A Binary Multi-Objective Chimp Optimizer With Dual Archive for Feature Selection in the Healthcare Domain

JAYASHREE PIRI[1], PUSPANJALI MOHAPATRA[1], (Member, IEEE),
MANAS RANJAN PRADHAN[2], (Member, IEEE), BISWARANJAN ACHARYA[3], (Member, IEEE),
AND TAPAS KUMAR PATRA[4], (Member, IEEE)

[1]Department of CSE, International Institute of Information Technology, Bhubaneswar 751029, India
[2]Skyline University College, Sharjah, United Arab Emirates
[3]Kalinga Institute of Industrial Technology Deemed University, Bhubaneswar 751024, India
[4]College of Engineering and Technology, Bhubaneswar 751029, India

Corresponding author: Tapas Kumar Patra (tapas.patra.in@ieee.org)

ABSTRACT Medical datasets frequently include vast feature sets with numerous features that are related to one another. As a result, the curse of dimensionality affects learning from a medical dataset to discover significant characteristics, making it necessary to minimize the feature set. Feature selection (FS) is a major step in classification and also in reducing the dimension. This study attempts a novel Binary Multi-objective Chimp Optimization Algorithm (BMOChOA) with dual archive and k-nearest neighbors (KNN) classifier for mining relevant aspects from medical data. In this research, 12 versions of BMOChOA are implemented based on the group information and types of chaotic functions used. The best Pareto front obtained from suggested BMOChOA variations is compared with three benchmark multi-objective FS methods by taking 14 popular medical datasets of variable dimensions. By analyzing the experimental outputs using four multi-objective performance evaluators, it is found that the proposed FS method is superior in finding the best trade-off between the two objective functions: the number of features and classification performance.

INDEX TERMS Chimp optimization, classification, healthcare, data mining, feature selection, multi-objective.

I. INTRODUCTION

The relevance of the diagnosis is known to all professionals. Timely and effective diagnostic treatment can save patients' lives, so it's necessary to have proper computer aided diagnostic (CAD) systems that doctors can use. In general, classification is the most important aspect of CAD systems. A vast amount of observations, mostly electronic health records (EHRs), are collected in the medical sector [1]. Nowadays, machine learning is becoming more common in healthcare. Because of the complex nature and size of data, the huge quantity of data created by Electronic Data Interchange (EDI) clinical transactions cannot be handled and evaluated by conventional techniques. Classifying medical data sets is a difficult challenge because they frequently contain many attributes and examples. The need for rapid and precise diagnosis is a consequence of the search for

The associate editor coordinating the review of this manuscript and approving it for publication was Xinyu Du.

more correct and faster categorization methods in CAD systems [2]. The classification precision depends greatly on the chosen feature set to allow classification methods to distinguish instances and detect similarities between examples of the same class. Noisy, duplicated, and irrelevant characteristics may also be present in high-dimensional medical data. These properties lead to unfavorable impacts on the learning process and on the efficiency of categorization by widening the search area too much. This idea is often called the "curse of dimensionality". There is, thus, high computation, sophisticated models, and extremely long learning times required for analysis and data mining [1].

The application of FS methods is one of the most effective answers to this problem [3]. FS techniques tend to pick a subset of highly significant features by removing or reducing duplicate and unnecessary characteristics. Therefore, FS techniques can lead to improved data understanding, reduced learning time, and simplified prediction models with potentially improved performance [4]. Because of its vast

search area and the intricate connections between features, FS is a complicated process. Searching techniques for identifying subsets of optimum features in an L-dimensional data set necessitate the inclusion of 2^L subsets from the combination of characteristics, making searching for big L extremely difficult. Thus, the FS issue is characterized as an NP-Hard task [5].

FS methods are classified mostly into wrapper, filter, and embedding techniques [6]. Our emphasis in this study is on wrapper processes. The FS process includes a search process for the quasi-optimal feature subsets and a prediction system for the wrapping technique [7]. Wrapper techniques give superior outcomes than filter approaches, but due to the continuous learning of the classification algorithm, these strategies take longer than the filter techniques [8]. By contrast, filter strategies are based on statistical methods and theories of information and attempt to recognize a near-optimal sub-string of characteristics having the strongest individual association with the result and the lowest inner correlation [6]. Furthermore, embedded methods are aimed at integrating the FS phase into the training phase of the classification [9]. Use of the wrapper techniques such as greedy search alternatives like sequential forward selection (SFS) [10] and sequential backward selection [11] tool have been published to date with numerous solutions to the FS task. These strategies do, however, have shortcomings such as slow convergence, optimal local trapping, and are computationally costly [12].

Approaches based on evolutionary algorithms were proposed to overcome the limitations described above. These architectures can identify more optimum solutions than greedy methods because of the population-based and improved global search capabilities. The genetic algorithm (GA) is utilized for the resolution of the FS issue in numerous research studies as a well-known metaheuristic method [13]–[16]. Moreover, other algorithms such as Grey Wolf Optimization (GWO) [17], Harris Hawk Optimizer (HHO) [18]–[20], Forest Optimizer (FO) [3], Dragonfly Algorithm (DA) [21], Jaya optimization method [22], Bacteria foraging method [23], spider monkey optimization [24], spotted hyena method [25], salp swarm algorithm [26], ant lion optimizer [27], and teaching-learning optimization [28] etc., have been employed to tackle FS tasks.

In principle and practice, the utilization of population-based research and the production of various unique solutions might make evolutionary algorithms suited to resolve multi-objective issues. Due to its structural character, FS may be classified as a multi-objective optimization problem (MOP), given that it takes account of at least two conflicting objectives. MOPs are generally supported by a number of non-dominated (ND) solutions which represent a compromise between opposing aims, providing various options for decision-makers. In addition, the FS was examined in many objectives compared to the single objective state in a small number of ways, according to related research [29], [30].

The Chimp Optimizing Algorithm (ChOA) is one of the advanced metaheuristic technique developed for the resolution of optimization issues. According to prior studies, this technique has properties such as low feature assessment, high-speed, and excellent global and local discovery [31], [32]. The capacity of this method for handling the FS assignment has yet to be explored to the best of our understanding.

This article aims mainly to build a ChOA-formed multi-objective wrapper FS technique with dual archive, which can concurrently decrease the amount of features and boost the classification performance, and provide a set of Pareto solutions. Because FS is a multi-target optimization problem, we propose for the first time a binary variant of Multi-objective ChOA called Binary Multi-Objective Chimp Optimization with a Sigmoid Transfer Function (BMOChOA-S).

In particular, the following goals have been examined in this article.

1) To gain insight into the strengths and weaknesses of recent works on metaheuristic-based FS tasks.
2) To introduce a BMOChOA-S to discover Pareto optimum solutions for the FS work.
3) To present a comparative analysis report on the performance of the BMOChOA towards the FS task considering healthcare data, by implementing twelve different variants of BMOChOA depending on the chimp group and the type of chaotic map used.
4) To evaluate the suggested technique with three well-considered multi-target approaches and examine whether the strategy proposed outweighs benchmarking methodologies for limiting the length of feature subset and increasing the classification accuracy.
5) To confirm the performance of the presented technique in terms of computational cost.

The structure of this article is organized accordingly. In Section II, we discuss the ChOA algorithm, the multi-target optimization idea, and current FS experiments. The suggested multi-target technique is explained in Section III. Section IV shows the experimental set-ups. Section V fully describes the results and comments. Finally, the last part provides the conclusion and suggests future guidelines for work.

II. BACKGROUND
First, the basic ChOA principles and multi-objective optimization techniques are outlined in this part. Then, a summary of significant work done in the FS world is presented.

A. ChOA
ChOA's conceptual foundation is based on chimpanzees' hunting habits. The standard ChOA separates the chimp group into four types: attacker, barrier, chaser, and driver. The attacker is the leader among them. The other three varieties of chimpanzees helped to hunt, which in turn reduced their status. Drivers track the prey but do not try to arrest them with

it. Barriers are placed in trees to construct a dam covering the movement of the prey. Chasers move to catch it quickly after the prey. In the end, the attackers predict the break-out course of the prey towards the chasers or down to the lower canopy. This crucial job (attack) has a favourable correlation with age, intelligence, and physical capacity. In addition, during the same hunt, chimps might swap responsibilities or retain their jobs over the whole procedure [33]. Broadly speaking, the technique of chimp hunting is separated into two major phases: exploration (driving, jamming, and chasing the prey) and exploitation (attacking the prey).

Mathematically, the two phases are described below:

1) DRIVING AND CHASING THE PREY

In the exploration and development phase, the prey is pursued. Eqs. 1 and 2 are used for mathematical modelling of driving and tracking prey.

$$b = \left| c3 P_{prey}(CI) - c2 P_{chimp}(CI) \right| \quad (1)$$

$$P_{chimp}(CI+1) = P_{prey}(CI) - c1.b \quad (2)$$

Here, CI: the current iteration count, $c1, c2,$ *and* $c3$: the coefficient vectors, P_{prey}: the prey's position, P_{chimp}: chimp's position. c1, c2, and c3 vectors are calculated by the Eqs. (3) – (5), respectively.

$$c1 = 2 \cdot f \cdot rnd1 - f \quad (3)$$

$$c2 = chaotic\ value \quad (4)$$

$$c3 = 2 \cdot rnd2 \quad (5)$$

$rnd1$ and $rnd2$ are the random vectors $\epsilon\ [0, 1]$ and c2 is a chaotic vector computed on the basis of several chaotic maps to show the influence of chimps' sexual drive on the hunting progression.

Hypothetically, separate independent groups with a shared purpose may be utilized in each population-based optimization method to provide a direct and random search result simultaneously. Any continuous function can be taken out to update distinct chimp groups. These functions have to be selected such that f is decreased throughout each run [34]. The procedure of driving and chasing the prey is pictorially shown in FIGURE 1 [31].

2) ATTACKING (EXPLOITATION)

The process of hunting is generally done by attackers. Occasionally, drivers, barriers, and chasers take part in hunting. Apparently, there is no information regarding the optimal position in a conceptual search area. To mathematically replicate the behaviour of chimps, the initial attacker (the best solution), driver, barrier, and chaser are better informed about the status of the prospective prey. So four of the best yet achieved solutions are kept, and the other chimps need to change their spots according to the ideal chimp positions. This is expressed as follows:

$$c11 = 2 \cdot f1 \cdot rnd1 - f1$$

$$c12 = 2 \cdot f2 \cdot rnd1 - f2$$

FIGURE 1. Driving and chasing the prey [31].

$$c13 = 2 \cdot f3 \cdot rnd1 - f3$$

$$c14 = 2 \cdot f4 \cdot rnd1 - f4 \quad (6)$$

$$b_{Attacker} = \left| c3 \cdot P_{Attacker}(CI) - c2 \cdot P_{chimp}(CI) \right|$$

$$b_{Barrier} = \left| c3 \cdot P_{Barrier}(CI) - c2 \cdot P_{chimp}(CI) \right|$$

$$b_{Chaser} = \left| c3 \cdot P_{Chaser}(CI) - c2 \cdot P_{chimp}(CI) \right|$$

$$b_{Driver} = \left| c3 \cdot P_{Driver}(CI) - c2 \cdot P_{chimp}(CI) \right| \quad (7)$$

$$P1(CI+1) = P_{Attacker}(CI) - c11 \cdot b_{Attacker}$$

$$P2(CI+1) = P_{Barrier}(CI) - c12 \cdot b_{Barrier}$$

$$P3(CI+1) = P_{Chaser}(CI) - c13 \cdot b_{Chaser}$$

$$P4(CI+1) = P_{Driver}(CI) - c14 \cdot b_{Driver} \quad (8)$$

$$P_{chimp}(CI+1) = \frac{P1 + P2 + P3 + P4}{4} \quad (9)$$

Here, CI is the current iteration. The dynamic coefficient c1 and vector b are computed using eq 6 and 7. With the passage of repetition, f drops non-linearly from 2.5 to 0. The dynamic coefficients for f taking two distinct versions of ChOA (ChOA1 and ChOA2) with various independent groups are given in APPENDIX VI [31].

B. MULTI-OBJECTIVE OPTIMIZATION (MOP)

Many real-world tasks generally consist of a group of goals that must be optimized at the same time. The solution to these problems is a series of solutions that really represent a compromise between distinct objectives. The set containing all the trade-off solutions to a given problem is called the Pareto optimal set or Pareto front. Mathematically,

$$max\ f_k(z), \quad k = 1, 2, \ldots\ldots, K$$

$$min\ h_l(z), \quad l = 1, 2, \ldots\ldots, L$$

$$Subject\ to: g_m(z) \geq 0, \quad m = 1, 2, \ldots, M$$

$$e_n(z) = 0, \quad n = 1, 2, \ldots, N$$
$$z_i lb \leq z_i \leq z_i ub$$

For a single objective optimization problem (SOP), a candidate's superiority in relation to other solutions is determined via a fitness comparison. The merit of a candidate solution is nonetheless judged in MOP by the concept of dominance. A solution P in the objective space of a C-objective problem dominates another solution Q if the following two criteria are true:

1) $\forall_C$: P is not worse than Q
2) $\exists_c$: c ∈ C and Pc is strictly better than Qc

C. RELATED WORKS

In practice, FS methods are divided into three primary types: embedded, filter, and wrapper. However, our concentration is on a wrapper-based approach. Wrapper techniques utilize a search procedure to discover almost optimal solutions and a classifier for rating the solutions that have been identified. Thus, wrapper methods may be subdivided into two: evolutionary and non-evolutionary groups based on the type of search strategy. Branch & Bound [35], SFS [10] and SBS [11] might be considered among the most well-known non-evolutionary solutions. Although these methods may be implemented relatively simply, they struggle with issues including convergence to local optimums and with substantial computing overheads for huge data. Structural failure also occurs in both SFS as well as SBS techniques; thus, the following steps may not remove (add) features which have previously been included (excluded) from the set [6]. Sequential Floating Selection (SFFS) and Sequential Backward Floating Selection (SBFS) were developed to tackle this issue [36]. Unfortunately, advancements in these methods have not solved the convergence problem to the optimal local level [37]. We might refer to Focus [37] and Relief [38] as two of many non-evolutionary filtering approaches. Filter processes such as mRmR [39] or MIFS [40] also attempt to enhance the efficiency of the FS algorithm by applying information theory ideas. This improvement is obtained by looking at the relevancy of an attribute to the outcome and any duplication among the characteristics.

Researchers used evolutionary techniques to resolve the difficulties described above and to apply better search strategies. These algorithms create and assess numerous solutions concurrently because they are population-based and have better global discovery than traditional approaches. Single objective wrapper techniques generally serve the purpose by restricting the length of the feature set, or by enhancing the classification efficiency, or by aggregating these targets [41]. For a better understanding of the single-objective evolutionary methods for solving FS task the interested reader can refer to DA [21], SSA [26], [42], HS [43], TLBO [28], grasshopper optimization [44], Jaya algorithm [22], [45], HHO [18], atom search [46], SMO [24], SHO [25], CS [47], ALO [48], ABC [30], FOA [49], FPA [50], and WOA [51], [52] etc.

The excellent quality of any meta-heuristic methodology is focused and constrained. Exploration and exploitation are two opposing conditions to be considered when developing meta-heuristics. The meta-heuristic algorithms perform well in some cases but poorly in others, so it is critical to strike a reasonable balance between exploration and exploitation to improve the algorithms' efficiency. Each nature-influenced methodology has its own positive and negative aspects, such that the right algorithm for a particular problem is not guaranteed. We cannot find the optimal solution for each kind of function with the individual optimization algorithm [53]. The implementation and proposal of modern meta-heuristics with high precision for actual implementations has therefore become a challenge to scientists, [54]. As a result, the hybridization of evolutionary methods has engaged many research people to solve FS problems. The aim of hybridization is to identify compatible alternatives in order to ensure the optimal output of optimization methods, which is accomplished by combining and coordinating the exploration and exploitation processes [55]. Hybridization of evolutionary methods is a common method for combining the strengths of independent architectures to address such shortcomings [54]. Some recently suggested hybrid evolutionary techniques for handling FS task are: ABC-GA [56], MA-HS [57], PSO-GE [58], GWO-PSO [41], TEO-SOA [59], GWO-CSA [60], PSO-FLA [61], SCA-ALO [62], TLBO-SSA [63], HHO-CS [64], SCA-HHO [65], GWO-HHO [66], and SCA-CS [67] etc.

The ChOA has been proposed recently to solve different optimization tasks [31]. ChOA is meant to mitigate two challenges in the resolution of high-dimensional issues: poor convergence speed and entrapment in the local optimum. Jia *et al.* [32] in 2021 have attempted an enhanced ChOA for solving optimization problems in the continuous domain. Gaurav Dhiman has suggested a fusion of SSA and SHO based ChOA for tackling the optimization applications in engineering [68]. Also, a hybrid SCA-ChOA method is proposed by Kaur *et al.* [69] for HLS of datapaths in digital filters and engineering applications. **To the best of our knowledge, no work exists in the literature for discrete ChOA in order to find solutions to problems like feature selection.** For the first time, we proposed a discrete version of the ChOA with a multi-objective essence for selecting relevant factors from variable sized healthcare data.

It is noticeable that these strategies provide one almost optimum result using single target techniques. However, the FS is fundamentally a MOP that attempts to achieve at least two opposing goals: to reduce the number of characteristics and to increase the efficiency of the classification. In these cases, the true response is for clients to provide a number of non-dominated (ND) alternatives to choose solutions which are suitable to their circumstances. We will look at the multi-objective techniques offered to the FS issue in the following part of this article.

In the past few years, multi-objective heuristics have been the focus of significant investigation due to the simultaneous evaluation of many typically conflicting objectives and the presentation of a series of ND solutions. In the study [70], the GA and non-domination concepts were used to address the FS issue with the NPGA multi-objective method. NSGA-I was used in [14] to take into account the aims of lowering the number of characteristics and decreasing the artificial neural network's classification error. Zhu *et al.* [71] have introduced a hybrid filter and wrapper procedure utilizing a mimetic method using a filter process to enhance wrapper MOEA solutions by the addition/removal of features depending on correlation condition. Huang *et al.* [72] in 2010 have sought to locate the Pareto front using the NSGA-II method. The MOPSO method was used to FS in the studies by Xue *et al.* [12]. Vignolo *et al.* [73] have introduced a novel MOGA-based method for FS in the face recognition domain. Three major goals are considered here: 1) increasing accuracy, 2) decreasing the number of characteristics, and 3) limiting mutual information (MI). In Ref [74], a multi-objective FS (DEMOFS) method employing differential evolution was presented. In Ref [75], where the accuracy of every class is treated as a cost function, the NSGA-II wrapper methodology was introduced. The study in Ref [76] investigates the MOFS-BDE, a novel multi-objective FS strategy that incorporates three operators: a unique binary mutation operator, a One-bit purifying Search operator, and a fast non-dominated sorting operator to increase performance. Zhang *et al.* [77] have suggested the first work of multi-objective PSO with Pareto dominance and an external repository for cost-based FS problems. The PSOMOFS, presented by Hu *et al.* [78] is a fuzzy multi-objective FS approach with PSO to solve the FS issue with fuzzy costs. To trim the elitist repository, this technique provides a fuzzy dominance relation to check the excellence of nominee particles and sets a fuzzy crowding distance (CD) metric to discover the global leader. In Ref [79], an innovative MOPSO technique was presented for addressing the FS challenge by using local search to enhance the repository solutions. An enhanced MOPSO technique is proposed in Ref [80] that utilizes t-score and precision as goal functions. Ragothaman and Sarojini [81] have tried to address the FS problem on healthcare datasets using a MOABC inspired by non-dominated sorting. Jimenez *et al.* [82] utilized the ENORA technique, which applied dominance and slot principles to identify the Pareto front in order to pick parameters in the online sales prediction method. In the article [30], the FS problem was solved using a multi-target ABC approach. In Ref [83], a mixed MO approach built on the SSA and the SHO has been proposed for FS. This approach prefers to mix SSA and SHO to balance variability and convergence to obtain the best Pareto front. The NSGA-II method-based hybrid filter wrapper was developed in the work [84]. Baliarsingh *et al.* [85] used a fisher score technique together with a multi-objective Penguin Optimizer. Numerous recent research papers, such as [86]–[88], are concurrently working on resolving the

FS issue and optimizing the classifier's parameters. For example, [86] utilized MOPSO to fulfil three objectives: optimizing the settings of SVM, choosing a kernel function, and also picking the right characteristics. A unique method using the multi-objective Grey Wolf optimizer (MOGWO) was introduced by Al-Tashi *et al.* [17] in 2020. The technique presented employs the repository to preserve non-dominated alternatives. Recently, in 2021, Piri *et al.* [20] have suggested a MO quadratic binary harris hawk optimizer for selecting required features from 12 medical datasets. Also, they have proved the superiority of the proposed technique by comparing the results with deep-based FS methods such as AE and TSFS. A multi-objective FOA has been introduced in the article [3] for tackling the FS task. When the depth of FS issues grows, the solution space grows exponentially, resulting in a large number of local optima. As a result, for large-scale FS issues, present evolutionary approaches still suffer from the problem of local optima stagnation. To address this, Xue *et al.* [89] have presented a self-adaptive PSO (SaPSO) method for FS, specifically for widescale data. To cope with high-dimensional FS challenges, Song *et al.* [90] have presented an innovative FS technique called "bare bones PSO" (BBPSO), integrating mutual information. They additionally devised a successful swarm initialization technique based on correlation in order to hasten swarm convergence. Also, the authors in Ref [91] have developed a novel three-phase hybrid FS method based on correlation-guided clustering and PSO for large dimensional data.

Most prior research has employed single objective evolutionary algorithms to address the FS issue, as can be observed from the literature, whereas few studies have concentrated on MOFS techniques compared to the other methods. Moreover, the effectiveness of the ChOA to solve MOPs has yet to be studied given the potential functionalities of ChOA, such as the ease of operators, the reduced number of fitness assessments, and the small number of parameters. Therefore, in this article, we suggest the BMOChOA method to solve the FS issue.

III. PROPOSED METHODOLOGY
As per the study in the previous part, the FS is a MOP that may intrinsically address at least two goals. That is, to limit the count of characteristics and boost the accuracy of classification. The binary version of ChOA to resolve discrete optimization problems like FS has not been established yet. Furthermore, no implementation of ChOA to address FS as MOP in the existing work has yet been suggested. This portion thus contains a Binary Multi-Objective ChOA (BMOChOA) to tackle the FS problem for the first time in healthcare data mining. FIGURE 2 shows the overall design of the presented BMOChOA-based FS job. In the following, FS is regarded as a binary optimization task in the context of a 0/1 or discrete presentation of candidate solutions. The ChOA is not directly suited to the FS challenge, as it aims to solve the problem of continuous optimization.
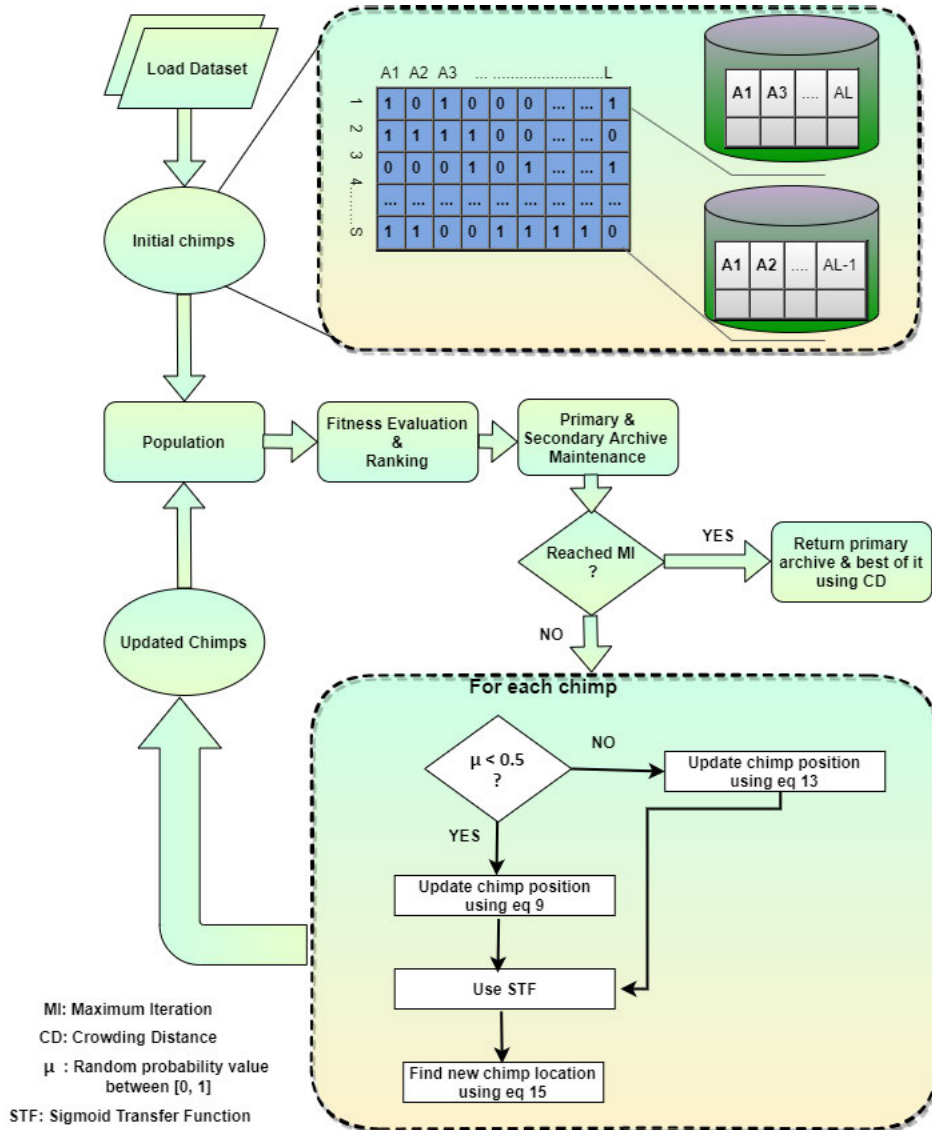
FIGURE 2. Working procedure of the proposed BMOChOA.

In order to attain the aim, various changes are thus necessary:

1) **Representation of chimp:** Each chimp location in BMOChOA is represented as a vector of length L, consisting of only 0s and 1s. If a particular bit value in the nominee position vector is 1, then the corresponding feature is considered to form the reduced dataset, otherwise not. The pictorial representation of the structure of each chimp's location is shown in FIGURE 2.

2) **Fitness Assessment:** This research considers FS as a bi-objective optimization problem. Therefore, each chimp in the population is evaluated by applying the following two goal functions:

$$Obj1(P) = \sum_{j=1}^{L} p_j, \quad if \ p_j = 1 \quad (10)$$

where P is the location string of a chimp of length L.

$$Obj2(P) = Classification \ Accuracy$$
$$= \frac{TPS + TNS}{TPS + TNS + FPS + FNS} \quad (11)$$

Here, TPS: True +ves, TNS: True −ves, FPS: False +ves, and FNS: False −ves [92].

The following procedures are taken to calculate the objective function Obj2 for each chimp:

- First, a compressed data set is formed by taking the characteristics from those indices of the chimp location string where the value is 1.
- Then the KNN model with a 10-fold CV is applied to compute the Obj2 value of each chimp. For training and testing, the total samples of a dataset are randomly split into 10 groups.

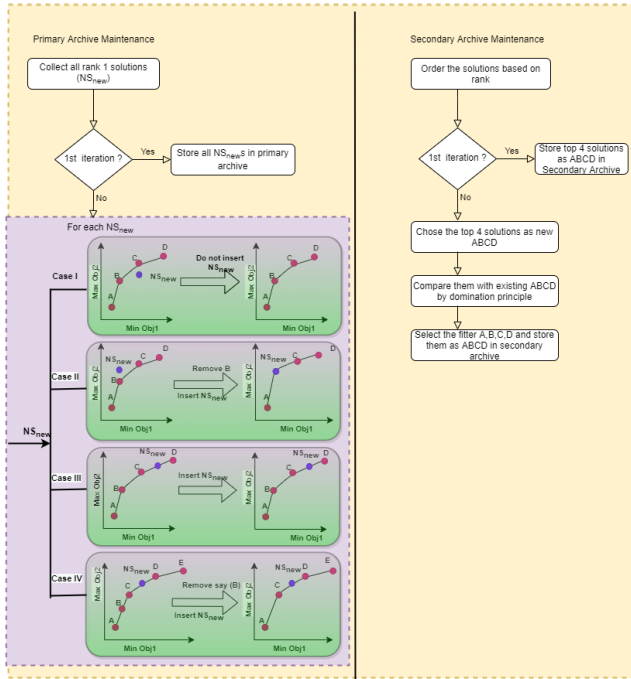3) **Maintenance of Primary Archive:** The primary archive must be updated following evaluations of each

chimp in the swarm because, after completion of each iteration, the BMOChOA provides a collection of Pareto optimum solutions rather than a single one. A fresh non-dominated solution NS_{new} of the in progress iteration is allowed to be inserted into the primary archive depending on the following scenarios:

- If the NS_{new} is dominated by at least one of the primary archive participants, it is not allowed to be entered.
- If NS_{new} dominates any current member of the primary archive (say Q), NS_{new} will substitute Q for improved archive construction.
- If NS_{new} does not dominate the current primary archive residents, and if NS_{new} is not dominated by the present archive participants, and there is sufficient room for the new solution, then put the NS_{new} in the primary archive.
- If NS_{new} and the present primary archive residents are not dominant, and for the new one, no area is there, then drop one alternative from the most congested location and move NS_{new} to the archive [20].

4) **Secondary Archive Management:** In this paper, after each generation, we consider a solution optimal if its rank value is 1. The primary archive is responsible for storing all the optimal solutions of a particular iteration. However, for practical simulation of chimps' behavior, the algorithm requires four best solutions (A: Attacker, B: Barrier, C: Chaser, D: Driver) in each iteration. But during the execution of each loop, it may happen that there are not four best solutions ranked 1 to find

ABCD. Therefore, in each iteration, we have ordered the chimps according to their rank and chose the top four solutions as ABCD. Again, after each generation, we may get a different set of ABCD that are better than the previous ABCD. So we need to preserve the location of ABCD after each loop. Here we feel the presence of a secondary archive to keep the ABCD of each iteration. After finishing a particular iteration, the new ABCD group is compared with the older one using the dominance principle, and the fresh one will replace the existing one if it is fitter. Then the new set of ABCD is used to update the chimp position in the next iteration. The schematic diagram for both primary and secondary archive management is shown in the FIGURE 3.

5) **Conversion from Continuous to Binary ChOA:** Existing research demonstrates the effectiveness of the transfer function (TF) to turn a continuous optimizer into a binary one because of its simplicity, low cost, and rapid and easy execution. The majority of researchers employed the common TFs in the form of S and V to alter the continuous optimizer into a discrete one. But this proposal uses the S-form (sigmoid ()) method to achieve the same.

$$T(y) = sigmoid(y) = \frac{1}{1 + e^{-2y}} \quad (12)$$

6) **Chimp Location Update:** Assume that μ is a random probability value between 0 and 1. Then if μ is less than 0.5, the equation 9 is applied to alter the current location of the respective chimp. Similarly, a chaotic value is used for updating the chimp if $\mu >= 0.5$ (given in equation 13). In this article we have taken 6 important chaotic maps and their descriptions are mentioned in APPENDIX VI.

$$P_{chimp}(CI + 1) = chaotic\_value \quad (13)$$

Here we have implemented two variants of BMOChOA, namely BMOChOA1 and BMOChOA2. Therefore, depending on the chaotic maps used for each type in order to update the chimp location, there are 12 different BMOChOAs that are coded. The APPENDIX VI contains the detailed naming.

After calculating the new location of the chimp $(\Delta p_i^L(CI + 1))$ in the continuous domain, the sigmoid TF comes into front to convert it to a probability value by using the equation 14.

$$T(\Delta p_i^L(CI + 1)) = sigmoid(\Delta p_i^L(CI + 1))$$
$$= \frac{1}{1 + e^{-2\Delta p_i^L(CI+1)}} \quad (14)$$

Then the BMOChOA alters the new location of each chimp to continue the next iteration, by applying the equation 15.

$$p_i^L(CI + 1) = \begin{cases} 1, & rd < T(\Delta p_i^L(CI + 1)) \\ 0, & otherwise \end{cases} \quad (15)$$

where, i: i^{th} bit of the chimp's position, L: original dimension, *rd*: random number between 0 and 1, and CI: current iteration

7) **Returning the best of Primary archive:** The primary archive holds all the solutions that are not mutually dominated after the given number of steps. As a screening method, the CD value is utilized here to pick an optimum combination of features from a group of non-dominated ones. When the entire population is present in the primary archive, the calculation of the CD needs a C number of arrangements of a maximum of S solutions, resulting in a time complexity of (CSlogS). For details regarding the CD calculation procedure, refer to the article [20].

8) **BMOChOA Algorithm**

Algorithm 1 CD-Based BMOChOA for FS

1: Set the initial size of the population (S) and the upper bound of iterations to be used in the simulation (MI).
2: Set the initial chimps' location.
3: Evaluate the chimps using fitness functions.
4: Select the Attacker, Barrier, Chaser, and Driver based on their rank and store them in a secondary archive.
5: Save all non-dominated chimp locations into a primary archive.
6: **for** $CNT \leftarrow 1$ to MI **do**
7:    **for each** chimp **do**
8:       **if** $\mu < 0.5$ **then**
9:          Update the position of the chimp using equation 9.
10:       **else**
11:          Update the position of the chimp using equation 13.
12:       **end if**
13:       Use sigmoid() to transform the chimp's position into a probability value.
14:       Find the new location of the chimp in binary domain using equation 15.
15:    **end for**
16:    Update: f, c1, c2, c3, and b.
17:    Evaluate the updated chimp position.
18:    Update $P_{Attacker}$, $P_{Barrier}$, P_{Chaser}, P_{Driver}.
19:    Update the primary and secondary archives.
20: **end for**
21: Return the primary archive along with the best of it using CD measure.

---

9) **Analysis of Time Complexity:** The time complexity of the presented wrapper-based BMOChOA for FS can be computed as follows:

$$O(BMOChOA)$$
$$= O(initialization)$$
$$+ MI * [O(updating\_chimps) + O(evaluation)$$
$$+ O(finding\_non - dominated\_solutions)$$

$$+ O(rank - based\ ordering\ of\ the\ population)$$
$$+ O(sorting\_of\_primary\ archive\ using\ CD)]$$
$$+ O(output) \tag{16}$$

The time complexity of the KNN for Q samples is $O(Q * L)$ [20]. For finding the non-dominated solutions after each repetition, we have used the concept of dominance tree, which minimizes the number of duplicate comparisons, resulting in the time complexity of O(CSlogS) [93]. The complexity of the initialization task can be derived by the equation 17:

$$O(Initialization) + O(fitness\_evaluation)$$
$$+ O(rank - based\ ordering)$$
$$+ O(finding\_non - dominated\_solutions)$$
$$= S * [O(L) + O(Q * L)] + O(SlogS) + O(CSlogS)$$
$$\approx O(CSlogS) \approx O(SlogS) \tag{17}$$

Here, S:-population size, C:-number of objectives, Q:-sample size, and L:-actual dimension.
The CD-based ordering of the primary archive in each loop takes O(CSlogS) [20]. Hence, the overall computational complexity of the introduced FS method can be expressed as:

$$O(BMOChOA) = O(SlogS)$$
$$+ MI * [O(S * L) + S * O(Q * L)$$
$$+ O(CSlogS) + O(SlogS)$$
$$+ O(CSlogS)] + O(1) \approx O(CSlogS)$$
$$\approx O(SlogS) \tag{18}$$

IV. EXPERIMENTAL SETUPS

1) **Datasets in Details:** The achievement of the BMOChOA is confirmed by the use of eleven conventional datasets of various dimensions from UCI and three microarray cancer datasets [94]. TABLE 1 explains each dataset's structure, including the number of attributes (9-7129), instances (32-2126) and categories (2-21). To overcome the numerical issue, every dataset is normalized.

2) **Methods for Comparison:** To verify the superior performance of the presented BMOChOA in producing the Pareto fronts, three of the most popular MOFS techniques are used in this article: 1) MOGA 2) MOPSO, and 3) NSGA-II. In the first approach, [15], MOGA, a genetic algorithm with a multi-objective flavour, is applied to solve the FS task. Here, the CD measure is used to maintain population diversity. The second method, MOPSO, [12], [95] is a very popular multi-objective technique to handle the FS task. Here too, a repository is used to keep the updated fittest solutions from every iteration. The third strategy, as suggested by Deb *et al.* [96], is an NSGA-II and is a widely recognised MO approach. The method consists

TABLE 1. Datasets in details.

SID	Name	#Instances	#Attributes	#Classes
D1	BreastCancerW	699	9	2
D2	ILPD	583	10	2
D3	PrimaryTumor	339	17	21
D4	Lymphography	148	18	3
D5	Diabetic	1151	19	2
D6	Hepatities	155	19	2
D7	Cardiotocography	2126	21	3
D8	Cervical Cancer	858	35	2
D9	LungCancer	32	56	3
D10	Arrhythmia	452	279	16
D11	Parkinsons	756	754	2
D12	Colon tumor	62	2000	2
D13	SRBCT	83	2308	4
D14	Leukemia	72	7129	2

TABLE 2. Parameter settings.

Parameters	BMOChOAs	MOGA	MOPSO	NSGA-II
Population size	100	100	100	100
#Generations	50	50	50	50
Archive size	50	50	50	50
Mutation probability	-	0.03	0.02	0.02
Crossover probability	-	-	-	0.7

of two approaches: the preservation of variation and the non-dominated sorting procedure.

The aforementioned techniques were built and tested in Python 3.7 on an Intel Core i3-7020U CPU @ 2.30 GHz and a 4.00 GB RAM machine.

3) **Criteria for Comparison:** To assess the performance of the suggested technique, four criteria have been taken into consideration: 1) Inverted Generational Distance (IGD), 2) Hypervolume (HV), 3) Spread, and 4) Success Counting (SCC). The descriptions of the four criteria are provided below:

- **IGD [97]:** This metric is for estimating the distance between the True Pareto Front (TPF) and the Calculated Pareto Front (CLF). It is defined as:

$$IGD(CLF, TPF) = \frac{1}{|TPF|}(\sum_{j=1}^{|TPF|} d_j^2)^{\frac{1}{2}} \quad (19)$$

where $d_j = min_{s \in CLF} \|F(s) - F(j)\|$. The lower the value, the closer is the CLF to the TPF and the quicker is the rate of convergence. To calculate the TPF of each dataset we followed the following steps [98]:

1. Each evolutionary algorithm (EA) is run R times on the test issue (R = 20 in this case).
2. In Step 1, all of the solution sets that were acquired are combined into a single set. After that, only non-dominated solutions are chosen. The selected non-dominated solutions are employed as TPF.
3. The IGD measure evaluates each solution set acquired in each run of each EA using the chosen TPF in Step 2.

- **Hyper Volume (HV) [97]:** The HV indicator is characterized as the volume of the area in an objective space dominated by the CLF and marked from above by a reference point $ref \in R^I$ such that for every $s \in CLF$, s dominates ref. It is calculated as:

$$HV(CLF, ref) = \lambda_I(\bigcup_{s \in CLF}[s; ref]) \quad (20)$$

where λ_I is the I-dimensional Lebesgue measure.

- **Spread [99]:** This metric is used to estimate the spread of solutions present in the PF. The higher the value, the better the spread of Pareto solutions.

$$Spread = \sqrt{\sum_{i=1}^{C}(max_{j=1}^{|CLF|}Obj_i^j - min_{j=1}^{|CLF|}Obj_i^j)^2} \quad (21)$$

- **SCC [3]:** On the basis of these criteria, the total ND solutions of all techniques are gathered into a set TPF to measure the efficiency of multi-objective strategies, and then the CLFs of all the methods are retrieved from the elements of the set TPF. Then the contribution level of each technique in the TPF is computed using the following equation.

$$SCC = \sum_{j=1}^{|TPF|} S_j \quad (22)$$

where, S_j value is 1 if the jth solution of the TPF is present in the CLF of the algorithm under consideration, otherwise it is 0.

4) **Parameters settings:** As the proposed FS technique is a wrapper-based approach to choose the best feature subset with greater classification performance, we have taken the KNN classifier with a 10-fold CV as a wrapper for evaluating the feature sub string. The KNN algorithm may easily be implemented. KNN just needs to run the K value and the distance type function. When it generates predictions, it really gathers and saves information about the training sets. Because of the lack of training before the forecasts, fresh data that does not influence the accuracy of the KNN system may be incorporated simply. To execute all the above-discussed FS methods, the parameter values shown in TABLE 2 are taken.

V. RESULTS AND DISCUSSION

A. PERFORMANCE COMPARISON BETWEEN 12 VARIANTS OF BMOChOA

All 12 versions of the BMOChOA are executed 25 times independently. TABLE 3 shows the IGD values of the best Pareto front obtained from each BMOChOA for 14 different healthcare datasets. It is clearly visible from TABLE 3, that the IGD values of the CLFs produced by BMOChOA16 are
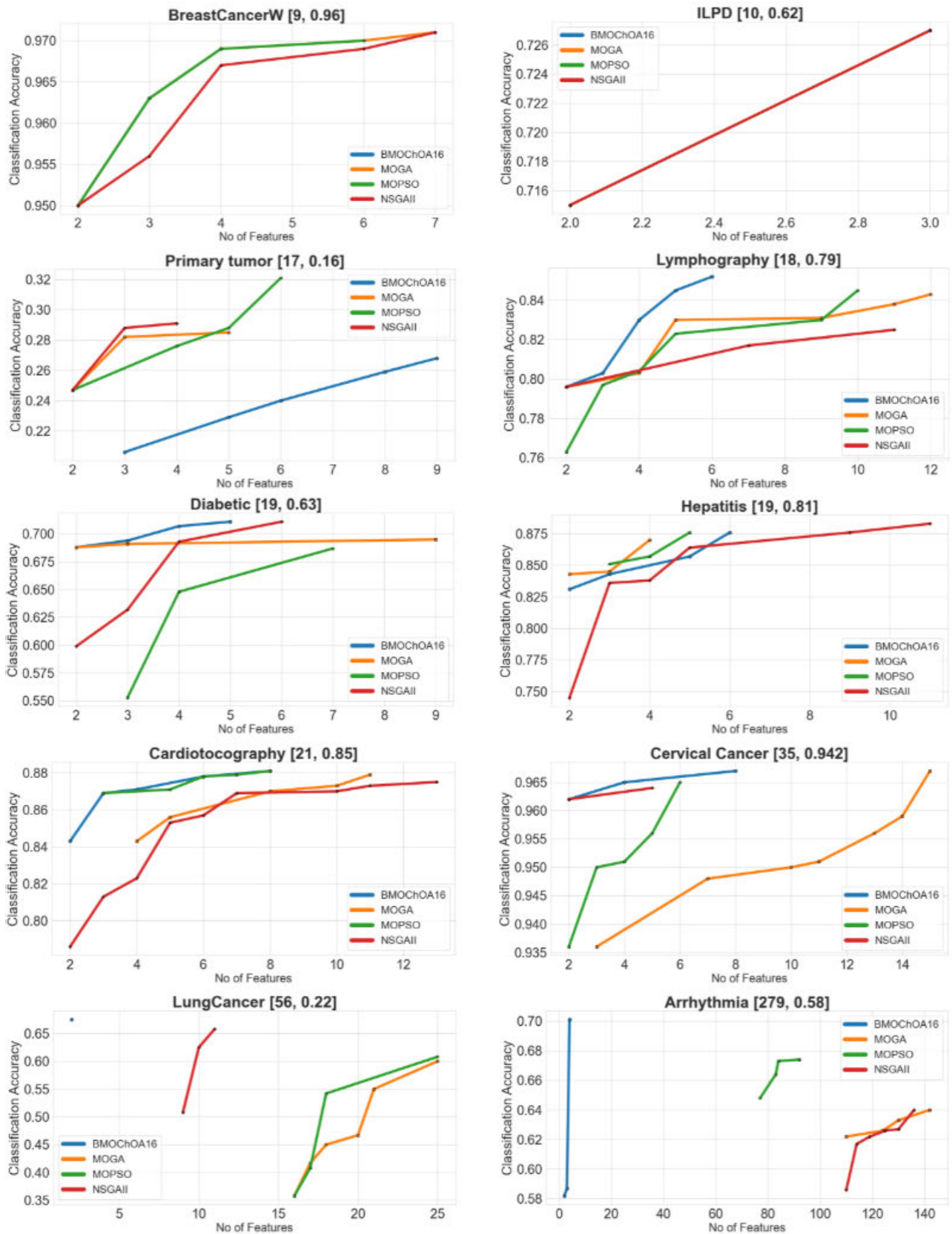
FIGURE 4. Pareto fronts of four methods for D1 to D10.

TABLE 3. IGD values of CLFs obtained by BMOChOAs.

Datasets	BMOChOA (IGD)											
	11	12	13	14	15	16	21	22	23	24	25	26
D1	0.071	0.062	0.059	0.151	0.046	**0**	0.046	0.042	0.125	0.05	0.046	**0**
D2	0	0	0.197	0.001	0.188	**0**	0	0	0.001	0.114	0.002	0.002
D3	0.122	0.026	0.019	0.035	0.068	**0.008**	0.05	0.058	0.06	0.041	0.232	0.036
D4	0.059	0.074	0.03	0.046	0.015	**0**	0.059	0.021	0.024	0.059	0.056	0.138
D5	0.093	0.094	0.089	0.023	0.03	**0.001**	0.022	0.296	0.088	0.093	0.03	0.091
D6	0.074	0.137	0.07	0.054	0.033	**0.029**	0.44	0.054	0.056	0.07	0.043	0.137
D7	0.059	0.074	0.03	0.046	0.09	**0.02**	0.296	0.088	0.093	0.03	0.091	0.138
D8	0.189	0.048	0.193	0.117	0.211	**0**	0.152	0.053	0.174	0.13	0.269	0.096
D9	0.912	0.594	0.918	1.076	1.078	**0**	0.482	0.575	0.874	0.929	0.937	0.808
D10	0.205	0.231	0.036	0.077	0.093	**0.015**	0.317	0.098	0.174	0.152	0.044	0.172
D11	0.053	0.034	0.027	0.03	0.262	**0**	0.031	0.072	0.055	0.036	0.26	0.15
D12	0.226	0.376	0.178	0.179	0.182	**0.06**	0.417	0.307	0.063	0.168	0.16	0.094
D13	0.193	0.054	0.106	0.258	1.263	**0.025**	0.103	0.079	0.041	0.029	0.193	0.258
D14	0.146	0.119	0.156	0.108	0.181	**0.048**	0.085	0.118	0.266	0.118	0.323	0.466

TABLE 4. Outputs of Wilcoxon signed rank test.

Datasets	BMOChOA										
	11	12	13	14	15	21	22	23	24	25	26
D1	0.000	2.00×10^{-8}	0.000	0.000	**0.1324**	0.000	0.000	**0.241**	0.000	**1.000**	0.000
D2	**1.000**	**0.0599**	0.000	0.000	0.000	9.00×10^{-4}	0.000	**0.0612**	0.000	**0.054**	**1.000**
D3	0.00341	0.000	0.000	0.021	0.000	**0.0981**	0.000	0.000	0.000	0.000	**0.094**
D4	**0.073**	0.000	0.002	0.000	**0.0587**	0.000	0.011	0.000	0.000	**1.000**	0.001
D5	0.000	0.000	0.000	0.000	0.003	0.000	0.0234	**0.094**	0.000	0.000	0.003
D6	0.0001	0.000	1.00×10^{-11}	0.0178	8.00×10^{-5}	0.000	0.000	0.000	0.013	0.001	0.042
D7	0.000	0.0134	6.00×10^{-7}	0.0003	8.00×10^{-5}	0.000	2.00×10^{-5}	5.00×10^{-7}	0.002	0.012	0.049
D8	3.00×10^{-5}	**0.4528**	0.000	0.000	0.000	**1.000**	8.00×10^{-4}	0.000	0.000	0.000	0.000
D9	0.000	0.000	0.000	**0.0722**	0.0151	0.0012	0.002	0.002	0.009	**0.098**	0.000
D10	0.000	9.00×10^{-5}	0.000	3.00×10^{-12}	**0.0956**	0.0345	0.000	0.022	0.000	0.004	0.038
D11	0.000	**0.081**	0.000	0.000	0.000	0.000	0.021	0.000	0.034	0.000	**1.000**
D12	0.0136	0.000	0.000	0.0001	0.0012	0.000	0.000	**0.067**	0.022	**1.000**	0.031
D13	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.043	0.000
D14	0.000	0.000	0.0147	**0.1777**	**1.000**	0.000	0.000	**0.072**	0.000	0.026	**0.092**
(b,e,w)	(12,2,0)	(11,3,0)	(14,0,0)	(12,2,0)	(10,4,0)	(11,2,1)	(14,0,0)	(9,5,0)	(14,0,0)	(9,5,0)	(6,4,4)

lower for all the 14 datasets as compared to other variants of BMOChOAs. The lower value of IGD indicates the higher convergence speed because it denotes the gap between the TPF and CLF. BMOChOA16 used a Tent Chaotic map for position updates of each chimp. The tent map is a piecewise non-smooth chaotic map. It covers the entire phase space, and the region is chaotic. The reason for getting better Pareto solutions from BMOChOA16 may be due to the ability of the tent map to search non-repeatedly all states within a specific range. This aids the BMOChOA16 to escape from local minima and reach the global best faster. In addition, we have used the Wilcoxon signed rank test to verify whether the BMOChOA16 is significantly superior to the other variants. The performance of the two approaches is significantly different if the P-value obtained from the Wilcoxon signed rank test is less than 0.05 and similar otherwise. The outputs of the Wilcoxon signed rank test along with the P-values are given in TABLE 4. The P-values of more than 0.05 are highlighted in the TABLE 4. The triplet (b, e, w) indicates that BMOChOA16 is significantly better, equal to, or worse than the other BMOChOA variants. By analyzing the entries in TABLE 4, we concluded that BMOChOA16 is significantly superior to the other versions. Therefore, in the following part of this article, we have used BMOChOA16 for comparison.

B. PERFORMANCE COMPARISON BETWEEN BMOChOA16, MOGA, MOPSO, AND NSGA-II

FIGURE 4 and 5 is for the visualization and comparison of Pareto fronts obtained by the four above discussed multi-objective methods for the datasets D1-D10 and D11-D14 respectively. The parato fronts produced by the proposed BMOChOA16 for Lymphography, Diabetic, Cardiotocography, Cervical Cancer, Lung Cancer, Arrhythmia, Parkinson, Colon tumor, and Leukemia data are present above the fronts of the other three benchmark approaches, indicating their closeness to the respective TFSs. In the case of the Cardiotocography dataset, there are many common members present in the two non-dominated solution sets output by BMOChOA16 and MOPSO. Similarly, for SRBCT samples, the performance of MOGA is also very close to the presented approach in producing the non-dominated solutions. The Pareto front for the LungCancer dataset has only one solution, but that solution is superior to others. For arrhythmia, Parkinson, and colon tumour data, there is a wide gap between the fronts of BMOChOA16 and the same of the other three methods, which indicates the supreme efficiency of the proposed approach over other standard methods.

All four methods are executed for 50 iterations, and after completing them, the CD measure is applied to pick the best
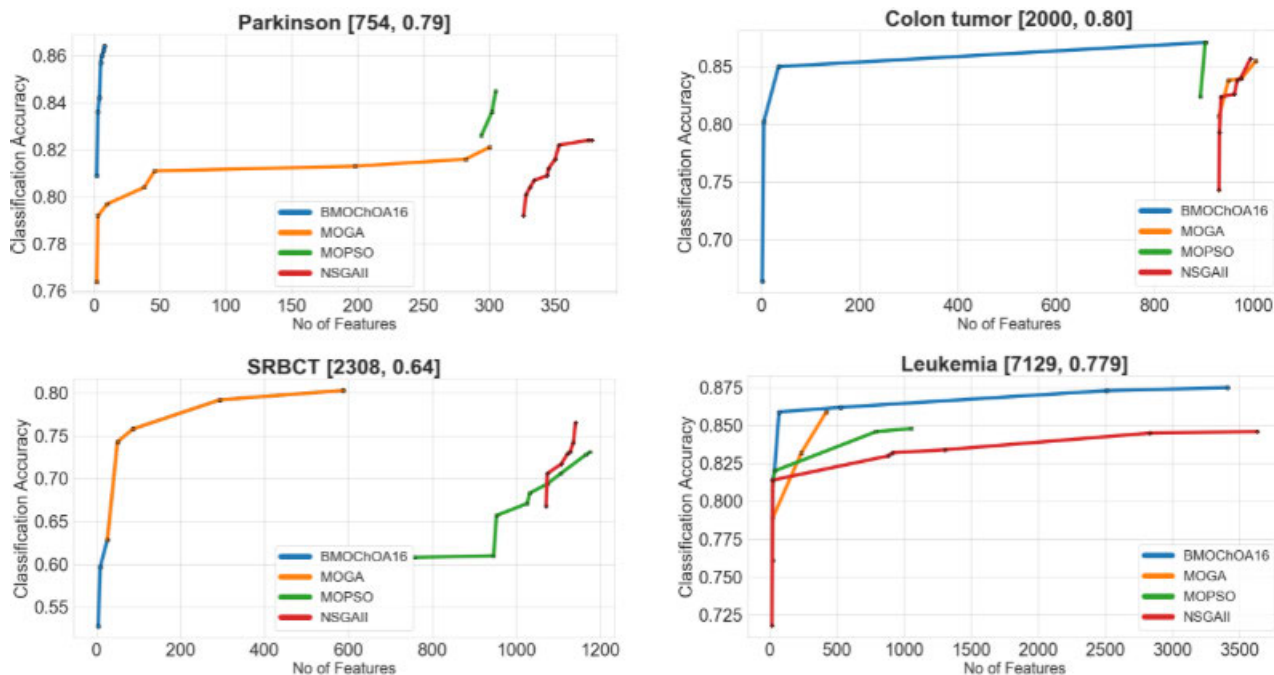
FIGURE 5. Pareto fronts of four methods for D11 to D14.

TABLE 5. List of solutions selected by CD measure.

Datasets	Before FS	After FS			
		MOGA	MOPSO	NSGA-II	BMOChOA16
D1	(9,0.96)	(2,0.95)	(6,0.97)	(2,0.95)	(4,0.969)
D2	(10,0.62)	(3,0.727)	(3,0.727)	(3,0.727)	(3,0.727)
D3	(17,0.16)	(2,0.247)	(6,0.321)	(2,0.247)	(3,0.206)
D4	(18,0.79)	(2,0.796)	(2,0.763)	(2,0.796)	(2,0.796)
D5	(19,0.63)	(2,0.688)	(4,0.648)	(6,0.711)	(2,0.688)
D6	(19,0.81)	(2,0.843)	(4,0.857)	(2,0.745)	(2,0.831)
D7	(21,0.85)	(11,0.879)	(8,0.881)	(2,0.786)	(8,0.881)
D8	(35,0.94)	(15,0.967)	(2,0.936)	(5,0.964)	(2,0.962)
D9	(56,0.22)	(16,0.358)	(16,0.358)	(11,0.658)	(2,0.675)
D10	(279,0.58)	(110,0.622)	(92,0.674)	(136,0.64)	(4,0.701)
D11	(754,0.79)	(300,0.821)	(305,0.845)	(326,0.792)	(2,0.809)
D12	(2000,0.8)	(931,0.807)	(903,0.871)	(931,0.743)	(903,0.871)
D13	(2308,0.64)	(588,0.803)	(1176,0.731)	(1141,0.765)	(588,0.803)
D14	(7129,0.779)	(421,0.859)	(1055,0.848)	(16,0.718)	(19,0.761)

out of the best from the least gathered area of the primary archive. The TABLE 5 is for listing the Pareto solutions selected by the CD value for all the 14 datasets. In the BreastCancerW data, CD based BMOChOA16 selected the best one, which gives 96.9% accuracy by taking less than 50% of the original features. However, the best solution of MOPSO gives 0.1% more accuracy than BMOChOA16 at the expense of 2 additional attributes. In the case of lung cancer data, the performance of the CD-based BMOChOA16 is excellent because it produced the best Pareto solution, which is able to classify the samples with an accuracy that is 45.5% more than the actual one by considering only 2 out of 56 features. A brilliant solution is obtained from the proposed method for the Arrhythmia dataset, which is able to give more than 12.1% classification accuracy than the original by focusing on only 1.4% of the actual feature set. For colon tumour and cardiotocography samples, both MOPSO and

BMOChOA16 have shown equal performance in selecting the best of the primary archive, and that is quite satisfactory. Similarly, the achievements of MOGA and BMOChOA16 are the same in the cases of SRBCT, diabetic, and lymphography data. In the Parkinson's dataset, BMOCHOA16 picked a solution which contains 2% more accuracy than the original by using only 2 out of 754 features.

The TABLE 6 and 7 have recorded the statistical evaluation results with respect to the count of features selected and the corresponding accuracy figure produced by BMOChOA16 and other three very famous algorithms, respectively. It is found that the BMOChOA16 is able to give higher average classification accuracy by using fewer features in the Lymphography, Diabetic, LungCancer, Parkinsons, and Leukemia datasets. All the methods have shown equal performance when considering the ILPD data. For the BreastCancerW dataset, the results are up to the mark

TABLE 6. Statistics about how many features were selected.

Datasets	Statistical Measures	BMOChOA16	MOGA	MOPSO	NSGA-II
BreastCancerW	Min	2	2	2	2
	Max	4	7	6	7
	Range	2	5	4	5
	Avg	3	4.4	3.75	4.66
	STD	1	2.074	1.707	1.966
ILPD	Min	2	2	2	2
	Max	3	3	3	3
	Range	1	1	1	1
	Avg	2.5	2.5	2.5	2.5
	STD	0.707	0.707	0.707	0.707
PrimaryTumor	Min	3	2	2	2
	Max	9	5	6	4
	Range	6	3	4	2
	Avg	6.2	2.66	4.25	2.6
	STD	2.38	1.21	1.707	0.894
Lymphography	Min	2	2	2	2
	Max	6	12	10	11
	Range	4	10	8	9
	Avg	4	7.166	5.5	6.66
	STD	1.58	4.07	3.271	0.014
Cardiotocography	Min	2	4	3	2
	Max	8	11	8	13
	Range	6	7	5	11
	Avg	4.6	7.6	5.8	6.77
	STD	2.408	3.049	1.923	3.8
Diabetic	Min	2	2	3	2
	Max	5	9	7	6
	Range	3	7	4	4
	Avg	3.5	4.25	4.66	3.75
	STD	1.29	3.20	2.08	1.707
Hepatitis	Min	2	2	3	2
	Max	6	4	5	11
	Range	4	2	2	9
	Avg	3.8	3	4	5.66
	STD	1.643	1	1	3.55
Cervical Cancer	Min	2	3	2	2
	Max	8	15	6	5
	Range	6	12	4	3
	Avg	4.66	10.428	4	3
	STD	3.05	4.23	1.581	1.73
LungCancer	Min	2	16	16	9
	Max	2	25	25	11
	Range	0	9	9	2
	Avg	2	19.5	18.6	10
	STD	-	3.271	3.646	1
Arrhythmia	Min	2	110	77	110
	Max	4	142	92	136
	Range	2	32	15	26
	Avg	3	126.5	84	122.33
	STD	1	13.304	6.164	9.852
Parkinsons	Min	2	2	294	326
	Max	8	300	305	378
	Range	6	298	11	52
	Avg	5	109.875	300.33	346.4
	STD	2.16	128.63	5.686	18.33
Colon tumor	Min	2	931	893	931
	Max	903	1006	903	995
	Range	901	75	10	64
	Avg	236.25	966	896.33	956.857
	STD	0.093	0.02	5.77	24.828
SRBCT	Min	4	26	757	1071
	Max	588	588	1176	1141
	Range	584	562	419	70
	Avg	151.14	209	1026.22	1111.28
	STD	217.22	236.91	129.869	28.633
Leukemia	Min	19	19	20	16
	Max	3409	421	1055	3632
	Range	3390	402	1035	3616
	Avg	941.28	174.25	475.5	1371.71
	STD	1413.61	193.096	528.366	1375.25

TABLE 7. Findings of statistical analysis on the four algorithms' classification accuracy.

Datasets	Statistical Measures	BMOChOA16	MOGA	MOPSO	NSGA-II
BreastCancerW	Min	0.95	0.95	0.95	0.95
	Max	0.969	0.971	0.97	0.971
	Range	0.019	0.021	0.02	0.021
	Avg	0.96	0.96	0.963	0.963
	STD	0.009	0.008	0.009	0.008
ILPD	Min	0.715	0.715	0.715	0.715
	Max	0.727	0.727	0.727	0.727
	Range	0.012	0.012	0.012	0.012
	Avg	0.721	0.721	0.721	0.721
	STD	0.008	0.008	0.008	0.008
PrimaryTumor	Min	0.206	0.247	0.247	0.247
	Max	0.268	0.285	0.321	0.291
	Range	0.062	0.038	0.074	0.044
	Avg	0.24	0.259	0.283	0.264
	STD	0.024	0.018	0.03	0.023
Lymphography	Min	0.796	0.796	0.763	0.796
	Max	0.852	0.843	0.845	0.825
	Range	0.056	0.047	0.082	0.029
	Avg	0.825	0.823	0.81	0.812
	STD	0.024	0.019	0.029	0.014
Cardiotocography	Min	0.843	0.843	0.869	0.786
	Max	0.881	0.879	0.881	0.875
	Range	0.038	0.036	0.012	0.089
	Avg	0.868	0.864	0.875	0.846
	STD	0.015	0.014	0.005	0.031
Diabetic	Min	0.688	0.688	0.553	0.599
	Max	0.711	0.695	0.687	0.711
	Range	0.023	0.007	0.134	0.112
	Avg	0.7	0.691	0.629	0.658
	STD	0.01	0.002	0.068	0.052
Hepatitis	Min	0.831	0.843	0.851	0.745
	Max	0.876	0.87	0.876	0.883
	Range	0.045	0.027	0.025	0.138
	Avg	0.85	0.852	0.861	0.84
	STD	0.017	0.015	0.013	0.05
Cervical Cancer	Min	0.962	0.936	0.936	0.962
	Max	0.967	0.967	0.965	0.964
	Range	0.005	0.031	0.029	0.002
	Avg	0.964	0.952	0.951	0.962
	STD	0.002	0.009	0.01	0.001
LungCancer	Min	0.675	0.358	0.358	0.508
	Max	0.675	0.6	0.608	0.658
	Range	0	0.242	0.25	0.15
	Avg	0.675	0.473	0.46	0.597
	STD	-	0.088	0.1	0.078
Arrhythmia	Min	0.582	0.622	0.648	0.586
	Max	0.701	0.64	0.674	0.64
	Range	0.119	0.018	0.026	0.054
	Avg	0.623	0.63	0.664	0.619
	STD	0.067	0.007	0.012	0.018
Parkinsons	Min	0.809	0.764	0.826	0.792
	Max	0.864	0.821	0.845	0.824
	Range	0.055	0.057	0.019	0.032
	Avg	0.847	0.802	0.835	0.811
	STD	0.019	0.018	0.009	0.01
Colon tumor	Min	0.664	0.807	0.824	0.743
	Max	0.871	0.855	0.871	0.857
	Range	0.207	0.048	0.047	0.114
	Avg	0.796	0.835	0.839	0.817
	STD	0.093	0.02	0.027	0.038
SRBCT	Min	0.528	0.629	0.608	0.668
	Max	0.803	0.803	0.731	0.765
	Range	0.275	0.174	0.123	0.097
	Avg	0.692	0.745	0.676	0.722
	STD	0.107	0.069	0.045	0.03
Leukemia	Min	0.761	0.761	0.807	0.718
	Max	0.875	0.859	0.848	0.846
	Range	0.114	0.098	0.041	0.128
	Avg	0.837	0.81	0.83	0.817
	STD	0.041	0.043	0.02	0.044

TABLE 8. Multi-objective performance measures.

Datasets	MOGA				MOPSO				NSGA-II				BMOChOA16			
	IGD	HV	Spread	SCC	IGD	HV	Spread	SCC	IGD	HV	Spread	SCC	IGD	HV	Spread	SCC
D1	0	0.202	0.681	5	0.057	0.241	0.737	4	0.046	0.171	0.603	2	0	0.202	0.681	5
D2	0	0.351	0.645	2	0	0.351	0.645	2	0	0.351	0.645	2	0	0.351	0.645	2
D3	0.04	0.176	0.651	1	0.043	0.266	0.775	2	0.026	0.205	0.612	3	0.008	0.217	0.7	0
D4	0.037	0.178	0.696	1	0.038	0.178	0.72	0	0.147	0.334	0.961	1	0	0.204	0.654	5
D5	0.086	0.245	0.91	1	0.082	0.312	0.845	0	0.026	0.254	0.786	0	0.001	0.241	0.667	3
D6	0.147	0.295	0.682	1	0.158	0.286	0.594	0	0.04	0.179	0.764	0	0.029	0.202	0.665	0
D7	0.04	0.2	0.62	0	0.03	0.16	0.55	4	0.04	0.10	0.57	0	0.02	0.17	0.66	5
D8	0.269	0.152	0.581	0	0.189	0.201	0.642	0	0.053	0.292	0.817	0	0	0.318	0.928	2
D9	1.114	0.199	0.569	0	1.039	0.238	0.639	0	0.895	0.29	0.548	0	0	0.1	0	0
D10	0.118	0.222	0.375	0	0.121	0.219	0.328	0	0.227	0.162	0.315	0	0.015	0.314	0.756	1
D11	0.043	0.139	0.834	0	0.265	0.26	0.199	0	0.071	0.101	0.195	0	0	0.15	0.601	6
D12	0.307	0.219	0.26	0	0.396	0.265	0.212	1	0.232	0.143	0.243	0	0.06	0.292	1.427	4
D13	0.73	0.226	1.081	5	0.106	0.121	0.362	0	0.144	0.144	0.237	0	0.025	0.176	1.103	7
D14	0.155	0.274	1.102	1	0.159	0.274	1.055	2	0.049	0.16	0.893	2	0.048	0.162	1.033	7

because the MOPSO and NSGA-II produce only 0.3% more average accuracy at the expense of 0.75 and 1.66 average number of attributes, respectively. Similarly, the proposed method has shown satisfactory output for cardiotocography data by giving a lower average number of features while compromising with only 0.7% less classification accuracy value than MOPSO. For arrhythmia data, our suggested approach gave an accuracy value of 0.623 with an average of only 3 features (1.07% of the original dimension). However, the MOPSO is able to produce 4.1% more average accuracy than BMOChOA16 by including 78 extra features. With respect to the high dimensional dataset like colon tumor, the BMOChOA16 is able to classify the samples with an accuracy of 0.796 with the help of only 12% of the original number of attributes. However, MOPSO has an increased result of 4.3% more accuracy than BMOChOA16 by taking 45% of the original width. In the case of the SRBCT dataset, MOGA gave an excellent average accuracy value compared to that of the suggested approach, at the expense of 3% more features. The population containing the number of chosen features by the presented method has a lower standard deviation than others, indicating the closeness of the Pareto solutions towards the mean one with respect to the objective function one. From the above analysis and discussion, it can be concluded that the proposed approach can efficiently generate a set of non-dominated solutions in most of the datasets by discarding the irrelevant attributes.

C. ANALYSIS OF CONVERGENCE CHARACTERISTICS

Both the diversity of the final estimate of the Pareto front (CLF) and the convergence to the true Pareto set of a MOP are two key concerns in evolutionary multi-objective optimization (EMO). In this article, we have used IGD for convergence proof, Spread for distribution analysis, HV for both convergence and distribution analysis, and SCC to quantify the contribution of the four multi-objective methods. After the end of the 50 repetitions of all the FS strategies, the performance measures for the Pareto fronts shown in FIGURE 4 and 5 are recorded in Table 8. As the IGD values of the obtained Pareto fronts by the proposed approach are less than the others in all the datasets, they are very close to their respective TPFs. Also, the non-dominated solution

sets resulting from BMOChOA16 for all the 14 datasets are quite well scattered as their spread and HV values are attractive. The SCC values of the CLFs by the suggested approach are high as compared to others in the case of most of the datasets, indicating their larger contribution towards getting the TPFs. We have also conducted a comparative study on the IGD values of the Pareto fronts achieved from 25 separate executions of BMOChOA16, MOGA, MOPSO, and NSGA-II by applying the Wilcoxon signed rank test. The statistical results of this test by taking a 0.05 significance level are listed in TABLE 9. In TABLE 9 the '++', '- -', and '==' for any column Approach1-Approach2 denote whether the Approach 1 is statistically superior to, statistically worse than, or significantly same as the Approach2 respectively. Entries of the TABLE 9 indicate that the efficiency of the proposed approach is significantly better than others for most datasets, except ILPD and cardiotocography, where it is equal. Finally, we can conclude that the presented BMOChOA16 with Tent Chaotic Map outperforms others in terms of optimizing critical aspects of healthcare data.

D. RUNNING TIME COMPARISON

In order to record the average running time of the four above-discussed methods in TABLE 10, they are executed 25 times separately for each of the 14 datasets. By observing the entries of the TABLE 10, it is found that for most of the datasets like BreastCancerW (5.26), ILPD (5.12), PrimaryTumor (3.34), Diabetic (7.17), Cardiotocography (9.45), Cervical Cancer (2.14), Arrhythmia (7.59), Parkinsons (14.23), and Leukemia (17.02) datasets, the proposed FS method took less time to complete 50 iterations. Furthermore, for many datasets, MOPSO and NSGA-II took longer to execute than others, because in MOPSO, both position and velocity are updated for each particle, whereas in NSGA-II, two populations are combined and different fronts are computed for the next loop at each iteration. In MOChOA, either exploration or exploitation is performed for each chimp according to the value of μ. This may be one of the causes of its lower execution time. However, in the case of the other three methods, both exploration and exploitation are carried out for each solution of the population.

TABLE 9. Results of wilcoxon signed rank test on IGD values.

Datasets	BMOChOA16-MOGA	BMOChOA16-MOPSO	BMOChOA16-NSGA-II
D1	==	++	++
D2	==	==	==
D3	++	++	++
D4	++	++	++
D5	++	++	++
D6	++	++	==
D7	==	==	==
D8	++	++	++
D9	++	++	++
D10	++	++	++
D11	++	++	++
D12	++	++	++
D13	++	++	++
D14	++	==	++

TABLE 10. Average running time (mins).

Datasets	MOGA	MOPSO	NSGA-II	BMOChOA16
D1	6.15	6.51	12.42	5.26
D2	5.45	8.46	11.65	5.12
D3	4.18	4.23	4.36	3.34
D4	4.01	3.56	7.23	4.45
D5	9.21	11.34	18.04	7.17
D6	4.59	3.14	8.34	4.62
D7	14.25	13.12	23.37	9.45
D8	4.09	5.14	5.59	2.14
D9	3.45	4.01	7.37	3.45
D10	12.23	13.56	25.37	7.59
D11	21.23	22.02	24.36	14.23
D12	5.47	5.56	10.33	5.49
D13	11.12	15.56	16.31	13.23
D14	23.30	26.10	27.55	17.02

E. MERITS OF THE PROPOSED FS METHOD

By doing a vigilant revision of all the experimental outcomes brought up in the previous subsections, it is observed that the suggested BMOChOA16 method for FS could be an efficient member in the domain of FS for discarding unnecessary aspects from healthcare samples.

- The offered FS approach, BMOChOA16, is found to be excellent in generating the top Pareto fronts concerning both the objectives as compared to the other three benchmark multi-objective approaches.
- The most distinct characteristic of the proposed method is that it is able to produce a repository of Pareto solutions, which are very unique because of the crowding distance measure.
- For classification tasks in FS, we used KNN as a wrapper because it is one of the finest supervised classifiers and has the lowest computational cost.
- The comparative experimental outcomes of all the four FS methods disclose that the proposed BMOChOA16 with Tent chaotic map is able to reach higher classification accuracy by considering low-sized feature subsets in less time.
- The rate of convergence of the proposed method is high as compared to others due to the lower IGD value of the CLFs in all the datasets. Also, the CLFs of the BMOChOA16 are able to dominate a larger area in the objective space due to the higher spread values.

TABLE 11. Dynamic co-efficient (f) Values.

Groups	ChOA1	ChOA2
1	$1.95 - 2CI^{1/4}/MI^{1/3}$	$2.5 - (2log(CI)/log(MI))$
2	$1.95 - 2CI^{1/3}/MI^{1/4}$	$2.5 + (-2CI^3/MI^3)$
3	$1.5 + (-3CI^3/MI^3)$	$0.5 + 2\exp[-(4CI/MI)^2]$
4	$1.5 + (-2CI^3/MI^3)$	$2.5 + 2(CI/MI)^2 - 2(2CI/MI)$

TABLE 12. Details of chaotic maps.

SNO	NAME	CHAOTIC MAP
1	Quadratic	$p_{i+1} = p_i^2 - c, \quad c = 1$
2	Gauss	$p_{i+1} = \begin{cases} 1 & \text{if } p_i = 0 \\ \frac{1}{mod(p_i,1)} & \text{otherwise} \end{cases}$
3	Logistic	$p_{i+1} = \alpha p_i(1 - p_i), \quad \alpha = 4$
4	Singer	$p_{i+1} = \mu(7.86p_i - 23.31p_i^2 + 28.75p_i^3 - 13.302875p_i^4), \quad \mu = 1.07$
5	Bernoulli	$p_{i+1} = 2p_i(mod\, 1)$
6	Tent	$p_{i+1} = \begin{cases} \frac{p_i}{0.7}, & \text{if } p_i < 0.7 \\ \frac{10}{3}(1 - p_i) & \text{if } p_i \geq 0.7 \end{cases}$

VI. CONCLUSION AND FUTURE WORKS

A binary multi-objective chimp optimization algorithm is introduced in this study for the first time to handle the bi-objective FS task in the healthcare domain. To establish the method as a strong competitor for the FS purpose, we have taken six chaotic maps with two variants of BMOChOA, resulting in 12 distinct versions of the suggested approach. The results of all the 12 BMOChOAs are compared, analyzed, and then BMOChOA16 with Tent chaotic map is statistically selected as the best one in identifying the non-dominated solution set which is closer to their TPFs. Finally, to verify the robustness of the offered FS method, it is compared with three well-known multi-objective FS algorithms, namely, MOGA, MOPSO, and NSGA-II. BMOChOA16 has proved to be the supreme method in terms of achieving fewer feature sizes and high classification accuracy when compared with the other three popular methods. The best quality Pareto fronts are obtained from the proposed method in less time for most of the datasets and are verified by using several multi-objective performance evaluation conditions. The soundness of the proposed BMOChOA16 is again statistically verified by performing a Wilcoxon signed rank test on the IGD values of the CLFs by all four methods.

Here we have considered the FS problem as a bi-objective wrapper-based optimization task. However, scalability and computational complexity can be taken into consideration. Also, the proposed wrapper method can be enhanced for the hybrid filter-wrapper optimization task by focusing on the mutual information and correlation among the features and target attribute. In this article, to choose the best of the best from the primary archive, we have applied the CD measure. However, other criteria, such as knee point, can be used for the same purpose. The efficiency of the proposed BMOChOAs is verified using only healthcare data. However, it can be applied to several other optimization tasks in the real world.

APPENDIX A
DYNAMIC COEFFICIENT
See Table 11.

TABLE 13. Naming style of BMOChOAs.

Name	Type	Chaotic map
BMOChOA11	BMOChOA1	Quadratic
BMOChOA12	BMOChOA1	Gauss
BMOChOA13	BMOChOA1	Logistic
BMOChOA14	BMOChOA1	Singer
BMOChOA15	BMOChOA1	Bernoulli
BMOChOA16	BMOChOA1	Tent
BMOChOA21	BMOChOA2	Quadratic
BMOChOA22	BMOChOA2	Gauss
BMOChOA23	BMOChOA2	Logistic
BMOChOA24	BMOChOA2	Singer
BMOChOA25	BMOChOA2	Bernoulli
BMOChOA26	BMOChOA2	Tent

APPENDIX B
CHAOTIC MAPS
See Table 12.

APPENDIX C
NAMING OF BMOChOAS
See Table 13.

REFERENCES

[1] E. Tuba, I. Strumberger, T. Bezdan, N. Bacanin, and M. Tuba, "Classification and feature selection method for medical datasets by brain storm optimization algorithm and support vector machine," *Proc. Comput. Sci.*, vol. 162, pp. 307–315, Jan. 2019.

[2] R. Samant and S. Rao, "A study on feature selection methods in medical decision support systems," *Int. J. Eng. Res. Technol.*, vol. 2, no. 11, pp. 615–619, 2013.

[3] B. Nouri-Moghaddam, M. Ghazanfari, and M. Fathian, "A novel multi-objective forest optimization algorithm for wrapper feature selection," *Expert Syst. Appl.*, vol. 175, Aug. 2021, Art. no. 114737.

[4] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A survey on evolutionary computation approaches to feature selection," *IEEE Trans. Evol. Comput.*, vol. 20, no. 4, pp. 606–626, Aug. 2016.

[5] E. Amaldi and V. Kann, "On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems," *Theor. Comput. Sci.*, vol. 209, nos. 1–2, pp. 237–260, 1998.

[6] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Elect. Eng.*, vol. 40, no. 1, pp. 16–28, Jan. 2014.

[7] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, nos. 1–2, pp. 273–324, 1997.

[8] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, May 2003.

[9] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 4, pp. 491–502, Apr. 2005.

[10] A. W. Whitney, "A direct method of nonparametric measurement selection," *IEEE Trans. Comput.*, vol. C–20, no. 9, pp. 1100–1103, Sep. 1971.

[11] T. Marill and D. Green, "On the effectiveness of receptors in recognition systems," *IEEE Trans. Inf. Theory*, vol. IT-9, no. 1, pp. 11–17, Jan. 1963.

[12] B. Xue, M. Zhang, and W. N. Browne, "Particle swarm optimization for feature selection in classification: A multi-objective approach," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1656–1671, Dec. 2013.

[13] I.-S. Oh, J.-S. Lee, and B.-R. Moon, "Hybrid genetic algorithms for feature selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1424–1437, Nov. 2004.

[14] L. S. Oliveira, R. Sabourin, F. Bortolozzi, and C. Y. Suen, "Feature selection using multi-objective genetic algorithms for handwritten digit recognition," in *Proc. Object Recognit. Supported User Interact. Service Robots*, 2002, pp. 568–571.

[15] J. Piri, P. Mohapatra, and R. Dey, "Fetal health status classification using MOGA–CD based feature selection approach," in *Proc. IEEE Int. Conf. Electron., Comput. Commun. Technol. (CONECCT)*, Jul. 2020, pp. 1–6.

[16] A. Kharrat, K. Gasmi, M. Ben Messaoud, N. Benamrane, and M. Abid, "Medical image classification using an optimal feature extraction algorithm and a supervised classifier technique," *Int. J. Softw. Sci. Comput. Intell.*, vol. 3, no. 2, pp. 19–33, Apr. 2011.

[17] Q. Al-Tashi, S. J. Abdulkadir, H. M. Rais, S. Mirjalili, H. Alhussian, M. G. Ragab, and A. Alqushaibi, "Binary multi-objective grey wolf optimizer for feature selection in classification," *IEEE Access*, vol. 8, pp. 106247–106263, 2020.

[18] J. Too, A. R. Abdullah, and N. M. Saad, "A new quadratic binary Harris hawk optimization for feature selection," *Electronics*, vol. 8, no. 10, p. 1130, Oct. 2019.

[19] Y. Zhang, R. Liu, X. Wang, H. Chen, and C. Li, "Boosted binary Harris hawks optimizer and feature selection," *Eng. With Comput.*, vol. 37, pp. 3741–3770, May 2020.

[20] J. Piri and P. Mohapatra, "An analytical study of modified multi-objective Harris hawk optimizer towards medical data feature selection," *Comput. Biol. Med.*, vol. 135, Aug. 2021, Art. no. 104558.

[21] J. Too and S. Mirjalili, "A hyper learning binary dragonfly algorithm for feature selection: A COVID-19 case study," *Knowl.-Based Syst.*, vol. 212, Jan. 2021, Art. no. 106553.

[22] H. Das, B. Naik, and H. S. Behera, "A Jaya algorithm based wrapper method for optimal feature selection in supervised classification," *J. King Saud Univ. Comput. Inf. Sci.*, to be published.

[23] B. Niu, W. Yi, L. Tan, S. Geng, and H. Wang, "A multi-objective feature selection method based on bacterial foraging optimization," *Natural Comput.*, vol. 20, no. 1, pp. 63–76, Mar. 2021.

[24] G. R. Kancharla, N. R. Eluri, S. Dara, and N. Ansari, "An efficient algorithm for feature selection problem in gene expression data: A spider monkey optimization approach," in *Proc. 2nd Int. Conf. Adv. Comput. Softw. Eng. (ICACSE)*, Mar. 2019, pp. 1–4.

[25] V. Kumar and A. Kaur, "Binary spotted hyena optimizer and its application to feature selection," *J. Ambient Intell. Humanized Comput.*, vol. 11, no. 7, pp. 2625–2645, Jul. 2020.

[26] A. E. Hegazy, M. A. Makhlouf, and G. S. El-Tawel, "Feature selection using chaotic salp swarm algorithm for data classification," *Arabian J. Sci. Eng.*, vol. 44, no. 4, pp. 3801–3816, Apr. 2019.

[27] J. Piri, P. Mohapatra, and R. Dey, "Multi-objective ant lion optimization based feature retrieval methodology for investigation of fetal wellbeing," in *Proc. 3rd Int. Conf. Inventive Res. Comput. Appl. (ICIRCA)*, Sep. 2021, pp. 1732–1737.

[28] M. Allam and M. Nandhini, "Optimal feature selection using binary teaching learning based optimization algorithm," *J. King Saud Univ. Comput. Inf. Sci.*, to be published.

[29] T. M. Hamdani, J.-M. Won, A. M. Alimi, and F. Karray, "Multi-objective feature selection with NSGA II," in *Proc. Int. Conf. Adapt. Natural Comput. Algorithms*. Berlin, Germany: Springer, 2007, pp. 240–247.

[30] E. Hancer, B. Xue, M. Zhang, D. Karaboga, and B. Akay, "Pareto front feature selection based on artificial bee colony optimization," *Inf. Sci.*, vol. 422, pp. 462–479, 2017.

[31] M. Khishe and M. R. Mosavi, "Chimp optimization algorithm," *Expert Syst. Appl.*, vol. 149, Jul. 2020, Art. no. 113338.

[32] H. Jia, K. Sun, W. Zhang, and X. Leng, "An enhanced chimp optimization algorithm for continuous optimization domains," *Complex Intell. Syst.*, pp. 1–18, Apr. 2021.

[33] C. Boesch, "Cooperative hunting roles among taï chimpanzees," *Hum. Nature*, vol. 13, no. 1, pp. 27–46, Mar. 2002.

[34] S. Mirjalili, A. Lewis, and A. S. Sadiq, "Autonomous particles groups for particle swarm optimization," *Arabian J. Sci. Eng.*, vol. 39, no. 6, pp. 4683–4697, 2014.

[35] P. M. Narendra and K. Fukunaga, "A branch and bound algorithm for feature subset selection," *IEEE Trans. Comput.*, vol. C-26, no. 9, pp. 917–922, Sep. 1977.

[36] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, 1994.

[37] S. C. Yusta, "Different Metaheuristic strategies to solve the feature selection problem," *Pattern Recognit. Lett.*, vol. 30, no. 5, pp. 525–534, Apr. 2009.

[38] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Machine Learning Proceedings*. Amsterdam, The Netherlands: Elsevier, 1992, pp. 249–256.

[39] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[40] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Trans. Neural Netw.*, vol. 5, no. 4, pp. 537–550, Jul. 1994.

[41] Q. Al-Tashi, S. J. A. Kadir, H. M. Rais, S. Mirjalili, and H. Alhussian, "Binary optimization using hybrid grey wolf optimization for feature selection," *IEEE Access*, vol. 7, pp. 39496–39508, 2019.

[42] A. E. Hegazy, M. Makhlouf, and G. S. El-Tawel, "Improved salp swarm algorithm for feature selection," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 32, no. 2, pp. 335–344, 2018.

[43] R. Dash, "An adaptive harmony search approach for gene selection and classification of high dimensional medical data," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 33, no. 2, pp. 195–207, Feb. 2021.

[44] H. Hichem, M. Elkamel, M. Rafik, M. T. Mesaaoud, and C. Ouahiba, "A new binary grasshopper optimization algorithm for feature selection problem," *J. King Saud Univ. Comput. Inf. Sci.*, to be published.

[45] P. D. Sheth, S. T. Patil, and M. L. Dhore, "Evolutionary computing for clinical dataset classification using a novel feature selection algorithm," *J. King Saud Univ. Comput. Inf. Sci.*, to be published.

[46] J. Too and A. R. Abdullah, "Chaotic atom search optimization for feature selection," *Arabian J. Sci. Eng.*, vol. 45, no. 8, pp. 6063–6079, Aug. 2020.

[47] A. C. Pandey, D. S. Rajpoot, and M. Saraswat, "Feature selection method based on hybrid data transformation and binary binomial cuckoo search," *J. Ambient Intell. Humanized Comput.*, vol. 11, no. 2, pp. 719–738, Feb. 2020.

[48] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary ant lion approaches for feature selection," *Neurocomputing*, vol. 213, pp. 54–65, Nov. 2016.

[49] M. Ghaemi and M.-R. Feizi-Derakhshi, "Feature selection using forest optimization algorithm," *Pattern Recognit.*, vol. 60, pp. 121–129, Dec. 2016.

[50] S. A.-F. Sayed, E. Nabil, and A. Badr, "A binary clonal flower pollination algorithm for feature selection," *Pattern Recognit. Lett.*, vol. 77, pp. 21–27, Jul. 2016.

[51] M. M. Mafarja and S. Mirjalili, "Hybrid whale optimization algorithm with simulated annealing for feature selection," *Neuro Comput.*, vol. 260, pp. 302–312, Oct. 2017.

[52] Y. Zheng, Y. Li, G. Wang, Y. Chen, Q. Xu, J. Fan, and X. Cui, "A novel hybrid algorithm for feature selection based on whale optimization algorithm," *IEEE Access*, vol. 7, pp. 14908–14923, 2018.

[53] D. H. Wolper and W. G. Macready, "No free lunch theorems for optimization," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 67–82, Apr. 1997.

[54] M.-Y. Cheng and D. Prayogo, "Symbiotic organisms search: A new metaheuristic optimization algorithm," *Comput. Struct.*, vol. 139, pp. 98–112, Jul. 2014.

[55] N. Singh, L. H. Son, F. Chiclana, and J.-P. Magnot, "A new fusion of salp swarm with sine cosine for optimization of non-linear functions," *Eng. With Comput.*, vol. 36, no. 1, pp. 185–212, Jan. 2020.

[56] M. G. Bindu and M. K. Sabu, "A hybrid feature selection approach using artificial bee colony and genetic algorithm," in *Proc. Adv. Comput. Commun. Technol. High Perform. Appl. (ACCTHPA)*, Jul. 2020, pp. 211–216.

[57] T. Bhattacharyya, B. Chatterjee, P. K. Singh, J. H. Yoon, Z. W. Geem, and R. Sarkar, "Mayfly in harmony: A new hybrid meta-heuristic feature selection algorithm," *IEEE Access*, vol. 8, pp. 195929–195945, 2020.

[58] S. Meera and C. Sundar, "A hybrid Metaheuristic approach for efficient feature selection methods in big data," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 3, pp. 3743–3751, Mar. 2021.

[59] H. Jia, Z. Xing, and W. Song, "A new hybrid seagull optimization algorithm for feature selection," *IEEE Access*, vol. 7, pp. 49614–49631, 2019.

[60] S. Arora, H. Singh, M. Sharma, S. Sharma, and P. Anand, "A new hybrid algorithm based on Grey wolf optimization and crow search algorithm for unconstrained function optimization and feature selection," *IEEE Access*, vol. 7, pp. 26343–26361, 2019.

[61] S. P. Rajamohana and K. Umamaheswari, "Hybrid approach of improved binary particle swarm optimization and shuffled frog leaping for feature selection," *Comput. Electr. Eng.*, vol. 67, pp. 497–508, Apr. 2018.

[62] R. Hans and H. Kaur, "Hybrid binary sine cosine algorithm and ant lion optimization (SCALO) approaches for feature selection problem," *Int. J. Comput. Mater. Sci. Eng.*, vol. 9, no. 1, Mar. 2020, Art. no. 1950021.

[63] S. Thawkar, "A hybrid model using teaching–learning-based optimization and Salp swarm algorithm for feature selection and classification in digital mammography," *J. Ambient Intell. Humanized Comput.*, vol. 12, pp. 8793–8808, Jan. 2021.

[64] E. H. Houssein, M. E. Hosney, M. Elhoseny, D. Oliva, W. M. Mohamed, and M. Hassaballah, "Hybrid Harris hawks optimization with cuckoo search for drug design and discovery in chemoinformatics," *Sci. Rep.*, vol. 10, no. 1, pp. 1–22, Dec. 2020.

[65] K. Hussain, N. Neggaz, W. Zhu, and E. H. Houssein, "An efficient hybrid sine-cosine Harris hawks optimization for low and high-dimensional feature selection," *Expert Syst. Appl.*, vol. 176, Aug. 2021, Art. no. 114778.

[66] R. Al-Wajih, S. J. Abdulkadir, N. Aziz, Q. Al-Tashi, and N. Talpur, "Hybrid binary grey wolf with Harris hawks optimizer for feature selection," *IEEE Access*, vol. 9, pp. 31662–31677, 2021.

[67] M. Khamees and A. A.-B. Rashed, "Hybrid SCA-CS optimization algorithm for feature selection in classification problems," *AIP Conf. Proc.*, vol. 2290, no. 1, 2020, Art. no. 040001.

[68] G. Dhiman, "SSC: A hybrid nature-inspired meta-heuristic optimization algorithm for engineering applications," *Knowl.-Based Syst.*, vol. 222, Jun. 2021, Art. no. 106926.

[69] M. Kaur, R. Kaur, N. Singh, and G. Dhiman, "SChoA: A newly fusion of sine and cosine with chimp optimization algorithm for HLS of datapaths in digital filters and engineering applications," *Eng. With Comput.*, pp. 1–29, Jan. 2021.

[70] C. Emmanouilidis, A. Hunter, J. Macintyre, and C. Cox, "A multi-objective genetic algorithm approach to feature selection in neural and fuzzy modeling," *Evol. Optim.*, vol. 3, no. 1, pp. 1–26, 2001.

[71] Z. Zhu, Y.-S. Ong, and J.-L. Kuo, "Feature selection using single/multi-objective memetic frameworks," in *Multi-Objective Memetic Algorithms*. Berlin, Germany: Springer, 2009, pp. 111–131.

[72] B. Huang, B. Buckley, and T.-M. Kechadi, "Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications," *Expert Syst. Appl.*, vol. 37, no. 5, pp. 3638–3646, May 2010.

[73] L. D. Vignolo, D. H. Milone, and J. Scharcanski, "Feature selection for face recognition based on multi-objective evolutionary wrappers," *Expert Syst. Appl.*, vol. 40, no. 13, pp. 5077–5084, 2013.

[74] B. Xue, W. Fu, and M. Zhang, "Differential evolution (DE) for multi-objective feature selection in classification," in *Proc. Companion Publication Annu. Conf. Genetic Evol. Comput.*, Jul. 2014, pp. 83–84.

[75] A. Khan and A. R. Baig, "Multi-objective feature subset selection using non-dominated sorting genetic algorithm," *J. Appl. Res. Technol.*, vol. 13, no. 1, pp. 145–159, Feb. 2015.

[76] Y. Zhang, D.-W. Gong, X.-Z. Gao, T. Tian, and X.-Y. Sun, "Binary differential evolution with self-learning for multi-objective feature selection," *Inf. Sci.*, vol. 507, pp. 67–85, Jan. 2020.

[77] Y. Zhang, D. W. Gong, and J. Cheng, "Multi-objective particle swarm optimization approach for cost-based feature selection in classification," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 14, no. 1, pp. 64–75, Jan. 2017.

[78] Y. Hu, Y. Zhang, and D. Gong, "Multiobjective particle swarm optimization for feature selection with fuzzy cost," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 874–888, Feb. 2021.

[79] H. B. Nguyen, B. Xue, I. Liu, P. Andreae, and M. Zhang, "New mechanism for archive maintenance in PSO-based multi-objective feature selection," *Soft Comput.*, vol. 20, no. 10, pp. 3927–3946, Oct. 2016.

[80] J. K. Mandal, S. Mukhopadhyay, and P. Dutta, *Multiobjective Optimization: Evolutionary to Hybrid Framework*. Springer, 2018.

[81] B. Ragothaman and B. Sarojini, "A multi-objective non-dominated sorted artificial bee colony feature selection algorithm for medical datasets," *Indian J. Sci. Technol.*, vol. 9, no. 45, pp. 1–5, Dec. 2016.

[82] F. Jiménez, G. Sánchez, J. M. García, G. Sciavicco, and L. Miralles, "Multi-objective evolutionary feature selection for online sales forecasting," *Neurocomputing*, vol. 234, pp. 75–92, Apr. 2017.

[83] A. Sharma and R. Rani, "C-HMOSHSSA: Gene selection for cancer classification using multi-objective meta-heuristic and machine learning methods," *Comput. Methods Programs Biomed.*, vol. 178, pp. 219–235, Sep. 2019.

[84] M. Hammami, S. Bechikh, C.-C. Hung, and L. Ben Said, "A multi-objective hybrid filter-wrapper evolutionary approach for feature selection," *Memetic Comput.*, vol. 11, no. 2, pp. 193–208, Jun. 2019.

[85] S. K. Baliarsingh, S. Vipsita, K. Muhammad, and S. Bakshi, "Analysis of high-dimensional biomedical data using an evolutionary multi-objective emperor penguin optimizer," *Swarm Evol. Comput.*, vol. 48, pp. 262–273, Aug. 2019.

[86] I. Behravan, O. Dehghantanha, and S. H. Zahiri, "An optimal SVM with feature selection using multi-objective PSO," in *Proc. 1st Conf. Swarm Intell. Evol. Comput. (CSIEC)*, Mar. 2016, pp. 76–81.

[87] A. Bouraoui, S. Jamoussi, and Y. Benayed, "A multi-objective genetic algorithm for simultaneous model and feature selection for support vector machines," *Artif. Intell. Rev.*, vol. 50, no. 2, pp. 261–281, Feb. 2017.

[88] B. C. Dos Santos, C. N. Nobre, and L. E. Zarate, "Multi-objective genetic algorithm for feature selection in a protein function prediction context," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jul. 2018, pp. 1–6.

[89] Y. Xue, B. Xue, and M. Zhang, "Self-adaptive particle swarm optimization for large-scale feature selection in classification," *ACM Trans. Knowl. Discovery from Data*, vol. 13, no. 5, pp. 1–27, Oct. 2019.

[90] X.-F. Song, Y. Zhang, D.-W. Gong, and X.-Y. Sun, "Feature selection using bare-bones particle swarm optimization with mutual information," *Pattern Recognit.*, vol. 112, Apr. 2021, Art. no. 107804.

[91] X.-F. Song, Y. Zhang, D.-W. Gong, and X.-Z. Gao, "A fast hybrid feature selection based on correlation-guided clustering and particle swarm optimization for high-dimensional data," *IEEE Trans. Cybern.*, early access, Mar. 17, 2021, doi: 10.1109/TCYB.2021.3061152.

[92] J. Piri and P. Mohapatra, "Imbalanced cardiotocography data classification using re-sampling techniques," in *Proc. Int. Conf. Mach. Intell. Data Sci. Appl.* Singapore: Springer, 2021, pp. 681–692.

[93] H. Fang, Q. Wang, Y.-C. Tu, and M. F. Horstemeyer, "An efficient non-dominated sorting method for evolutionary algorithms," *Evol. Comput.*, vol. 16, no. 3, pp. 355–384, Sep. 2008.

[94] Z. Zhu, Y.-S. Ong, and M. Dash, "Markov blanket-embedded genetic algorithm for gene selection," *Pattern Recognit.*, vol. 40, no. 11, pp. 3236–3248, Nov. 2007.

[95] Z. Yong, G. Dun-Wei, and Z. Wan-Qiu, "Feature selection of unreliable data using an improved multi-objective PSO algorithm," *Neurocomputing*, vol. 171, pp. 1281–1290, Jan. 2016.

[96] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Aug. 2002.

[97] C. Audet, J. Bigeon, D. Cartier, S. L. Digabel, and L. Salomon, "Performance indicators in multiobjective optimization," *Eur. J. Oper. Res.*, vol. 292, no. 2, pp. 397–422, Jul. 2021.

[98] H. Ishibuchi, H. Masuda, Y. Tanigaki, and Y. Nojima, "Difficulties in specifying reference points to calculate the inverted generational distance for many-objective optimization problems," in *Proc. IEEE Symp. Comput. Intell. Multi-Criteria Decis.-Making (MCDM)*, Dec. 2014, pp. 170–177.

[99] M. Li and J. Zheng, "Spread assessment for evolutionary multi-objective optimization," in *Proc. Int. Conf. Evol. Multi-Criterion Optim.* Berlin, Germany: Springer, 2009, pp. 216–230.

JAYASHREE PIRI received the B.Tech. degree in information technology from BPUT, India, in 2007, and the M.E. degree in information technology from Jadavpur University, Kolkata, India, in 2012. She is currently pursuing the Ph.D. degree with the Department of Computer Science, IIIT Bhubaneswar, India. She has more than ten years of teaching experience. She has published more than ten research articles in peer-reviewed journals, such as *Computers in Biology and Medicine* and presented various papers in international conferences. Her main research interests include multi-objective optimization, evolutionary computing, medical data mining, and machine learning.

PUSPANJALI MOHAPATRA (Member, IEEE) received the B.E. degree in electrical engineering from IGIT Sarang, Odisha, India, in 1999, the M.Tech. degree in computer science from Utkal University, Bhubaneswar, in 2002, the M.B.A. degree in systems, IT and HR from BPUT Rourkela, India, in 2009, and the Ph.D. degree in CSE from Siksha 'O' Anusandhan University, India. She is currently an Assistant Professor at the International Institute of Information Technology, Bhubaneswar, and has more than 18 years of teaching experience. She has more than 35 articles in reputed journals, conferences, and book chapters. Her research interests include biomedical data mining, time series data analysis, and different fields like machine learning, deep learning, and evolutionary computing. She is also associated with various educational and research societies, like IEEE, ACM, IAENG, and OITS.

MANAS RANJAN PRADHAN (Member, IEEE) received the Master of Technology degree in computer science from Utkal University, India, and the Ph.D. degree in computer science from the University of Mysore, India. He has vast experience in teaching, research, and academic administration in India and abroad. He is currently working at Skyline University College, Sharjah, United Arab Emirates. As an Academic Leader, he has worked as the Head of Program at UPES, India, and the Dean-Faculty of IT and science at INTI International University, Malaysia. He has been associated with the IT industry for industry-academic collaboration, internships, placements, and workshops. He has executed the IBM Center of Education for Cloud Computing and Business Analytics at INTI International University, under Laureate International Universities, USA. He has presented and published many research papers in various conferences and journals. He has three Indian patents and three Australian patents to his credit. His research interests include business analytics, datamining, data warehouse, retail/ecommerce analytics, artificial intelligence, machine learning, and business process modeling. He has got the Mentor Award for the i-talent project contest from Confederation of Indian Industry (CII). He has played a vital role in organizing three international conferences, such as NGCT-2015 (UPES, India), ICQMOIT-2008 (ICFAI, India), and ICD-2019 (SUC, United Arab Emirates).

BISWARANJAN ACHARYA (Member, IEEE) received the M.C.A. degree from IGNOU, New Delhi, India, in 2009, and the M.Tech. degree in computer science and engineering from the Biju Pattanaik University of Technology (BPUT), Rourkela, Odisha, India, in 2012. He is currently pursuing the Ph.D. degree in computer application with the Veer Surendra Sai University of Technology (VSSUT), Burla, Odisha, India. He is currently an Academic with the Kalinga Institute of Industrial Technology Deemed University. He has a total of ten years of experience in both academia at some reputed universities, like Ravenshaw University and the software development field. He has published some of the patents to his credit. He has published many research articles in international reputed journals as well as serving as a reviewer for many peer-reviewed journals. He has more than 50 patents on his credit. His research interests include multiprocessor scheduling along with different fields, such as data analytics, computer vision, machine learning, and the IoT. He is also associated with various educational and research societies, like IEEE, IACSIT, CSI, IAENG, and ISC.

TAPAS KUMAR PATRA (Member, IEEE) received the Master of Engineering degree from the National Institute of Technology, Rourkela, and the Ph.D. degree from the Indian Institute of Science, Bangalore. He is currently working as an Associate Professor at the College of Engineering and Technology, Bhubaneswar. He has more than 25 years of teaching, academic administration, and research experience. Furthermore, he has guided many award-winning projects at the national and international levels. He has published many papers in reputed journals and conferences. His research interests include electronic systems, communication, wireless networking, the IoT, artificial intelligence, machine learning, computer vision, VLSI, and embedded systems. He is an Active Member of various societies, such as IEEE, COMSOC, ACM, LMISTE, and LMISOI. He is a Microsoft Certified Professional (MCP) and a Microsoft Certified System Engineer (MCSE). He has received the Best Paper Award in IEEE COMSWARE-2007. In addition, he was awarded the Gold Medal in the Intel India Embedded Challenge, in 2011, as a Finalist.