, HS+LAHC, GA, MA and GWO produce 100% accuracy. Here, the result from GA is slightly lesser than that of the proposed algorithm by 2.5%. Results from other methods like BA, PSO, RDA and WOA are much less than that of HS+LAHC and HS, having secured last rank with just 80% accuracy. Apart from that in the case of PenglungEW dataset, our proposed method gives 100% classification accuracy along with PSO. In this case, results from WOA, HS and GWO are the closest to our proposed algorithm. Furthermore, remaining wrappers like MA, EO, RDA and BA produce results below 90% percent.

In the case of Vote dataset, it is observed that HS+LAHC outclasses every other existing meta-heuristic based wrappers by holding the top position with 100% percent classification accuracy. Remaining all other methods such as BA, GA and HS produce classification accuracies

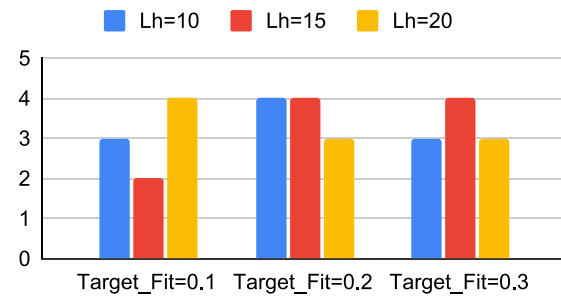
ranging from 93% to 95% with MA being an exception. In BreastEW dataset, although classification accuracy of HS+LAHC is highest with 98.2%, other methods also produce extremely competitive outcomes with a difference of just variation of 2% to 3%. In the case of Exactly dataset, HS+LAHC produces best result of 100% accuracy along with GWO, PSO and WOA. Here, on one side RDA, EO, MA and GA produce competitive results but on the other hand, HS and BA secure the worst two classification accuracies. HS+LAHC and MA outperform other nature based wrappers for the Exactly2 dataset with a very narrow margin. In the case of Sonar dataset not only our proposed algorithm achieves the highest accuracy along with WOA but also remaining other optimization algorithms have produced results, which are quite competitive.

In the case of Lymphography dataset, though the proposed method attains the second position after GA, the results produced by proposed

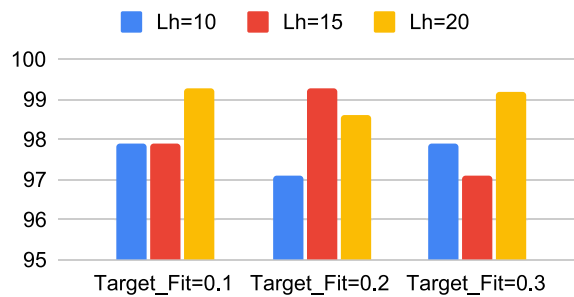
HMCR=0.8 vs Accuracies



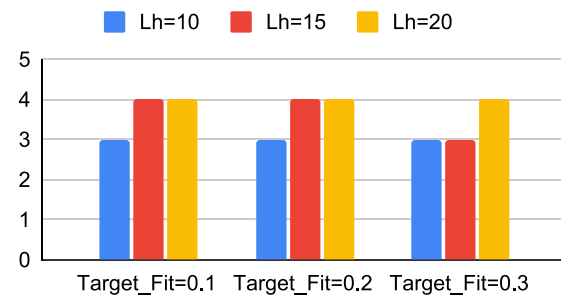
HMCR=0.8 vs #Features



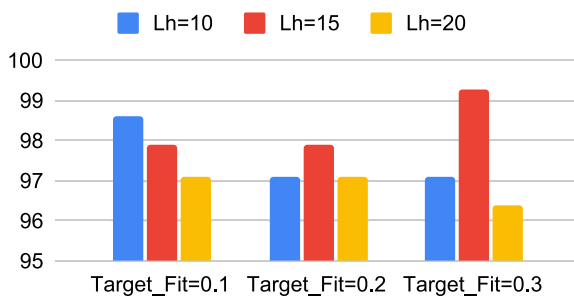
HMCR=0.9 vs Accuracies



HMCR=0.9 vs #Features



HMCR=0.95 vs Accuracies



HMCR=0.95 vs #Features

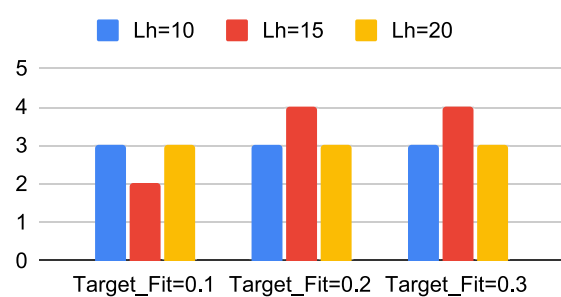


Fig. 3. Results obtained for different sets of hyperparameters on the BreastCancer dataset. Here, HMCR is a component of the HS, whereas History Length (Lh) and Target Fitness (Target_Fit) are components of LAHC.

Table 4

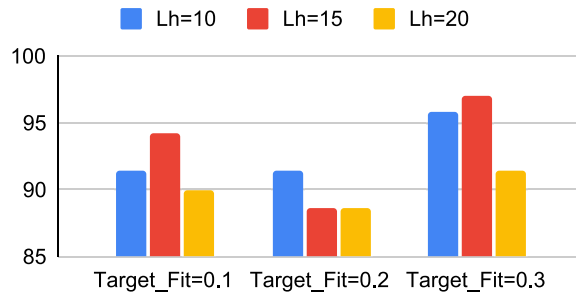
Comparison of proposed method with other optimization algorithms in terms of classification accuracy. Accuracy is in %.

| Dataset | HS+LAHC | BA | MA | EO | GA | GWO | PSO | RDA | WOA | HS |
|--------------|--------------|------|--------------|--------------|--------------|--------------|--------------|------|--------------|--------------|
| Breastcancer | 98.2 | 96.4 | 97.1 | 97.1 | 97.1 | 99.2 | 98.5 | 97.8 | 97.8 | 93.5 |
| CongressEW | 97.7 | 96.5 | 98.8 | 97.7 | 100.0 | 98.8 | 98.8 | 98.8 | 98.8 | 94.2 |
| Exactly | 100.0 | 68.0 | 96.0 | 95.5 | 80.0 | 100.0 | 100.0 | 92.5 | 100.0 | 69.0 |
| Exactly-2 | 79.0 | 77.5 | 79.0 | 76.0 | 77.5 | 78.5 | 78.5 | 76.0 | 76.0 | 76.0 |
| BreastEW | 98.2 | 95.6 | 95.6 | 96.4 | 95.6 | 96.4 | 95.6 | 94.7 | 95.6 | 93.8 |
| Lymphography | 93.3 | 90.0 | 93.3 | 90.0 | 96.6 | 86.6 | 93.3 | 90.0 | 90.0 | 80.0 |
| M-of-n | 100.0 | 84.0 | 100.0 | 97.5 | 100.0 | 100.0 | 83.5 | 92.0 | 94.0 | 80.0 |
| PenglungEW | 100.0 | 73.3 | 86.6 | 86.6 | 86.6 | 93.3 | 100.0 | 80 | 93.3 | 93.3 |
| Sonar | 95.2 | 90.4 | 92.8 | 92.8 | 90.4 | 90.4 | 95.2 | 90.4 | 95.2 | 85.7 |
| Vote | 100.0 | 95.0 | 86.6 | 96.6 | 98.3 | 98.3 | 95.0 | 98.3 | 98.3 | 96.6 |
| HeartEW | 88.8 | 83.3 | 81.4 | 88.8 | 83.3 | 87.0 | 85.1 | 81.4 | 83.3 | 81.4 |
| Ionosphere | 95.7 | 92.8 | 90.0 | 92.8 | 94.2 | 88.5 | 92.8 | 92.8 | 92.8 | 92.8 |
| Wine | 100.0 | 97.2 | 97.2 | 100.0 | 100.0 | 100.0 | 100.0 | 94.4 | 97.2 | 100.0 |
| Zoo | 100.0 | 90.0 | 100.0 | 100.0 | 100.0 | 100.0 | 95.0 | 95.5 | 100.0 | 100.0 |
| SpectEW | 87.0 | 83.3 | 90.7 | 92.5 | 87.0 | 88.8 | 87.0 | 79.6 | 92.5 | 87.0 |

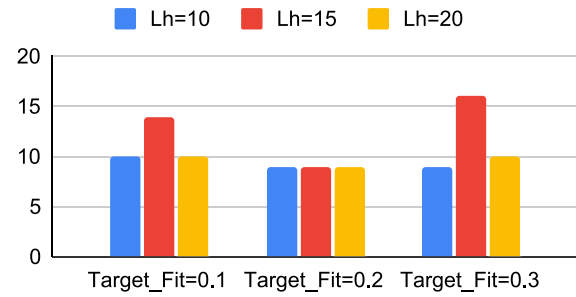
algorithm as well as other nature based wrappers are quite up to the mark. For Ionosphere, our method stands at first position and the remaining methods also produce competitive results. In the case of BreastCancer dataset also, our method stands at the topmost position with the other wrappers producing satisfactory results. For HeartEW

dataset, HS+LAHC and EO produces the highest accuracy of 88.88% along with remaining methods producing very good results. Our proposed method secures the worst position on the basis of classification accuracy for CongressEW dataset with accuracy of 97.7%. Here in this case GA secures the best outcome with an accuracy of 100% and all

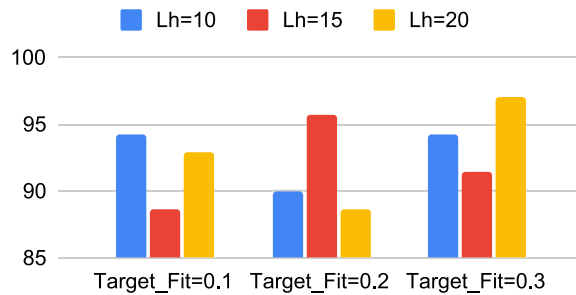
HMCR=0.8 vs Accuracies



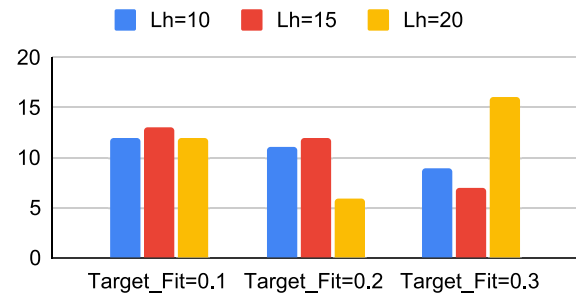
HMCR=0.8 vs #Features



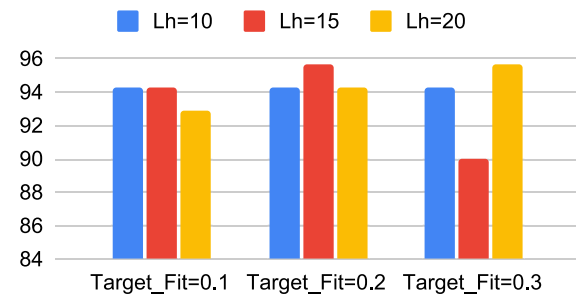
HMCR=0.9 vs Accuracies



HMCR=0.9 vs #Features



HMCR=0.95 vs Accuracies



HMCR=0.95 vs #Features

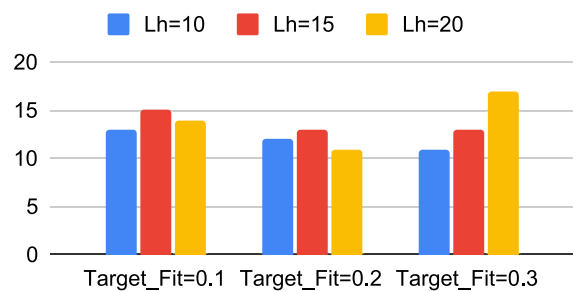


Fig. 4. Results obtained for different sets of hyperparameters on the Ionosphere dataset. Here, HMCR is a component of the HS, whereas History Length (Lh) and Target Fitness (Target_Fit) are components of LAHC.

other methods producing competitive outcomes. Our proposed algorithm has achieved 100% accuracy for the Wine and Zoo datasets, along with other algorithms like EO, GA, GWO, PSO, and HS. For the SpectEW dataset, our algorithm has produced the third highest classification accuracy, outperforming BA and RDA, but being outperformed by GA, PSO, and HS.

4.4.2. Results in terms of feature dimension reduction

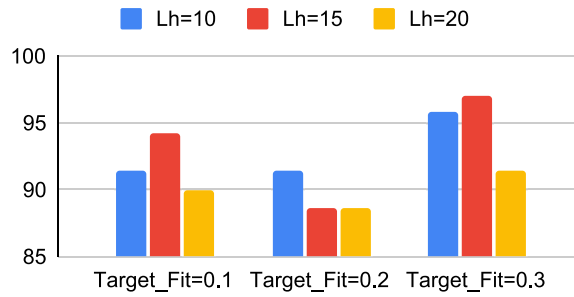
In Table 6, we observe the number of selected features for each of the algorithms. We observe the proposed method gives competitive reduction in number of features as compared to other algorithms. It is noteworthy that, for most of the datasets the proposed method may not give the most reduced number of features but gives competitive reduction of features at the same time maintaining an edge in accuracy. We also note that the difference in number of features is very nominal for all of the datasets and there is no abnormal high features for any of the datasets. One strong possible reason for this behavior may be attributed to the fact that we use chaotic initialization of the search space. This allows our algorithm for strong exploration over the search space, whereas in case of other algorithms the search space is constrained due to random initialization. Secondly, the local search method used here (i.e., LAHC) ensures proper exploitation, thereby making sure

the proposed method is properly exploited to upgrade the solution's quality.

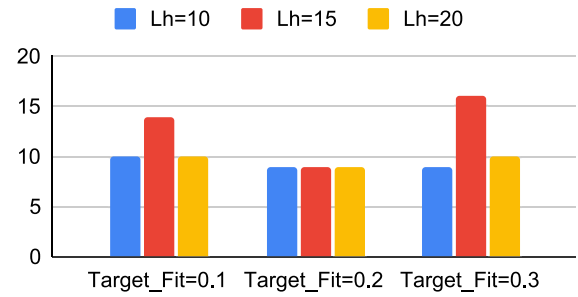
From Table 6 we can conclude that HS+LAHC produces satisfactory results in terms of the number of selected features. Our proposed method secures the best position on the basis of total number of dimensions for CongressEW dataset along with RDA. Other methods: BA, EO, GA, GWO, PSO, RDA, WOA and HS produce satisfactory results. Here MA selects the highest number of features. In the case of M-of-n dataset, HS+LAHC along with PSO, MA and GWO produces second best result by selecting 7 features whereas BA and HS secure the best result. Here, the results from GA, EO and WOA are slightly worse than that of our proposed algorithm.

Apart from that in the case of PenglungEW dataset, our proposed method gives second best outcome in terms of dimensionality. In this case, the result from RDA is the best with 27 features. But remaining methods' performance are extremely poor as compared to our proposed algorithm since all of these methods select features in the range of 100 to 240 with MA securing the last position. HS+LAHC, MA, EO and GWO outperform most of the remaining nature-based wrappers for the Exactly2 dataset by selecting the second least number of features after GA, RDA and WOA. BA and PSO also select competitive number of features. In the case of Lymphography dataset, HS+LAHC attains the

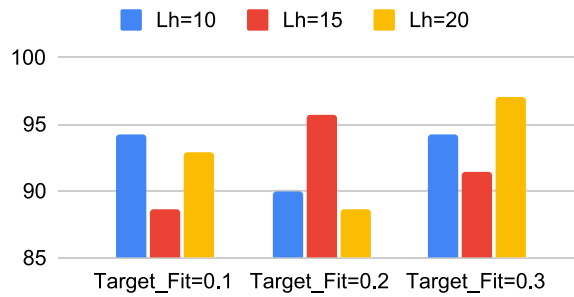
HMCR=0.8 vs Accuracies



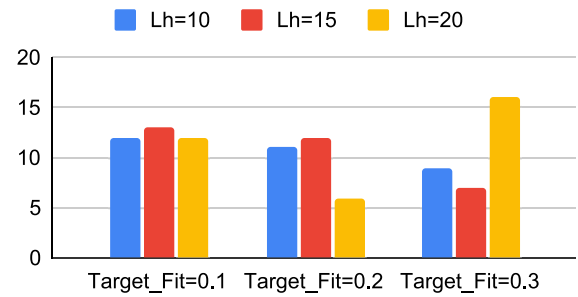
HMCR=0.8 vs #Features



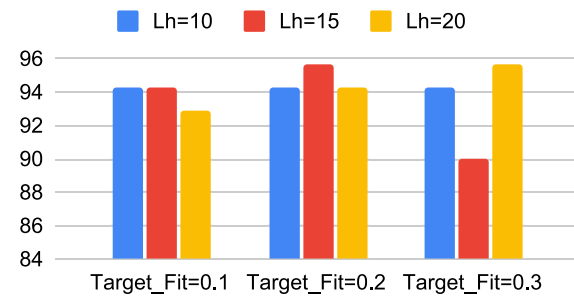
HMCR=0.9 vs Accuracies



HMCR=0.9 vs #Features



HMCR=0.95 vs Accuracies



HMCR=0.95 vs #Features

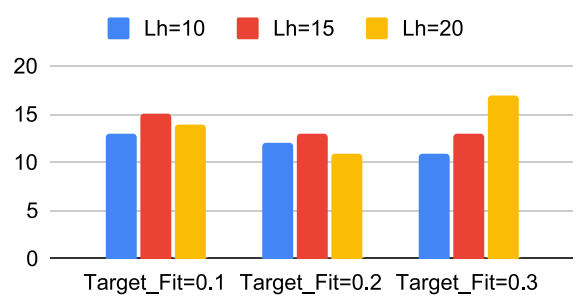


Fig. 5. Results obtained for different sets of hyperparameters on the Sonar dataset. Here, HMCR is a component of the HS, whereas History Length (Lh) and Target Fitness (Target_Fit) are components of LAHC.

Table 5

Performance of HS+LAHC in terms of precision, recall, F1 score, fitness value and execution time.

| Dataset | Precision | Recall | F-1 score | Fitness value | Execution time (min) |
|--------------|-----------|---------|-----------|---------------|----------------------|
| Breastcancer | 0.93529 | 0.93571 | 0.93544 | 97.1 | 45 |
| CongressEW | 0.94315 | 0.94253 | 0.94271 | 94.8 | 40 |
| Exactly | 0.66297 | 0.70000 | 0.65441 | 97.6 | 55 |
| Exactly-2 | 0.70765 | 0.74000 | 0.72027 | 77.5 | 55 |
| BreastEW | 0.87607 | 0.87719 | 0.87633 | 96.1 | 96 |
| Lymphography | 0.81267 | 0.80000 | 0.80278 | 91.9 | 20 |
| PenglungEW | 0.52000 | 0.60000 | 0.56000 | 97.7 | 20 |
| Sonar | 0.68481 | 0.66667 | 0.66972 | 95.3 | 35 |
| Vote | 0.96667 | 0.96667 | 0.96667 | 98.7 | 40 |
| Ionosphere | 0.80702 | 0.97872 | 0.88462 | 92.8 | 66 |
| HeartEW | 0.84515 | 0.82857 | 0.81300 | 87.5 | 15 |
| M-of-n | 0.79000 | 0.80000 | 0.79000 | 96.3 | 60 |
| Wine | 0.95000 | 0.94000 | 0.94000 | 98.0 | 20 |
| Zoo | 0.85000 | 0.85000 | 0.83000 | 97.1 | 9 |
| SpectEW | 0.79000 | 0.80000 | 0.79000 | 85.4 | 96 |

Table 6

Comparison of proposed method with other optimization algorithm in terms of the number of selected features.

| Dataset | HS+LAHC | MA | BA | EO | GA | GWO | HS | PSO | RDA | WOA |
|--------------|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Breastcancer | 2 | 4 | 3 | 4 | 4 | 3 | 2 | 2 | 1 | 3 |
| CongressEW | 1 | 11 | 4 | 3 | 5 | 4 | 5 | 2 | 1 | 4 |
| Exactly | 7 | 8 | 3 | 1 | 2 | 8 | 3 | 7 | 1 | 1 |
| Exactly-2 | 2 | 2 | 3 | 2 | 1 | 2 | 3 | 3 | 1 | 1 |
| BreastEW | 6 | 10 | 9 | 11 | 8 | 11 | 9 | 11 | 1 | 5 |
| Lymphography | 4 | 11 | 5 | 6 | 9 | 7 | 7 | 6 | 3 | 6 |
| PenglungEW | 45 | 236 | 103 | 157 | 140 | 104 | 108 | 120 | 27 | 102 |
| Sonar | 28 | 29 | 22 | 23 | 23 | 30 | 20 | 18 | 11 | 27 |
| Vote | 7 | 11 | 4 | 5 | 3 | 3 | 5 | 4 | 2 | 5 |
| Ionosphere | 12 | 17 | 13 | 14 | 8 | 15 | 13 | 16 | 12 | 12 |
| HeartEW | 6 | 6 | 6 | 7 | 4 | 6 | 6 | 7 | 10 | 6 |
| M-of-n | 7 | 7 | 6 | 8 | 8 | 7 | 6 | 7 | 13 | 8 |
| Wine | 4 | 3 | 5 | 5 | 7 | 6 | 5 | 8 | 5 | 4 |
| Zoo | 5 | 6 | 6 | 6 | 7 | 5 | 5 | 5 | 6 | 7 |
| SpectEW | 10 | 10 | 8 | 8 | 4 | 10 | 8 | 8 | 1 | 8 |

second position after RDA, the results produced by proposed algorithm as well as other nature-based wrappers are quite up to the mark (in between 5 to 9 features selected) except MA selecting the highest number of features.

For Ionosphere dataset, our method stands at second position and the remaining methods also produce competitive results. For Breast-Cancer dataset also, our method and PSO stand at second position after RDA with the other remaining wrappers producing satisfactory results. For HeartEW dataset again, HS+LAHC produces the second least number of dimensions along with MA, BA, GWO, HS and WOA. Other methods also yield competitive results. In BreastEW dataset, our proposed algorithm secures the third position by selecting just 6 features. It can be observed that other algorithms such as BA and GWO have produced considerably good results.

In the case of Exactly dataset, proposed algorithm produces the sixth best result by selecting 7 features along with PSO. Here, on one hand RDA, WOA, BA and HS produce feature sets of relatively similar dimensions but on the other side, GWO and MA have selected most number of features. In the case of Sonar dataset though our proposed algorithm achieves the eighth best result, the number of features selected is very competitive with the other methods such as MA, BA, HS, GWO and GA etc. Here, RDA has selected the least number of features. The proposed method for the Wine dataset requires 4 features, which is the second lowest number of dimensions (the first being the MA, which requires 3 features). For the Zoo dataset, the proposed method outperforms other algorithms. However, for the SpectEW dataset, the proposed method requires the most features along with the GWO algorithm.

4.5. Statistical test

For evaluating the proposed method's discriminative ability statistically, we present a statistical test in Table 7. To perform this test, we collect statistical evidences to decide whether we have enough evidence to decide the given piece of information is significant or not. We hypothesize "The proposed algorithm yields similar results when compared with the existing algorithms". To reject this hypothesis, we conduct Wilcoxon Rank-Sum Test (Ahmed et al., 2022) where we compare each of the algorithms with the proposed algorithm. To calculate the p -value, we consider 15 independent runs on each dataset for each of the algorithms and compare it with the proposed algorithm. We note the classification accuracies of each of the independent run. If we find the p -value < 0.05 (5%), we reject the hypothesis by concluding that we do not have enough evidence to withhold the said hypothesis.

From results in Table 7, we observe that out of 135 cases, we can reject the hypothesis for 85 cases. A statistically significant result implies the stability of the algorithm over different splits of data. A stable algorithm would produce similar results for any of the train-test split. Therefore, a statistical validation of this type would imply the stability of algorithm for different splits of data. Thus we can claim that when compared to the existing algorithms our method performs statistically superior.

5. Comparative analysis of the proposed method

Three types of comparisons are offered in this section. The first one uses various meta-heuristic optimization algorithms aided by LAHC based local search algorithm and our proposed algorithm in terms of the classification accuracy and the number of features selected. Secondly, we compare our proposed method to the state-of-the-art methods to prove the effectiveness of our algorithm. The final comparison is performed when HS algorithm is combined with LAHC (proposed) and other local search algorithms separately to show the HS+LAHC combination is the best one among others.

5.1. Comparison of HS+LAHC with various optimization algorithms aided by LAHC

From the results given in Table 8, we can clearly see that our method outperforms various optimization algorithms hybridized with LAHC in terms of classification accuracy. Our method performs best on 10 out of 15 datasets. Whereas, the results given in Table 9 make it very evident that in terms of the number of features selected, our method outperforms all others, with the exception of the RDA+LAHC.

5.2. Comparative results when HS algorithm is combined with different local search methods

From Tables 10 and 11, it can be observed that though our proposed algorithm has a similar accuracy to HS with adaptive beta hill climbing (HS+ABHC) or with SA (HS+SA), but it outperforms these algorithms in terms of the number of features required to achieve the same level of classification accuracy. For 8 out of 15 datasets, our proposed algorithm has achieved competitive level of classification accuracy with lowest number of features. This proves the effectiveness of the HS+LAHC combination in comparison with other combinations.

5.3. Comparison of HS+LAHC with various state-of-the-art methods

In Table 12, we compare the proposed HS+LAHC algorithm with various widely-used and/or recently proposed optimization algorithms utilized for the purpose of feature selection using UCI datasets. We compare our method with five widely used state-of-the-art metaheuristic algorithms such as binary PSO (BPSO) (Chuang, Chang, Tu, & Yang, 2008), binary Salp Swarm Algorithm (BSSA) (Faris, Mafarja, Heidari, Aljarah, Ala'm, Mirjalili, & Fujita, 2018), Memetic framework with memory-based fuzzy local search (MFFLS) (Chatterjee et al., 2022), Discrete EO combined with SA (DEOSA) (Guha et al., 2023), Sinusoidal Binary Dragonfly Algorithm (SBDA) (Hammouri et al., 2020), the fourth version of Binary Horse herd Optimization Algorithm (BHOA) with S-shaped and one-point crossover (BHOA_S4_Cr1) (Awadallah, Hammouri et al., 2022). Apart from these, we have compared the

Table 7

Comparison of proposed method with other optimization algorithms based on p-values.

| Dataset | GWO | MA | WOA | PSO | RDA | EO | HS | BA | GA |
|--------------|---------|---------|---------|---------|---------|---------|----------|---------|---------|
| Lymphography | 6.1E-05 | 6.1E-05 | 0.00061 | 0.00335 | 6.1E-05 | 0.00115 | 6.1E-05 | 0.00030 | 6.1E-05 |
| BreastCancer | 0.90622 | 0.27685 | 0.25238 | 0.42120 | 0.52447 | 0.59948 | 0.07299 | 0.97796 | 0.00115 |
| BreastEW | 0.84692 | 0.03015 | 0.19509 | 0.45577 | 0.00281 | 0.26355 | 0.00115 | 0.01842 | 6.1E-05 |
| CongressEW | 0.82411 | 0.00030 | 0.12608 | 0.54684 | 0.11138 | 0.19849 | 0.04762 | 0.01048 | 0.21368 |
| Exactly | 0.03272 | 0.00462 | 0.03008 | 0.00285 | 6.1E-05 | 0.01296 | 6.1E-05 | 0.00092 | 6.1E-05 |
| Exactly2 | 0.23366 | 0.03015 | 0.04827 | 0.02660 | 0.01674 | 0.24184 | 0.01647 | 0.01120 | 0.01674 |
| HeartEW | 0.04125 | 0.00305 | 0.00671 | 0.00427 | 6.1E-05 | 0.20776 | 0.000122 | 0.00335 | 0.20776 |
| Ionosphere | 0.56140 | 0.01025 | 0.13537 | 0.84692 | 0.27182 | 0.35913 | 0.00671 | 0.00537 | 0.01507 |
| M-of-n | 0.59611 | 0.04376 | 6.1E-05 | 0.00326 | 0.00095 | 0.00145 | 6.1E-05 | 0.00096 | 0.01670 |
| PenglungEW | 0.65094 | 0.00012 | 0.37868 | 0.87417 | 0.04199 | 0.19381 | 0.00097 | 0.02303 | 0.87384 |
| Sonar | 0.04791 | 0.00836 | 0.71972 | 0.25238 | 0.02154 | 0.02154 | 0.00085 | 0.00061 | 0.06372 |
| Vote | 0.04125 | 0.00418 | 0.02275 | 0.01806 | 0.00998 | 0.00262 | 0.00282 | 0.00018 | 6.1E-05 |
| Wine | 0.73888 | 0.19566 | 0.37953 | 0.24821 | 0.04611 | 0.76302 | 0.00324 | 0.01969 | 0.00091 |
| Zoo | 0.41421 | 0.56370 | 1.00000 | 0.04550 | 0.10247 | 0.17971 | 0.00276 | 0.00504 | 0.15729 |
| SpectEW | 0.03267 | 0.30331 | 0.82332 | 0.03775 | 0.04546 | 0.05856 | 0.00379 | 0.00262 | 0.18036 |

Table 8

Comparison of the proposed method with other optimization algorithms hybridized with LAHC in terms of classification accuracy. Accuracy values are in %.

| Dataset | HS+LAHC | BA+LAHC | MA+LAHC | EO+LAHC | GA+LAHC | GWO+LAHC | PSO+LAHC | RDA+LAHC | WOA+LAHC | HS |
|--------------|--------------|---------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Breastcancer | 98.2 | 96.4 | 97.8 | 97.8 | 97.1 | 99.2 | 99.2 | 97.8 | 98.5 | 93.5 |
| CongressEW | 97.7 | 96.5 | 98.8 | 97.7 | 100.0 | 98.8 | 98.8 | 98.8 | 98.8 | 94.2 |
| Exactly | 100.0 | 68.0 | 99.5 | 98.5 | 80.0 | 100.0 | 100.0 | 99.0 | 100.0 | 69.0 |
| Exactly-2 | 79.0 | 77.5 | 79.0 | 76.0 | 77.5 | 78.5 | 78.5 | 76.0 | 76.0 | 76.0 |
| BreastEW | 98.2 | 95.6 | 95.6 | 96.4 | 95.6 | 96.4 | 95.6 | 94.7 | 95.6 | 93.8 |
| Lymphography | 93.3 | 90.0 | 96.6 | 96.6 | 96.6 | 93.3 | 93.3 | 96.6 | 93.3 | 80.0 |
| M-of-n | 100.0 | 84.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 80.0 |
| PenglungEW | 100.0 | 73.3 | 86.6 | 86.6 | 86.6 | 93.3 | 100.0 | 84.8 | 93.3 | 93.3 |
| Sonar | 95.2 | 90.4 | 92.8 | 92.8 | 90.4 | 90.4 | 95.2 | 90.4 | 95.2 | 85.7 |
| Vote | 100.0 | 95.0 | 100.0 | 96.6 | 98.3 | 98.3 | 98.3 | 98.3 | 98.3 | 96.6 |
| HeartEW | 88.8 | 83.3 | 90.7 | 90.7 | 83.3 | 87.0 | 90.7 | 88.8 | 92.5 | 81.4 |
| Ionosphere | 95.7 | 92.8 | 90.0 | 92.8 | 94.2 | 91.86 | 92.8 | 92.8 | 92.8 | 92.8 |
| Wine | 100.0 | 97.2 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 94.4 | 97.22 | 100.0 |
| Zoo | 100.0 | 90.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| SpectEW | 87.0 | 83.3 | 92.5 | 94.4 | 87.0 | 87.0 | 90.7 | 90.7 | 92.5 | 87.0 |

Table 9

Comparison of the proposed method with other optimization algorithms hybridized with LAHC in terms of the number of features selected.

| Dataset | HS+LAHC | MA+LAHC | BA+LAHC | EO+LAHC | GA+LAHC | GWO+LAHC | PSO+LAHC | RDA+LAHC | WOA+LAHC |
|--------------|---------|---------|---------|---------|---------|----------|----------|----------|----------|
| Breastcancer | 2 | 3 | 3 | 3 | 4 | 3 | 5 | 1 | 2 |
| CongressEW | 1 | 11 | 4 | 3 | 5 | 4 | 2 | 1 | 4 |
| Exactly | 7 | 7 | 3 | 5 | 2 | 8 | 7 | 7 | 1 |
| Exactly-2 | 2 | 2 | 3 | 2 | 1 | 2 | 3 | 1 | 1 |
| BreastEW | 6 | 10 | 9 | 11 | 8 | 11 | 11 | 1 | 5 |
| Lymphography | 4 | 11 | 5 | 8 | 9 | 11 | 7 | 8 | 12 |
| PenglungEW | 45 | 236 | 103 | 157 | 140 | 104 | 120 | 149 | 102 |
| Sonar | 28 | 29 | 22 | 23 | 23 | 30 | 18 | 11 | 27 |
| Vote | 7 | 7 | 4 | 5 | 3 | 3 | 4 | 2 | 5 |
| Ionosphere | 12 | 17 | 13 | 14 | 8 | 14 | 16 | 12 | 12 |
| HeartEW | 6 | 5 | 6 | 6 | 4 | 6 | 2 | 7 | 4 |
| M-of-n | 7 | 7 | 6 | 7 | 8 | 7 | 7 | 7 | 7 |
| Wine | 4 | 3 | 5 | 5 | 8 | 5 | 8 | 5 | 4 |
| Zoo | 5 | 6 | 6 | 6 | 7 | 5 | 8 | 6 | 7 |
| SpectEW | 10 | 8 | 8 | 12 | 4 | 10 | 12 | 10 | 8 |

performance of the proposed method to one of the newly published metaheuristic algorithms, dubbed Binary Gaining Sharing Knowledge (BGSK) (Agrawal, Ganesh, Oliva, & Mohamed, 2022) based optimization, in order to further illustrate the usefulness of HS+LAHC. PBGSK is the name given to the improved version. This method has become more well-known as a result of its adaptability and effectiveness in solving optimization issues. We have here compared our proposed method with two popular variants of PBGSK.

6. Discussion

The convergence graphs for the proposed hybrid feature selection algorithm are presented in Fig. 6. In each case, we see the proposed algorithm converges with progression in iteration. It should be again noted that the mean fitness of candidate solutions is considered here

to plot the graphs. The increase in this value showcases the ability of the algorithm to converge for all candidate solutions. This has been possible as we use a memory based local optimization technique. The use of such an algorithm ensures proper exploitation of the search space which in turn stabilizes the entire framework. It should be noted that the proposed framework additionally uses chaotic initialization, which in turn lets the algorithm explore the search space optimally.

The convergence curves related to other metaheuristic optimization algorithms considered in this work for the comparison purpose are shown in Fig. 7. In most of the cases we see that the average fitness converges to an optimum with progression in iteration. It is noteworthy that when we carefully observe the figures presented, we see most of the algorithms in the initial iterations converge to their optimal solution which remains unchanged during the later iterations. Whereas in case of the proposed HS+LAHC algorithm, the algorithm converges to its

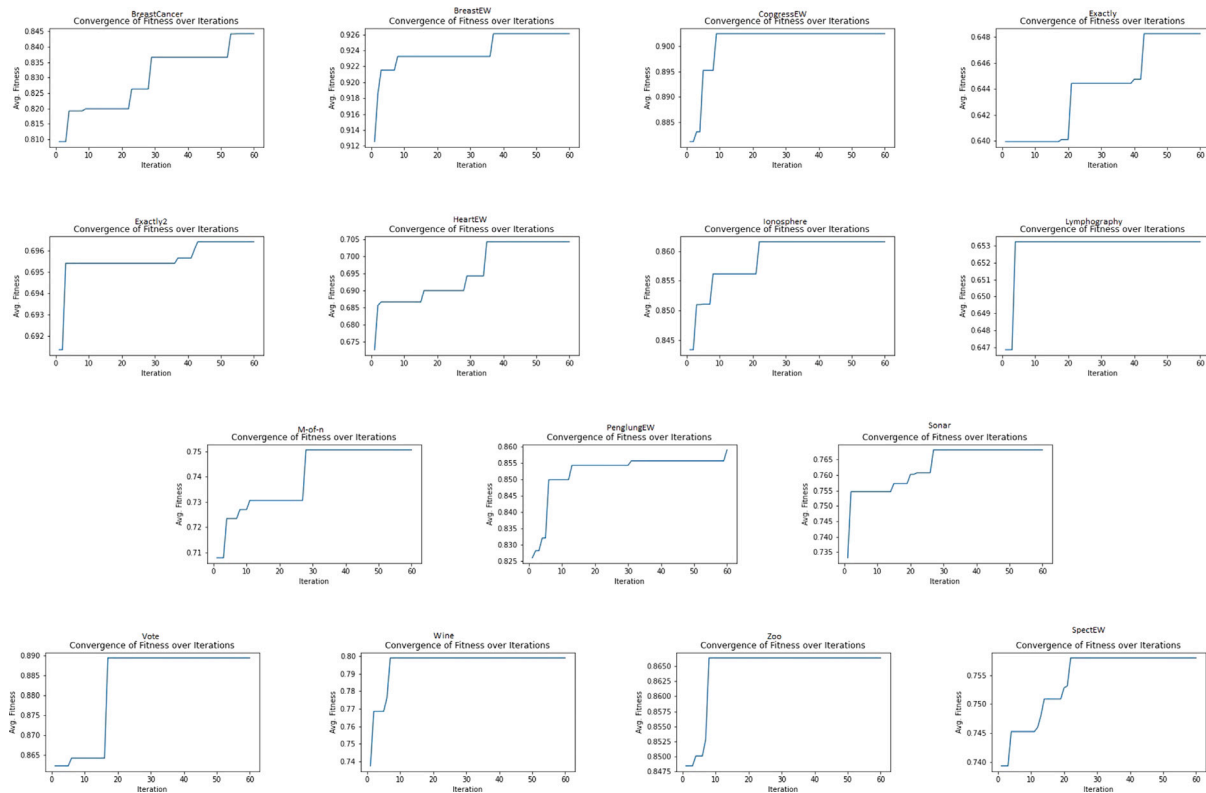


Fig. 6. The convergence graphs concerning all datasets using the proposed HS+LAHC algorithm. The x-axis corresponds to the progression in iteration, whereas the y-axis corresponds to the mean of the fitness value taking into account all the agents.

Table 10

Comparison among HS+ABHC, HS+SA and HS+LAHC in terms of classification accuracy.

| Dataset | HS+ABHC | HS+SA | HS+LAHC |
|--------------|--------------|--------------|--------------|
| Breastcancer | 97.8 | 97.8 | 98.2 |
| CongressEW | 97.7 | 97.7 | 97.7 |
| Exactly | 100.0 | 100.0 | 100.0 |
| Exactly-2 | 76.0 | 80.0 | 79.0 |
| BreastEW | 94.7 | 98.2 | 98.2 |
| Lymphography | 93.3 | 90.0 | 93.3 |
| PenglungEW | 80.0 | 100.0 | 100.0 |
| Sonar | 88.0 | 92.8 | 95.2 |
| Vote | 95.0 | 96.6 | 100.0 |
| Ionosphere | 94.2 | 95.7 | 95.7 |
| HeartEW | 88.8 | 87.0 | 88.8 |
| M-of-n | 100.0 | 100.0 | 100.0 |
| Wine | 100.0 | 100.0 | 100.0 |
| Zoo | 100.0 | 100.0 | 100.0 |
| SpectEW | 88.8 | 90.7 | 87.03 |

Table 11

Comparison among HS+ABHC, HS+SA and HS+LAHC in terms of the number of features selected.

| Dataset | HS+ABHC | HS+SA | HS+LAHC |
|--------------|---------|-------|---------|
| Breastcancer | 4 | 3 | 2 |
| CongressEW | 8 | 6 | 1 |
| Exactly | 6 | 6 | 7 |
| Exactly-2 | 1 | 8 | 2 |
| BreastEW | 11 | 10 | 6 |
| Lymphography | 8 | 5 | 4 |
| PenglungEW | 147 | 124 | 45 |
| Sonar | 26 | 26 | 28 |
| Vote | 4 | 6 | 7 |
| Ionosphere | 16 | 9 | 12 |
| HeartEW | 3 | 4 | 6 |
| M-of-n | 7 | 6 | 7 |
| Wine | 5 | 5 | 4 |
| Zoo | 7 | 6 | 5 |
| SpectEW | 11 | 8 | 10 |

optimal solution with progression in iteration sequentially for most of the datasets. We also observe that for certain datasets like Exactly, Ionosphere and M-of-n, some algorithms converge to their optimal solution in the initial iteration itself and thereby not improvising its solution over the course of iterations. We should specifically note the performance of HS which is our base algorithm on which the proposed method is improvised upon. We observe HS converges to optimal solution initially and does not improvise its solution. In this scenario we can claim that we have gained a significant advantage of using chaos guided initialization and the LAHC based local search to improve the overall solution.

In Fig. 8, we present the box-plots regarding the classification accuracies of 15 independent runs for each of the optimization algorithms. The yellow line in the plot shows the median, whereas the top and bottom ends of the box show the second and first quartile of the

distribution respectively. In most of the cases, we observe that the size of the box for the proposed algorithm is relatively smaller than the competitor algorithms. This showcases the ability of the proposed algorithm to produce similar results for different splits of data. Also we note that the distribution is not very much skewed in one end expect a very few cases which is depicted by the relative position of yellow line w.r.t the box formed. In addition to that we observe the less number of outliers. All these account to the stability of each of these algorithms. This is possible because our method maintains the crucial trade-off between exploration and exploitation which in turn helps the algorithm to produce stable results. It is also noteworthy that the use of local search can only improvise the results and cannot degrade the original results, thereby justifying the use of a local search mechanism in the HS algorithm for achieving better and stable results.

Table 12

Comparison of the proposed method with some state-of-the-art methods in terms of classification accuracy and the number of features. Accuracy values are in %.

| Dataset | HS+LAHC | DEOSA | PBGSK_S1 | PBGSK_S2 | MFFLS | BPSO | BSSA | BHOA_S4_Cr1 | SBDA |
|--------------|------------------|------------------|------------|-------------|------------------|-----------------|-----------------|-------------------|---------------------|
| Breastcancer | 98.2(2) | 96.4(3) | 97.4(4.8) | 97.4(4.8) | 98.0(5) | 96.2(4) | 97.6(4) | 96.5(3.1) | 99.3(5) |
| CongressEW | 97.7(1) | 96.5(5) | 96.2(4.4) | 96.4(5) | 97.7(7) | 96.3(3) | 96.2(6) | 97.0(5.1) | 97.5(5.4) |
| Exactly | 100.0(7) | 100.0(7) | 91.8(7) | 95.6(6.4) | 100.0(6) | 100.0(6) | 98.0(7) | 100.0(5.9) | 100.0(6.1) |
| Exactly-2 | 79.0(2) | 76.0(1) | 76.8(4) | 76.4(2.8) | 78.5(7) | 76.8(1) | 75.8(3) | 74.9(4) | 75.7(5) |
| BreastEW | 98.2(6) | 96.4(8) | 95.1(12.4) | 95.1(11.2) | 98.0(4) | 97.1(9) | 94.8(16) | 97.0(16) | 97.5(12.2) |
| Lymphography | 93.3 (4) | 96.6(9) | 84.0(8.2) | 84.0(7.8) | 96.6(5) | 94.2(5) | 89.0(10) | 86.4(6.6) | 95.4(6.8) |
| M-of-n | 100.0(7) | 100.0(6) | 99.2(6.6) | 97.4(7) | 100.0(3) | 88.8(6) | 99.1(7) | 100.0(6) | 100.0(6) |
| PenglungEW | 100.0(45) | 100.0(86) | 82.5(149) | 82.5(159.4) | 100.0(33) | 96.3(130) | 87.7(171) | 89.1(157.6) | 100.0(117.5) |
| Sonar | 95.2(28) | 97.6(25) | 83.2(25.4) | 81.7(24.6) | 100.0(13) | 94.8(22) | 93.7(31) | 92.5(32.1) | 99.3(24.3) |
| Vote | 100.0(7) | 96.6(1) | 96.2(4.8) | 96.1(7) | 100.0(3) | 96.0(3) | 95.1(5) | 97.0(4) | 97.2(4) |
| HeartEW | 88.8(6) | 90.7(5) | 85.1(5.6) | 85.7(5.2) | 85.1(7) | 83.7(3) | 86.0(6) | 85.9(7) | 86.7(6) |
| Ionosphere | 95.7(12) | 95.7(7) | 88.8(12.4) | 87.6(12) | 95.0(9) | 94.8(7) | 91.8(16) | 93.4(12) | 98.4(12.6) |
| Wine | 100.0(4) | 100.0(3) | 94.3(5.2) | 94.3(5.2) | 100.0(2) | 97.7(5) | 99.3(6) | 96.6(3.4) | 100.0(4.4) |
| Zoo | 100.0(5) | 100.0(4) | 93.3(5.8) | 93.7(6.2) | 100.0(2) | 96.0(5) | 100.0(7) | 100.0(5.8) | 100.0 (1.9) |
| SpectEW | 87.0(10) | 90.7(9) | 87.1(9.6) | 87.01(10) | 90.7(8) | 88.8(6) | 83.6(11) | 87.3(10.8) | 92.5(8.5) |

Table 13

Description of microarray datasets used for experimentation.

| Dataset | #Samples | #Attributes | #Classes | Use |
|-----------|----------|-------------|----------|---|
| DLBCL | 77 | 7070 | 2 | Diffuse large B-cell lymphoma detection |
| SRBCT | 83 | 2309 | 4 | Small-round-blue-cell tumor analysis |
| Leukaemia | 72 | 5148 | 2 | Leukaemia Detection |

Table 14

Comparison of proposed method with other optimization algorithms based on classification accuracy. Accuracy is measured in %.

| Dataset | HS+LAHC | BA | MA | EO | GA | GWO | PSO | RDA | WOA | HS |
|-----------|--------------|--------------|--------------|------------|--------------|--------------|--------------|--------------|--------------|--------------|
| DLBCL | 100.0 | 93.75 | 100.0 | 93.75 | 93.75 | 93.75 | 100 | 93.75 | 93.75 | 93.71 |
| SRBCT | 100.0 | 100.0 | 82.3 | 100 | 100.0 | 100.0 | 100.0 | 94.1 | 100.0 | 100.0 |
| Leukaemia | 100.0 | 100.0 | 93.3 | 93.3 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 86.6 |

Table 15

Comparison of proposed method with other optimization algorithms based on feature dimension.

| Dataset | HS+LAHC | BA | MA | EO | GA | GWO | PSO | RDA | WOA | HS |
|-----------|---------|-----|-----|-----|-----|-----|-----|-----------|-----------|-----|
| DLBCL | 144 | 92 | 159 | 139 | 95 | 185 | 149 | 51 | 150 | 91 |
| SRBCT | 138 | 92 | 210 | 140 | 124 | 173 | 104 | 122 | 75 | 151 |
| Leukaemia | 126 | 126 | 219 | 138 | 128 | 168 | 117 | 10 | 89 | 170 |

7. Results on microarray datasets

In the preceding texts, we observe the usefulness of the proposed feature selection method over various optimization algorithms on standard UCI datasets. In this section, we determine the applicability of the proposed method on very high-dimensional datasets. For this, we consider microarray-based gene expression datasets. Research in microarray data analysis helps discoveries in various sectors like pharmacology, medicine and biology. These datasets are used for cancer prediction from the available gene expression information. Table 13 presents a brief description of the microarray datasets used here for experimentation. In the DLBCL dataset, 32 samples are extracted from people who were cured while the rest were from people who were suffering from this disease. For the SRBCT dataset, 29 cases are of Ewing sarcoma (EWS), 11 cases are of Burkitt lymphoma (BL), 18 cases are of neuroblastoma (NB) and 25 cases are of rhabdomyosarcoma (RMS). In the case of the Leukemia dataset, 47 cases are of acute lymphoblastic leukaemia (ALL) and the rest cases are of acute myeloid leukemia (AML). It should be noted that in these datasets, the number of features is very high while the number of instances is quite low. Obtained classification accuracies and the number of selected features by the present hybrid feature selection method are presented in Tables 14 & 15 respectively. The results provided in these tables prove the usefulness of the proposed method in selecting relevant features with high classification accuracies from very high dimensional datasets also.

Traditionally gene-expression has been a matter of great interest of computer scientists due to the “curse of dimensionality” with these many features. Considering that the low samples and huge number of features microarray datasets inherently poses an additional challenge during feature selection. As we discuss the previous section, considering all $2^N - 1$ feature subsets is never possible, where N is the number of features. Such a diverse search space requires the redundant genes to be left out for efficient and effective diagnosis. These datasets, in comparison with the standard UCI datasets, contain much more irrelevant features. Therefore, removal of these features becomes an important task. In microarray data analysis, the main objective is to capture the inter-dependencies amongst the genes to capture the association of genes for the presence of a particular type of cancer. To be specific, there are a few biomarkers present in these datasets those are actually useful for cancer prediction. From the above mentioned tables and discussion, we can conclude our hybrid feature selection method works well on the high dimensional data also.

8. Experiment on COVID-19 dataset

COVID-19 is a highly contagious disease caused by the SARS-CoV-2 virus. Common symptoms include fever, cough, fatigue, and difficulty breathing. The accurate detection of the disease is challenging and several screening techniques have been developed. Many people have been infected worldwide, and there have been several deaths due to this virus. Detecting COVID-19 and keeping infected individuals in quarantine is a top priority for every country. The COVID-19 dataset used here is publicly available at <https://github.com/Atharva-Peshkar/Covid-19-Patient-Health-Analytics>, and it is in the csv format. The dataset contains information about 1086 patients with 74 attributes. The results obtained using the proposed FS method have been compared with other meta-heuristics-based FS methods. In Table 16, we have shown the said comparison in terms of the classification accuracy and the number of selected features (shown in the bracket).

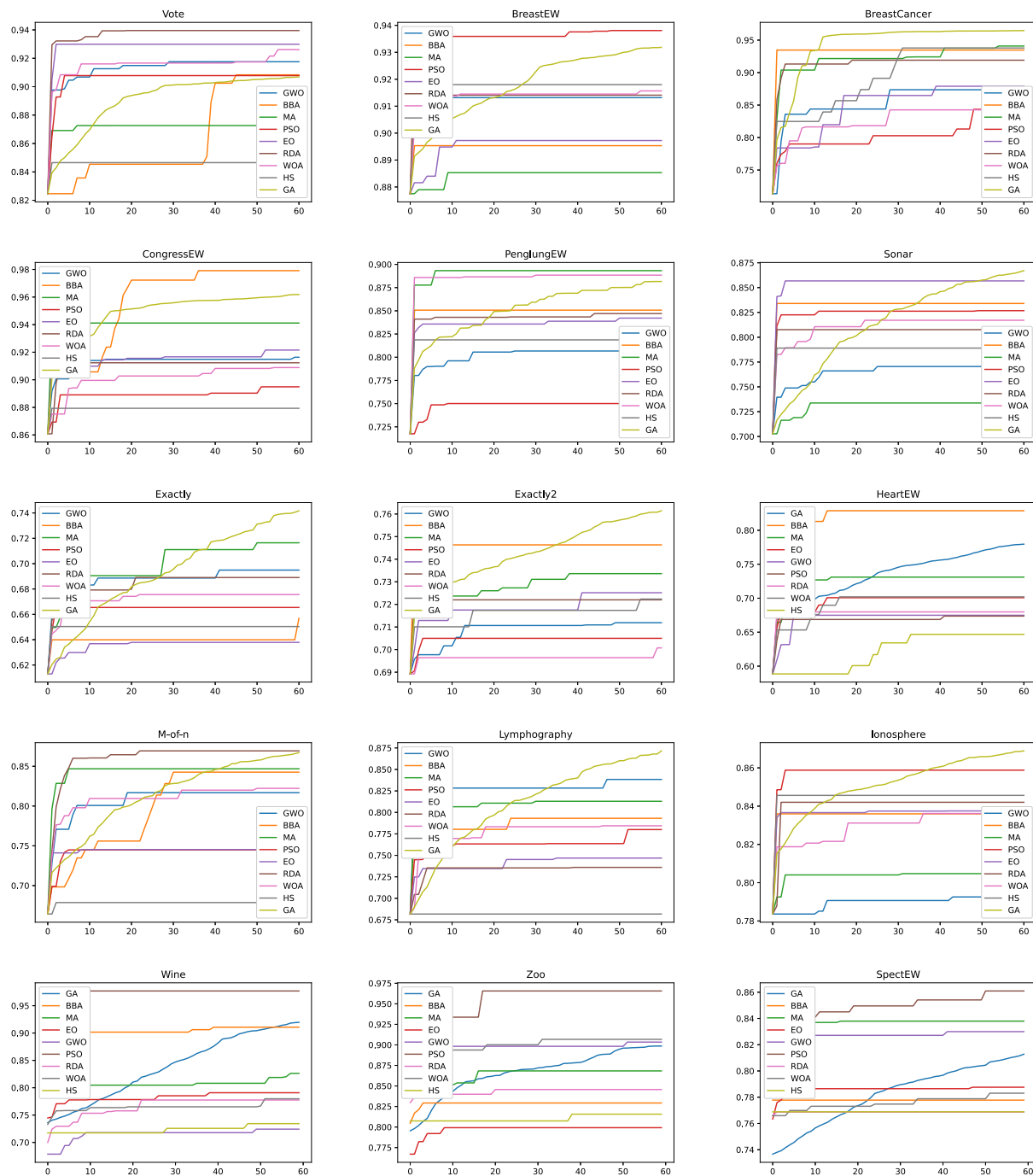


Fig. 7. The convergence graphs concerning all datasets using other optimization algorithms. The x-axis corresponds to the progression in iteration, whereas the y-axis corresponds to the mean of the fitness value taking into account all the agents.

Table 16

Results on the COVID-19 dataset. Values in the bracket indicate the number of selected features.

| Dataset | HS+LAHC | BA | MA | EO | GA | GWO | PSO | RDA | WOA | HS |
|----------|-----------------|----------|----------|----------|-----------------|----------|----------|----------|----------|----------|
| Covid 19 | 98.1(23) | 96.3(27) | 94.9(59) | 95.8(29) | 98.1(30) | 96.3(36) | 97.6(22) | 96.7(10) | 96.7(20) | 97.6(20) |

9. Conclusion

This work proposes an improved version of the HS algorithm for feature selection that means this algorithm aims to eliminate the redundancy from the raw datasets. The HS algorithm has been modified with chaos-guided population initialization to improve search space

exploration, and with a local search method called LAHC for better exploitation. The proposed method has been tested on various UCI datasets along with a COVID-19 dataset, and Microarray-based gene expression datasets used for cancer gene/biomarker identification. The results show improved performance compared to state-of-the-art metaheuristic-based feature selection algorithms. The results have been

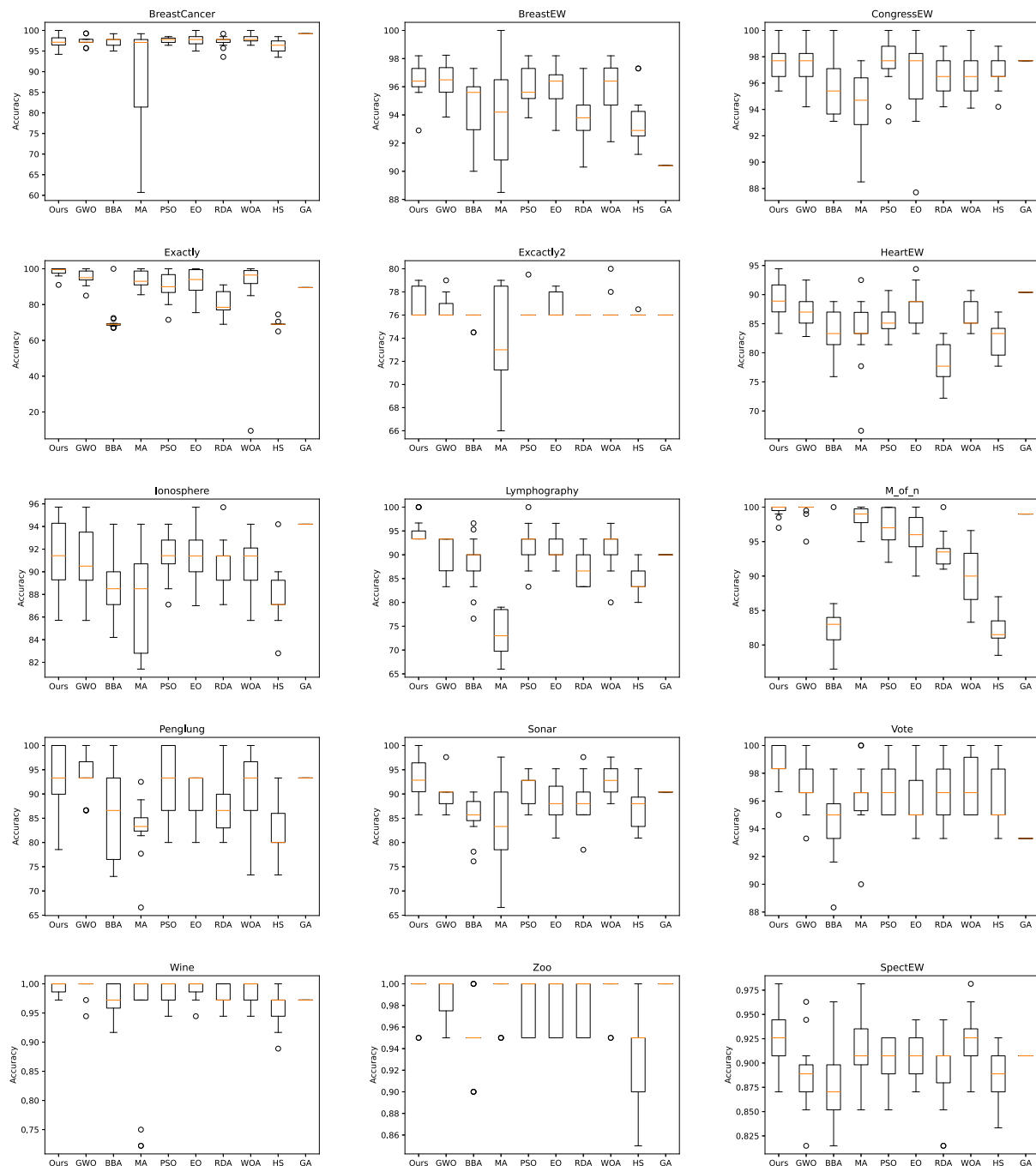


Fig. 8. Box-plots regarding various feature selection algorithms including the proposed one. The box-plots are drawn considering classification accuracies of 15 independent runs on each dataset. The y-axis depicts the accuracy whereas the x-axis depicts the optimization algorithm.

analyzed using sensitivity, convergence, and boxplot to gain deeper insights on the behavior of the algorithm. However, the local search method increases average training time, which can be addressed in the future with better dynamic exploration and exploitation strategies. One possible extension could be to use statistical correlation metrics to select the hyperparameters for the local search. We plan to improve the algorithm by incorporating mutation strategies to enhance solution reliability. The method has potential applications in fields such as bioinformatics, data mining, and image processing, where in many applications, we commonly find the existence of high dimensional feature vectors.

CRediT authorship contribution statement

Anurup Naskar: Conceptualization, Methodology, Software, Investigation, Validation, Writing – original draft, Writing – review & editing. **Rishav Pramanik:** Conceptualization, Methodology, Software, Investigation, Validation, Writing – original draft, Writing – review & editing. **S.K. Sabbir Hossain:** Conceptualization, Methodology, Software, Investigation, Writing – original draft, Writing – review & editing. **Seyedali Mirjalili:** Conceptualization, Methodology, Investigation Visualization, Resources, Writing – review & editing, Supervision, Project administration. **Ram Sarkar:** Conceptualization, Methodology,

Investigation, Visualization, Resources, Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to thank the Center for Microprocessor Applications for Training Education and Research (CMATER) research laboratory of the Department of Computer Science and Engineering, Jadavpur University, Kolkata, India for providing infrastructural support to this work.

References

- Agrawal, P., Ganesh, T., Oliva, D., & Mohamed, A. W. (2022). S-shaped and v-shaped gaining-sharing knowledge-based algorithm for feature selection. *Applied Intelligence*, 1–32.
- Ahmed, S., Ghosh, K. K., Singh, P. K., Geem, Z. W., & Sarkar, R. (2020). Hybrid of harmony search algorithm and ring theory-based evolutionary algorithm for feature selection. *IEEE Access*, 8, 102629–102645.
- Ahmed, S., Sheikh, K. H., Mirjalili, S., & Sarkar, R. (2022). Binary simulated normal distribution optimizer for feature selection: Theory and application in COVID-19 datasets. *Expert Systems with Applications*, 200, Article 116834.
- Akbal, E., Barua, P. D., Dogan, S., Tuncer, T., & Acharya, U. R. (2022). DesPatNet25: Data encryption standard cipher model for accurate automated construction site monitoring with sound signals. *Expert Systems with Applications*, 193, Article 116447.
- Al-Betar, M. A., Aljarah, I., Awadallah, M. A., Faris, H., & Mirjalili, S. (2019). Adaptive β -hill climbing for optimization. *Soft Computing*, 23(24), 13489–13512.
- Awadallah, M. A., Al-Betar, M. A., Braik, M. S., Hammouri, A. I., Doush, I. A., & Zitar, R. A. (2022). An enhanced binary rat swarm optimizer based on local-best concepts of PSO and collaborative crossover operators for feature selection. *Computers in Biology and Medicine*, 147, Article 105675.
- Awadallah, M. A., Hammouri, A. I., Al-Betar, M. A., Braik, M. S., & Abd Elaziz, M. (2022). Binary horse herd optimization algorithm with crossover operators for feature selection. *Computers in Biology and Medicine*, 141, Article 105152.
- Bae, J. H., Kim, M., Lim, J., & Geem, Z. W. (2021). Feature selection for colon cancer detection using k-means clustering and modified harmony search algorithm. *Mathematics*, 9(5), 570.
- Bavkar, S., Iyer, B., & Deosarkar, S. (2021). Optimal EEG channels selection for alcoholism screening using EMD domain statistical features and harmony search algorithm. *Biocybernetics and Biomedical Engineering*, 41(1), 83–96.
- Bhattacharyya, T., Chatterjee, B., Singh, P. K., Yoon, J. H., Geem, Z. W., & Sarkar, R. (2020). Mayfly in harmony: A new hybrid meta-heuristic feature selection algorithm. *IEEE Access*, 8, 195929–195945.
- Burke, E. K., & Bykov, Y. (2017). The late acceptance hill-climbing heuristic. *European Journal of Operational Research*, 258(1), 70–78.
- Chatterjee, B., Ahmed, S., Bhattacharyya, T., & Sarkar, R. (2022). 2LS: Memetic framework with memory based fuzzy local search. *Expert Systems with Applications*, 209, Article 118293.
- Chuang, L.-Y., Chang, H.-W., Tu, C.-J., & Yang, C.-H. (2008). Improved binary PSO for feature selection using gene expression data. *Computational Biology and Chemistry*, 32, 29–37. <http://dx.doi.org/10.1016/j.compbiolchem.2007.09.005>.
- Cui, X., Li, Y., Fan, J., Wang, T., & Zheng, Y. (2020). A hybrid improved dragonfly algorithm for feature selection. *IEEE Access*, 8, 155619–155629.
- Das, S., Singh, P. K., Bhowmik, S., Sarkar, R., & Nasipuri, M. (2016). A harmony search based wrapper feature selection method for holistic bangla word recognition. *Procedia Computer Science*, 89, 395–403.
- Dash, R. (2021). An adaptive harmony search approach for gene selection and classification of high dimensional medical data. *Journal of King Saud University-Computer and Information Sciences*, 33(2), 195–207.
- Dogan, S., Barua, P. D., Kutlu, H., Baygin, M., Fujita, H., Tuncer, T., et al. (2022). Automated accurate fire detection system using ensemble pretrained residual network. *Expert Systems with Applications*, 203, Article 117407.
- Dokeroglu, T., Deniz, A., & Kiziloz, H. E. (2022). A comprehensive survey on recent metaheuristics for feature selection. *Neurocomputing*.
- Eberhart, R., & Kennedy, J. (1995). Particle swarm optimization. In *Proceedings of the IEEE international conference on neural networks*, Vol. 4 (pp. 1942–1948). Citeseer.
- Faramarzi, A., Heidarinejad, M., Stephens, B., & Mirjalili, S. (2020). Equilibrium optimizer: A novel optimization algorithm. *Knowledge-Based Systems*, 191, Article 105190.
- Faris, H., Mafarja, M. M., Heidari, A. A., Aljarah, I., Ala'm, A.-Z., Mirjalili, S., et al. (2018). An efficient binary salp swarm algorithm with crossover scheme for feature selection problems. *Knowledge-Based Systems*, 154, 43–67.
- Geem, Z. W., Kim, J. H., & Loganathan, G. V. (2001). A new heuristic optimization algorithm: harmony search. *Simulation*, 76(2), 60–68.
- Gholami, J., Pourpanah, F., & Wang, X. (2020). Feature selection based on improved binary global harmony search for data classification. *Applied Soft Computing*, 93, Article 106402.
- Ghosh, K. K., Begum, S., Sardar, A., Adhikary, S., Ghosh, M., Kumar, M., et al. (2021). Theoretical and empirical analysis of filter ranking methods: Experimental study on benchmark DNA microarray data. *Expert Systems with Applications*, 169, Article 114485.
- Ghosh, M., Guha, R., Sarkar, R., & Abraham, A. (2020). A wrapper-filter feature selection technique based on ant colony optimization. *Neural Computing and Applications*, 32(12), 7839–7857.
- Ghosh, K. K., Singh, P. K., Hong, J., Geem, Z. W., & Sarkar, R. (2020). Binary social mimic optimization algorithm with X-shaped transfer function for feature selection. *IEEE Access*, 8, 97890–97906.
- Guha, S., Das, A., Singh, P. K., Ahmadian, A., Senu, N., & Sarkar, R. (2020). Hybrid feature selection method based on harmony search and naked mole-rat algorithms for spoken language identification from audio signals. *IEEE Access*, 8, 182868–182887.
- Guha, R., Ghosh, K. K., Bera, S. K., Sarkar, R., & Mirjalili, S. (2023). Discrete equilibrium optimizer combined with simulated annealing for feature selection. *Journal of Computer Science*, Article 101942.
- Guha, R., Ghosh, M., Chakrabarti, A., Sarkar, R., & Mirjalili, S. (2020). Introducing clustering based population in binary gravitational search algorithm for feature selection. *Applied Soft Computing*, 93, Article 106341.
- Guha, R., Ghosh, M., Mutsuddi, S., Sarkar, R., & Mirjalili, S. (2020). Embedded chaotic whale survival algorithm for filter-wrapper feature selection. *Soft Computing*, 24(17), 12821–12843.
- Hammouri, A. I., Mafarja, M., Al-Betar, M. A., Awadallah, M. A., & Abu-Doush, I. (2020). An improved dragonfly algorithm for feature selection. *Knowledge-Based Systems*, 203, Article 106131.
- Holland, J. H. (1992). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT Press.
- Hossain, S., Mukhopadhyay, S., Ray, B., Ghosal, S. K., & Sarkar, R. (2022). A secured image steganography method based on ballot transform and genetic algorithm. *Multimedia Tools and Applications*, 1–30.
- Huang, Y.-F., & Chen, P.-H. (2020). Fake news detection using an ensemble learning model based on self-adaptive harmony search algorithms. *Expert Systems with Applications*, 159, Article 113584.
- Inbarani, H. H., Bagyamathi, M., & Azar, A. T. (2015). A novel hybrid feature selection method based on rough set and improved harmony search. *Neural Computing and Applications*, 26(8), 1859–1880.
- Karadal, C. H., Kaya, M. C., Tuncer, T., Dogan, S., & Acharya, U. R. (2021). Automated classification of remote sensing images using multileveled MobileNetV2 and DWT techniques. *Expert Systems with Applications*, 185, Article 115659.
- Kashan, A. H. (2015). A new metaheuristic for optimization: optics inspired optimization (OIO). *Computers & Operations Research*, 55, 99–125.
- Kirkpatrick, S., Gelatt, C. D., Jr., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598), 671–680.
- Kundu, R., Chattopadhyay, S., Cuevas, E., & Sarkar, R. (2022). AltWOA: Altruistic whale optimization algorithm for feature selection on microarray datasets. *Computers in Biology and Medicine*, 144, Article 105349.
- Li, J., & Liu, H. (2017). Challenges of feature selection for big data analytics. *IEEE Intelligent Systems*, 32(2), 9–15.
- Li, Z., Zou, D., & Kong, Z. (2019). A harmony search variant and a useful constraint handling method for the dynamic economic emission dispatch problems considering transmission loss. *Engineering Applications of Artificial Intelligence*, 84, 18–40.
- Malakar, S., Roy, S. D., Das, S., Sen, S., Velásquez, J. D., & Sarkar, R. (2022). Computer based diagnosis of some chronic diseases: A medical journey of the last two decades. *Archives of Computational Methods in Engineering*, 1–43.
- Mirjalili, S. (2015). The ant lion optimizer. *Advances in Engineering Software*, 83, 80–98.
- Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in Engineering Software*, 69, 46–61.
- Moayedikia, A., Ong, K.-L., Boo, Y. L., Yeoh, W. G., & Jensen, R. (2017). Feature selection for high dimensional imbalanced class data using harmony search. *Engineering Applications of Artificial Intelligence*, 57, 38–49.
- Moscato, P., et al. (1989). On evolution, search, optimization, genetic algorithms and martial arts: Towards memetic algorithms. *Caltech Concurrent Computation Program, C3P Report*, 826, 1989.
- Mukhopadhyay, S., Hossain, S., Malakar, S., Cuevas, E., & Sarkar, R. (2022). Image contrast improvement through a metaheuristic scheme. *Soft Computing*.
- Pramanik, R., Biswas, M., Sen, S., de Souza Júnior, L. A., Papa, J. P., & Sarkar, R. (2022). A fuzzy distance-based ensemble of deep models for cervical cancer detection. *Computer Methods and Programs in Biomedicine*, [ISSN: 0169-2607] 219, Article 106776. <http://dx.doi.org/10.1016/j.cmpb.2022.106776>.

- Pramanik, R., Dey, S., Malakar, S., Mirjalili, S., & Sarkar, R. (2022). TOPSIS aided ensemble of CNN models for screening COVID-19 in chest X-ray images. *Scientific Reports*, 12(1), 1–19.
- Pramanik, R., Sarkar, S., & Sarkar, R. (2022). An adaptive and altruistic PSO-based deep feature selection method for pneumonia detection from chest X-rays. *Applied Soft Computing*, [ISSN: 1568-4946] 128, Article 109464. <http://dx.doi.org/10.1016/j.asoc.2022.109464>.
- Radman, A. (2021). Combination of BESO and harmony search for topology optimization of microstructures for materials. *Applied Mathematical Modelling*, 90, 650–661.
- Rajagopal, A., Ramachandran, A., Shankar, K., Khari, M., Jha, S., Lee, Y., et al. (2020). Fine-tuned residual network-based features with latent variable support vector machine-based optimal scene classification model for unmanned aerial vehicles. *IEEE Access*, 8, 118396–118404.
- Saha, S., Ghosh, M., Ghosh, S., Sen, S., Singh, P. K., Geem, Z. W., et al. (2020). Feature selection for facial emotion recognition using cosine similarity-based harmony search algorithm. *Applied Sciences*, 10(8), 2816.
- Sarkar, S. S., Sheikh, K. H., Mahanty, A., Mali, K., Ghosh, A., & Sarkar, R. (2021). A harmony search-based wrapper-filter feature selection approach for microstructural image classification. *Integrating Materials and Manufacturing Innovation*, 10(1), 1–19.
- Shi, B., Wang, Q., Yin, S., Yue, Z., Huai, Y., & Wang, J. (2021). A binary harmony search algorithm as channel selection method for motor imagery-based BCI. *Neurocomputing*, 443, 12–25.
- Storn, R. (1996). On the usage of differential evolution for function optimization. In *Proceedings of North American fuzzy information processing* (pp. 519–523). Ieee.
- Tahani, M., & Babayan, N. (2019). Flow regime algorithm (FRA): a physics-based meta-heuristics algorithm. *Knowledge and Information Systems*, 60(2), 1001–1038.
- Too, J., Mafarja, M., & Mirjalili, S. (2021). Spatial bound whale optimization algorithm: an efficient high-dimensional feature selection approach. *Neural Computing and Applications*, 33(23), 16229–16250.
- Too, J., & Mirjalili, S. (2021). A hyper learning binary dragonfly algorithm for feature selection: A COVID-19 case study. *Knowledge-Based Systems*, 212, Article 106553.
- Tran, B., Xue, B., & Zhang, M. (2018). Variable-length particle swarm optimization for feature selection on high-dimensional classification. *IEEE Transactions on Evolutionary Computation*, 23(3), 473–487.
- Tsakirakis, E., Marinaki, M., Marinakis, Y., & Matsatsinis, N. (2019). A similarity hybrid harmony search algorithm for the team orienteering problem. *Applied Soft Computing*, 80, 776–796.
- Tuncer, T., Dogan, S., & Acharya, U. R. (2021). Automated EEG signal classification using chaotic local binary pattern. *Expert Systems with Applications*, 182, Article 115175.
- Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1), 67–82.
- Yao, R., Zhang, W., Evans, R., Cao, G., Rui, T., Shen, L., et al. (2022). Inequities in health care services caused by the adoption of digital health technologies: scoping review. *Journal of Medical Internet Research*, 24(3), Article e34144.