Adaptive Hybridized Meta-Heuristic Algorithm for Subspace Clustering on High Dimensional Data

Pradeep Kumar D [1], Sowmya B J [2], Anita Kanavalli [2], Amaresh T N [1], Anish S [1], Chinmay S Nadgir [1], Gagan A Nischal [1], Supreeth S [3,*], and Shruthi G [3]

[1] Department of Computer Science and Engineering, M S Ramaiah Institute of Technology (Affiliated to VTU), Bengaluru, Karnataka, India; Email: pradeepkumard@msrit.edu (P.K.D.), tn.amaresh2002@gmail.com (A.T.N.), anishgowda43@gmail.com (A.S.), chinmaynadgir@gmail.com (C.S.N.), 1ms19cs044@gmail.com (G.A.N.)
[2] Department of Artificial Intelligence & Data Science, M S Ramaiah Institute of Technology (Affiliated to VTU), Bengaluru, Karnataka, India; Email: researchrit1985@gmail.com (S.B.J), anitak@msrit.edu (A.K.)
[3] School of Computer Science and Engineering, REVA University, Bengaluru, Karnataka, India; Email: g.shruthi466@gmail.com (S.G.)
*Correspondence: supreeth1588@gmail.com (S.S.)

Abstract*—Nature-inspired algorithms have been successful for more efficient clustering of unlabeled data, and have effectively been used to improve a wide variety of numerical optimization problems, and when these algorithms are combined with suitable objective functions, the centroid for clusters is determined iteratively. Centroids are the points that are closest to the center in a cluster. These algorithms are not without their shortcomings, such as slow convergence or fixating on local minima. These are just some of the minor inconveniences that might be caused during our procedures of trying to create a hybrid. A recent trend that has been observed is the hybridization of these algorithms to overcome the shortcomings of the vanilla versions of the algorithm for efficient optimization and clustering. In this work, a novel version of such a Hybrid Meta-heuristic algorithm, developed from the Firefly and Whale optimization algorithms, for faster convergence and better optimization compared to its vanilla counterparts is presented. The firefly and whale algorithms are hybridized such that the drawbacks of one algorithm are taken care of and compensated by the advantages of the other. The outcomes show that the hybrid algorithm of whale and firefly converges faster and is more efficient in comparison with other nature inspired algorithms and its efficiency is further established from the results on standard datasets and as well for finding the clusters with in the different subspaces.

***Keywords*—subspace clustering algorithm, nature inspired algorithm, hybrid meta-heuristic algorithm, efficient clustering, firefly algorithm, whale algorithms**

I. Introduction

The process of successfully creating a hybridized algorithm out of the fundamental nature inspired algorithms using a hybridized model of one or more algorithms.

The literature review indicates that much of these meta-heuristic based algorithms [1] usually get trapped in their local optima's, to overcome these problems, researchers in the field have tried to explore the approach of hybridizing algorithms. The hybrid algorithm usually comes up with a better ability for convergence at global optimum, so we can combine the globalization ability of one algorithm with the better global search ability and a complementing algorithm with a faster convergence mechanism is the best approach to go about with the hybridization of the algorithms.

Ensuring that the hybridized nature inspired algorithm results in a much more efficient solution than the hybrids that have been formed in the past. The drawbacks of a given algorithm are balanced out by embedding the features of other algorithms and thus proving the efficiency of the algorithm by performing empirical analysis on the given subspace cluster.

Ensuring the hybrid is created in such a way that the features of one of the algorithms do not affect or in come in the way of the other algorithm when both are embedded together on top of each other. The objectives are:

1) Creating a hybridized algorithm by combining one or more fundamental nature-based algorithms.
2) Performing Empirical analysis on a given dataset to determine the efficiency of the algorithm.
3) Using subspace clustering to seek clusters within a subspace of a dataset.

The Research article goes with Abstract followed by the Introduction section. The related papers are cited and referred in the Literature Survey. The methodology and the process of the Algorithm are described in the Design and Implementation Section followed by the Result section.

Manuscript received April 17, 2023; revised May 22, 2023; accepted July 31, 2023; published October 20, 2023.

II. Literature Review

A development of feature selection [2], subspace clustering looks for clusters in various subspaces of the same dataset. Subspace clustering requires a search strategy and evaluation standards, just like feature selection. Additionally, subspace clustering has to restrict the assessment criteria's applicability in some way so that distinct subspaces are taken into account for each cluster in this section.

Flexible transformative advancement calculation [3] whales (Adaptive world class technique whale enhancement calculation, AWOA), a strategy for finding food and twisting update area presented versatile changing weighting capacity, abundant variety of populace while fortifying neighborhood streamlining capacity; adaptive differential variety aggravation is presented in the constriction wrapping stage to give the looking through power in the later stage; The adaptive step size factor function and adaptive differential mutation factor are introduced in the AWOA algorithm [4]. The adaptive difference mutation factor solves the method's weakness in the latter stages and keeps the algorithm from slipping towards the local optimal solution. The adaptive step size factor expands the search region and improves population variety. The performance of the algorithm has been improved by a variety of improved strategies. The simulation findings indicate that the late iteration sluggish convergence and weak search issues may be successfully addressed by the AWOA method described in this work. Increase diversity and flexibility of the original algorithm population, convergence speed, optimization capability, and optimization performance.

Agarwal *et al.* gave an in-depth analysis of 12 bio-inspired algorithms in [5]. With regard to input parameters, significant evolutionary methodologies, and application domains, this research provides researchers with a unified platform to analyze and contrast conventional and contemporary nature-inspired algorithms. Twelve nature-inspired algorithms were thoroughly examined as part of their research [6]. The work highlighted the important characteristics of these algorithms in terms of input parameters, evolutionary process, and applications. This paper's main objective was to inform the scientific community about the large-scale global optimization capabilities of contemporary algorithms for multimodal and unimodel continuous functions.

It is being determined how well the Bat Algorithm (BA), Artificial Bee Colony (ABC) algorithm, Cuckoo Search (CS) algorithm, Firefly Algorithm (FA), and Flower Pollination Algorithm (FPA), five contemporary nature-inspired algorithms, perform.

Oduntan and Thulasiraman [7] expressed clustering is a difficult issue with different useful applications that range numerous examination areas. An audit of existing writing shows that there are numerous assorted grouping calculations for various issue spaces The CHCA consolidates the component algorithms collaboratively and IHCA infuses tactics from the component algorithms.

Tzanetos and Dounias [8] expressed for a long time, Machine Learning made it feasible for people to plan the examples that administer introducing issues and furthermore, gave strategies to cluster and characterize huge measure of strange information.

Recent years have seen an increase in the need for novel techniques, notably Evolutionary Strategies, as a result of optimization problems that may be analytically described but are challenging to answer using simple or naive heuristic methods. The strategies used by plants and animals in nature to survive are the inspiration for these techniques.

All Swarm Intelligence techniques are gathered, including those that do not draw their inspiration from swarms, flocks, or groups but rather from animal behaviors. To highlight the key benefits of Nature Inspired Intelligence, the applications of these two subcategories are examined. the preliminary findings are then provided.

Agarwal and Mehta [9] expressed nature-motivated calculations have acquired tremendous prevalence as of late to handle hard genuine world (NP-hard and NP complete) issues and address complex streamlining capacities whose real arrangement doesn't exist. The paper presents a far reaching survey of 12 nature enlivened calculations. Nature motivated registering is a procedure that is enlivened by processes, saw from nature. These registering strategies prompted the improvement of calculations called Nature Inspired Algorithms (NIA). These calculations are subject of computational knowledge. The motivation behind growing such calculations is to streamline designing issues. Nature propelled calculations are persuaded from regular biological system and reproduce the way of behaving of normal living and non-living things. This paper introduced a broad survey of 12 nature-propelled calculations. The work featured the significant highlights of these calculations with regards to their feedback boundaries, transformative instrument, and applications

Dhiman and Kumar [10] presented a novel meta-heuristic computation named Spotted Hyena Optimizer (SHO) energized by the approach to acting of spotted hyenas. The essential thought driving this computation is the social connection between spotted hyenas and their agreeable approach to acting. This Paper presented another large number based smoothing out computation called the Spotted Hyena Optimizer (SHO). The fundamental thoughts which convince the procedure are breathes a new life into by well-disposed arranged movement and the hunting behavior of spotted hyenas. In this paper, the SHO computation is proposed for addressing 29 test abilities to research the examination, misleading, neighborhood optima abhorrence, and intermixing conduct. Furthermore, six certifiable planning plan issues are used to investigate the viability of the SHO computation moreover.

The Whale Streamlining Calculation (WOA) is one of the most popular metaheuristic computations used recently, according to Dutta and Banerjee [11], however any such simple metaheuristic calculation has some drawbacks. In this study, a MWOA (altered WOA) has been suggested as a solution to this drawback. The LSA and BA calculations are combined with the modified WOA. This WOA becomes more precise with LSA, and this framework becomes speedier with BA. By using this MWOA, the problem of premature mixing and catching of neighboring minima is also reduced.

A study on the Bat Algorithm (BA), a Swarm-based metaheuristic strategy, is presented by Zebari *et al.* [12]. BA has been influenced by the scavenging behavior of little bats. Artificial bats acting as search specialists are used by the algorithm to interact with the search for information while simulating the natural pulse loudness and emission rate of actual bats. It has developed into a potent swarm intelligence tactic for handling optimization problems in continuous and discrete spaces. These days, it has been effectively applied to take care of issues in practically all areas of optimization, and it is viewed as exceptionally productive. Accordingly, the writing has extended fundamentally, a wide scope of different applications and contextual analyses have been made based on the bat algorithm.

An attempt to synthesize numerous prominent research suggestions in [13], order for any newcomer to have a better understanding of the trip thus far. Natural evolution, swarm intelligence, biological, science-based, and other nature-inspired algorithms are classified here. This study looked at ACO, ABC, EAM, FA, FPA, GA, GSA, JAYA, PSO, SFLA, TLBO, and WCA, which are all well-known nature-inspired algorithms. The goal of this review is to provide an in-depth look at a variety of nature-inspired algorithms, including their source of inspiration, fundamental operators, control parameters, features, variations, and areas of app where they have been effectively implemented. It will also aid in finding and narrowing down the approaches that are most appropriate for the concern.

Chen *et al.* [14] proposed a flexible transformative advancement calculation whale, a strategy for finding food and twisting update area, presenting versatile changing weighing capacity, abundant differential variety of populace while fortifying neighborhood streamlining capacity. The AWOA method is introduced in this study to overcome the slow convergence speed and poor search power in the latter stages of iteration. The enhanced method keeps the whale algorithm's distinctive search mechanism, and it offers a wide variety of applications and good generalization ability. In Ref. [15–17], the author tried to synthesize numerous prominent research suggestions for any newcomer to have a better understanding of various natural evolution-based, swarm intelligence-based, biological-based, science-based and other nature-based algorithms. In this paper a survey of numerous nature-inspired algorithms has been given. Given that no method can solve all optimization [18, 19] problems perfectly. It could be better at solving some problems than others. Vasant *et al.* [20] studied the Bat Algorithm, a swarm-based metaheuristic algorithm inspired by the scavenging conduct of miniature bats. Bats are attractive creatures and their higher capability of echolocation has fascinated the interest of researchers from different fields. The echolocation component is a sort of sonar: bats, mostly miniature bats, make a noisy and short beat of sound and sort out the distance of an object by utilizing the echo of the sound they emitted returning to their ears. This noteworthy positioning technique gives bats the ability to choose between an obstacle and a prey, permitting them to chase even in entire haziness.

Based on this literature survey, Firefly algorithm and Whale Optimization algorithms have been chosen for hybridization in this paper. The hybridization of these two algorithms makes perfect sense as they have faster convergence and better optimization capability and they also overcome their limitations as they are compensated by each other.

To propose novel closed-form equations [21] for CS and STS prediction of RCCP, the MARS-GOA hybrid intelligent model, based on the combination of the Grasshopper Optimization Algorithm (GOA) and Multivariate Adaptive Regression Splines (MARS), was created. GOA was first created to act as a search engine for the suggested algorithm to optimize MARS control parameters by reducing prediction error. According to statistical metrics, the suggested hybrid MARS-GOA beat ELM, M5p, and standard MARS for the prediction of both the CS and STS of the RCCP (CoD = 0.811, PMARE = 22.146% and U95 = 34.670, respectively). TCS, TTS, and Monte-Carlo Simulation (MCS) reliability investigation results showed that the dependability of the RCCP mix design is mostly determined by the dataset's size and level of ambiguity.

For ten benchmark deterministic optimizations [22] and four RBDO problems of truss structures, respectively, the suggested enhanced BBA's and hybrid WSM-BBA's correctness is tested. The solved examples demonstrate the method's superiority to traditional RBDO approaches for addressing difficult problems involving discrete variables and its computing efficiency.

AOA-NM [23], a special hybrid optimizer, is introduced, and its performance is examined by using it to solve two real-world optimization problems for manufacturing and structures, as well as 10 benchmark engineering design challenges. The CEC2020 numerical test functions are examined statistically, and a non-parametric Wilcoxon sum-rank test is used. Additionally, the statistical findings in each of the cases under study are contrasted with the well-known MHs method. AOA-NM is recommended as finding the best solutions for the majority of the CEC 2020 test functions, and increased performance against comparison methods is shown both quantitatively and qualitatively. The AOA-NM mean statistical findings for all ten engineering design issues revealed an average 22.11% improvement in solution compared to other compared algorithms, with the best improvement reported up to 56.2931% and 127.55% for two of the problems.

In order to quickly approximate the global Best Position (BP) [24], the proposed hybrid Multi-level Cross-entropy-based Moth-Flame (MCMF) method leverages MCEO as a global search engine during the initial stages of optimization. Applying the suggested Search Space Borders Confining Factor (SSBCF), the search space boundaries are then adaptively constrained inside the effective region surrounding the present BP.

III. Design and Implementation

Creating a hybridized algorithm out of the fundamental nature inspired algorithms using a hybridized model of one or more algorithms for subspace cluster analysis. The drawbacks of an existing algorithm are balanced out by embedding the features of other algorithms and thus proving the efficiency of the algorithm by performing empirical analysis on the given subspace clusters. The aim is creating this hybrid algorithm as a flow between the whale optimization tactic and the firefly algorithm and thus resulting with an algorithm that could result in more accurate global minima with each iteration.

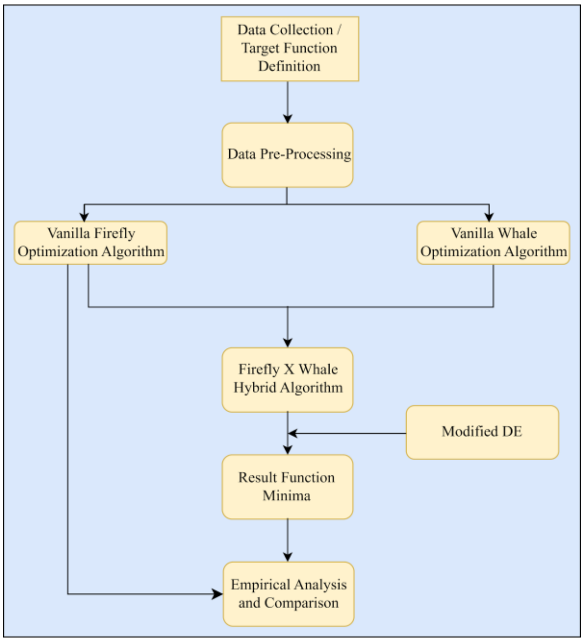


Figure 1. System architecture of hybridization model.

According to Fig. 1, after the initial data collection and target function definition, the pre-processing steps are performed which includes, eliminating or replacing any missing values, outliers, corrupt data entries, and non-informative features. The resulting dataset is normalized, and given as an input to the hybrid algorithm which is developed from the vanilla Firefly Algorithm and vanilla Whale Optimization Algorithm. The operators such as modified DE, sine-Cosine, or golden sine are used to perform the empirical analysis of the algorithm. Finally, the results are collected and graphical analysis is performed. It has various modules such as

1) **Collection and Cleaning of data:** Initial Collection of Data and Objective functions is done, followed by pre-processing the data to eliminate any missing values, outliers and corrupt data entries, and non-informative features.
2) **Algorithm Hybridization:** The resulting Dataset is normalized and given as input to the Vanilla algorithms with an evaluation function for minimizing centroid distance. Alternatively, the target objective function is supplied to the algorithm to be minimized. A hybrid is developed from the vanilla algorithms, and the same dataset/objective functions are supplied to the hybrid.
3) **Empirical Analysis:** Empirical analysis is a method for studying and interpreting data that is based on empirical research. Instead than relying on ideas and conceptions, the empirical method is based on actual data, measurements, and results. The data is gathered, and then an empirical analysis is done.

The whale optimization algorithm was chosen as the algorithm to hybridize and upon some iteration, found that mapping the intensity function of the fireflies with the distance updation metric of the whale algorithm could provide an efficient hybrid. Finally tested the hybrid optimization algorithm against the results of firefly and whale and concluded that the hybrid surpassed both of the vanilla algorithms in terms of efficiency by a significant margin.

A. Algorithmic Description

The equations necessary for the implementation of the hybrid algorithm are mentioned in the Eqs. (1)–(3).

$$I = I_0 e^{-\gamma r} \tag{1}$$

$$\beta = \beta_0 e^{-\gamma r^2} \tag{2}$$

$$\beta = \frac{\beta_0}{1+\gamma r^2} \tag{3}$$

where (I) is the brightness, (γ) is the light absorption coefficient, and (β) is the attractiveness of the fireflies.

$$\zeta_i^{t+1} = \zeta_i^t + \beta_0 e^{-\gamma r_{ij}^2} \left(\zeta_j^t - \zeta_i^t\right) + \alpha \varepsilon_i^t \tag{4}$$

Eq. (4) is used to calculate the instinctive behavior of the fireflies, where α is a random multiplier and ζ_i^t is a gaussian distribution vector.

$$\begin{gathered} \boldsymbol{X}(t+1) = \boldsymbol{D}' e^{bl} \cos(2\pi l) + \boldsymbol{X}^*(t) \\ \boldsymbol{D}' = \left|\boldsymbol{X}^*(t) - \boldsymbol{X}(t)\right| \end{gathered} \tag{5}$$

Eq. (5) is used to describe the spiral path generated by the whale algorithm during the exploitation phase, where X determines the position vector for each whale, b is a constant and l is a random number.

$$\begin{gathered} \eta = \frac{2}{1+e^{\beta sign\left(\Sigma\left(X^* - X(t)\right)\right)\|D\|}} \\ \eta_1' = \frac{2}{1+e^{r_3 \beta D'}} \\ \eta_2' = \frac{2}{1+e^{r_4 \beta D''}} \end{gathered} \tag{6}$$

where η, η_1', η_2' are the attractiveness coefficients generated in the Eq. (6) specifically for the new hybrid algorithm. Where X is the positional vector of the hybrid agent, $|A|$ is the absolute value of the random vector A.

Algorithm 1. Hybrid Algorithm

Initialize the whale population at random.
Determine the most effective agent X* by
evaluating the fitness values of whales.
while $t < t_{max}$ do
Determine a's value.
Perform for each search agent
if h<0.5 then
if |A| < 1 then
$X(t+1) = X^*(t) - A.D$
else
$X(t+1) = X_{rand?}(t) - A.D\eta$
else
if |A| < 1 then
$X(t+1) = D\eta_1 e^{bl} \cos(2\eta_1) + X^*(t)$
else
$X(t+1) = D\eta_2 e^{bl} \sin(2\eta_1) + X^*(t)$
end for
end while
Review X(t+1)'s fitness and update X*.(t)
Post-process results and visualization

The hybrid Algorithm 1 is an encapsulation of the features which are better in each of them, the hybrid aims to comprise the hunting property of the whale where usually the hunter leader encircles and some of the hunter whales try to poke a fish out at random. The attractiveness parameter of the firefly algorithm helps the fireflies converge at a much faster rate and we have made use of that in our hybrid. The positions of the centroids are assigned randomly at first and then from there when they are about to update the position of the centroids it uses a random probability function to decide the course of action while the firefly's attractiveness parameter is used as a "weight factor" essentially signifying the degree to which the move at a particular direction should be made depending on the other nodes. Thus, we end up with an algorithm which converges quickly but doesn't get stuck at local minima.

IV. Results and Inferences

The hybrid algorithm is evaluated on the 14 CEC Benchmark Functions which include the names of

- Beale function
- Six-hump camel-back function
- Easom function
- Eggholder function
- Holder table function
- Levy n.13 function
- Mccormick function
- Michalewicz function
- Rosenbrock function
- Schwefel function
- Styblinski-tang function
- Cross-in-tray function
- Bukin n.6 function
- Schubert function

Functions though can synthesize large amounts of data for evaluation of clusters. To avoid the over optimized results and test our hybrids performance on real time observations we also run it on some real time datasets as shown in Tables I–IV, which include the names of,

- Aggregation dataset
- Aniso dataset
- Appendicitis dataset
- Blood dataset
- Banknote dataset
- Diagnosis dataset
- Ecoli dataset
- Flame dataset
- Glass dataset
- Heart dataset
- Iris dataset
- Liver dataset
- Moons dataset
- Partition based dataset
- Path based dataset
- Seeds dataset
- Smiley dataset
- Sonar dataset
- Varied dataset
- Varied-density dataset
- Vertebral dataset
- Vertebral 2 dataset
- Vertebral 3 dataset
- Wine dataset
- Jain dataset
- Circles dataset
- Iris dataset
- Blobs dataset
- Balance dataset

A. CEC Objective Functions

For optimization, chose 14 of the CEC 2015 benchmark functions that were most suitable for testing. The functions are now described visually, together with the minimal value and the moment at which it is reached. 14 functions were executed. The sample of one functions is as shown in the Table I.

B. 30 Real Life/ Artificial Datasets

Used 30 datasets comprising both real-life datasets and artificial datasets. These are used to determine the performance of the hybrid and also to show the practical applications for the hybrid, i.e., clustering. Sample of the result is as shown in Table II.

TABLE I. CEC OBJECTIVE FUNCTIONS

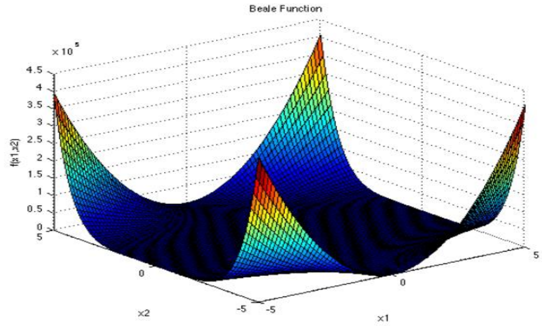
Beale function	
Function	$f(\mathbf{x}) = \left(1.5 - x_1 + x_1 x_2\right)^2 + \left(2.25 - x_1 + x_1 x_2^2\right)^2 + \left(2.625 - x_1 + x_1 x_2^3\right)^2$
Minimum	$f(x^*) = 0$, $x^* = 3, 0.5$
Visualization	

TABLE II. AGGREGATION DATASET

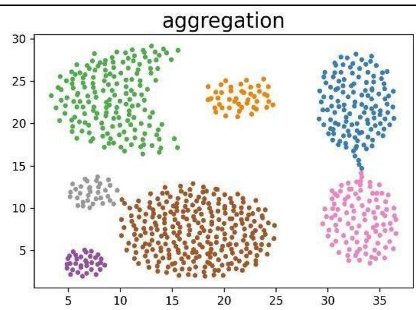
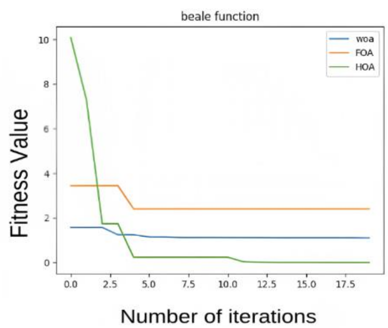
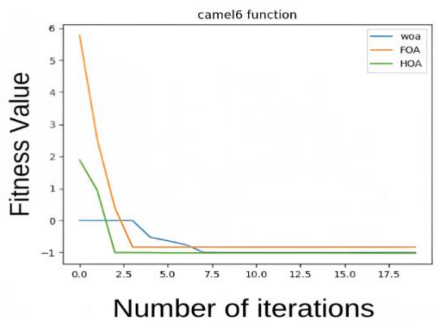
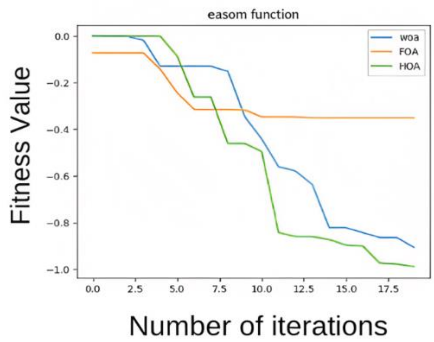
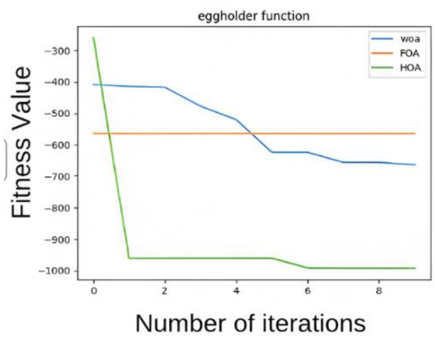
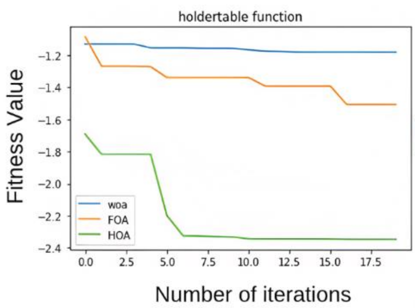
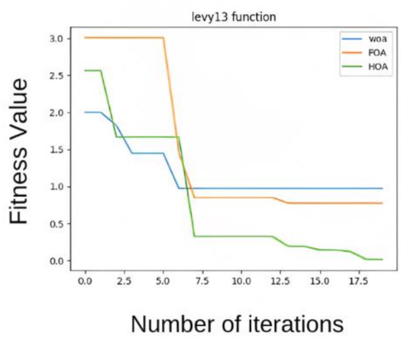
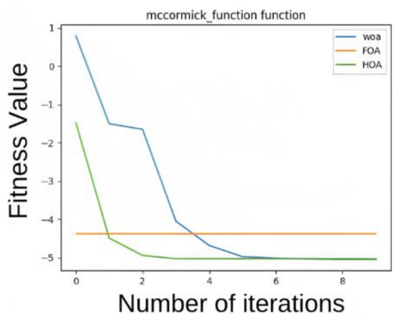
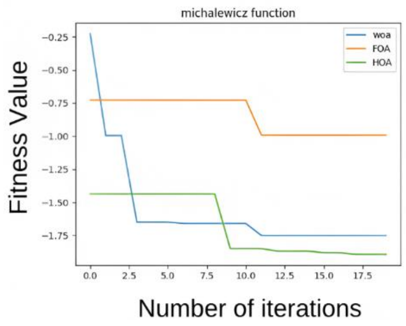
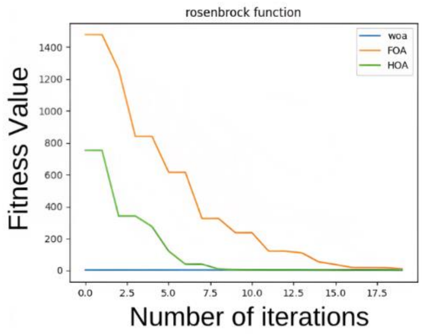
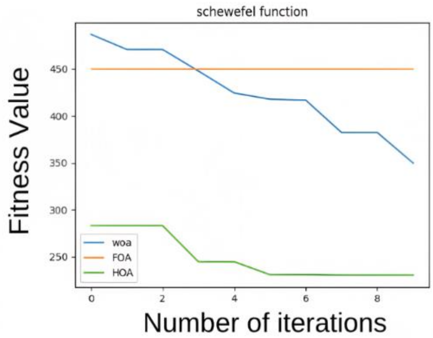
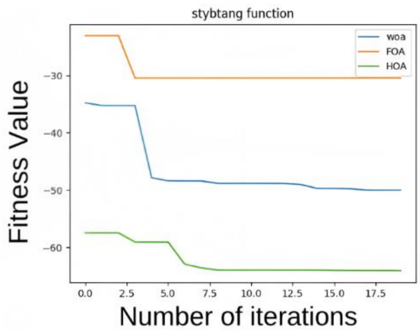
Aggregation Dataset				
No. of Clusters	7			
Dataset Contents		**Column1**	**Column2**	**Label**
	0	15.55	28.65	2
	1	14.9	27.55	2
	2	14.45	28.35	2
	3	14.15	28.8	2
	4	13.75	28.05	2
Dataset Description	Count	788.000000	788.000000	788.000000
	Mean	19.566815	14.171764	2.737310
	Std	9.922042	8.089683	1.664573
	Min	3.350000	1.950000	0.000000
	25%	11.150000	7.037500	2.000000
	50%	18.225000	11.725000	3.000000
	75%	30.700000	21.962500	4.000000
	Max	36.550000	29.150000	6.000000
Visual Representation of Clusters				

TABLE III. EMPIRICAL ANALYSIS ON BENCHMARK FUNCTIONS

Name of the function	Empirical result of the function	Inferences
Beale function		Hybrid converges to the minima faster than both whale and firefly No localization by any algorithm can be seen

Six-hump camel-back function		Hybrid converges to the minima faster than both whale and firefly No localization by any algorithm can be seen
Easom function		Hybrid converges to the minima faster than both whale and firefly The firefly algorithm gets stuck at a local minima. X-variable-number of iterations Y-variable-Fitness value
Eggholder function		Hybrid converges to the minima faster than both whale and firefly The firefly algorithm gets stuck at a local minima.
Holder Table Function		Hybrid converges to the minima faster than both whale and firefly No localization by any algorithm can be seen
Levy N.13 Function		Hybrid converges to the minima faster than both whale and firefly The Firefly and whale algorithm gets stuck at a local minima.

<table>
<tr>
<td>Mccormick function</td>
<td>
</td>
<td>Hybrid converges to the minima faster than both whale and firefly
The Firefly algorithm gets stuck at a local minima.</td>
</tr>
<tr>
<td>Michalewicz Function</td>
<td>
</td>
<td>● Hybrid converges to the minima faster than both whale and firefly
● The Firefly algorithm gets stuck at a local minima.</td>
</tr>
<tr>
<td>Rosenbrock Function</td>
<td>
</td>
<td>Hybrid converges to the minima faster than both whale and firefly
No localization by any algorithm can be seen</td>
</tr>
<tr>
<td>Schwefel Function</td>
<td>
</td>
<td>● Hybrid converges to the minima faster than both whale and firefly
● The Firefly algorithm gets stuck at a local minima.</td>
</tr>
<tr>
<td>Styblinski-Tang Function</td>
<td>
</td>
<td>Hybrid converges to the minima faster than both whale and firefly
The Firefly algorithm gets stuck at a local minima.</td>
</tr>
</table>

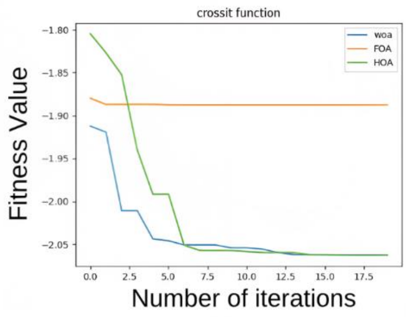
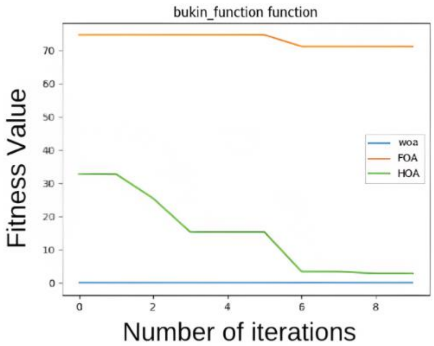
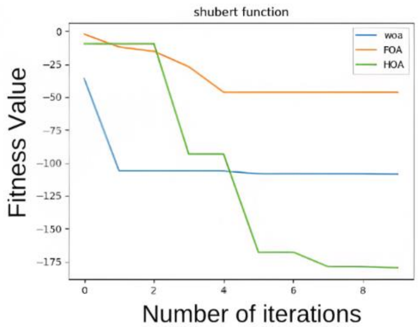
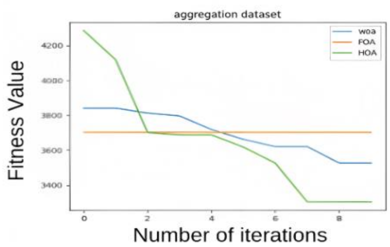
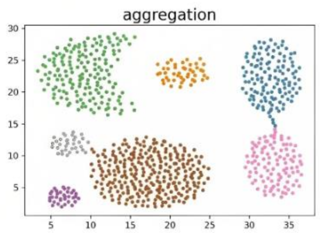
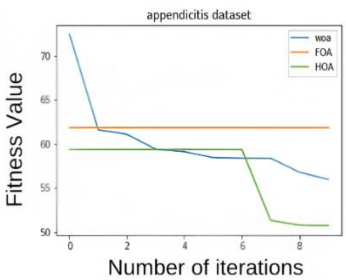
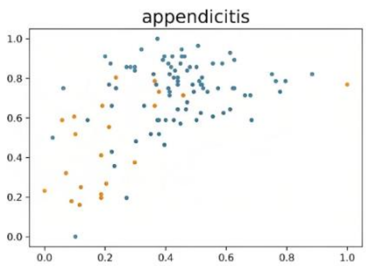
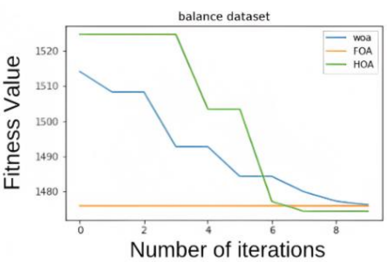
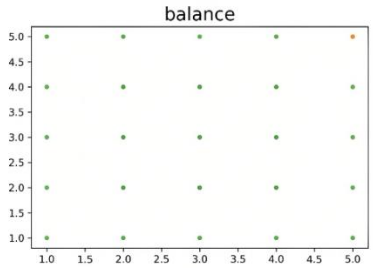
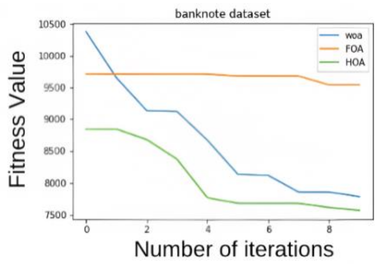
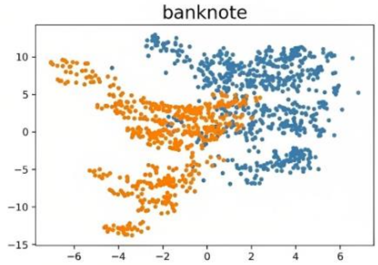
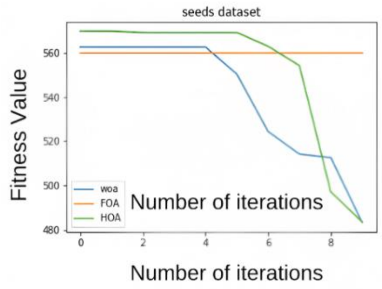
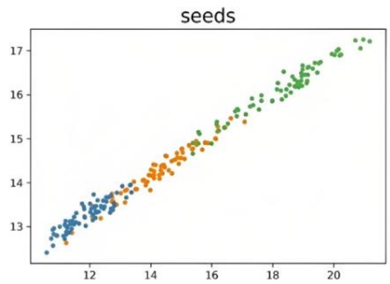
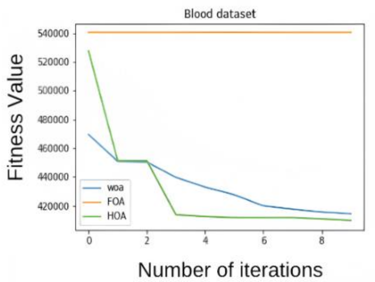
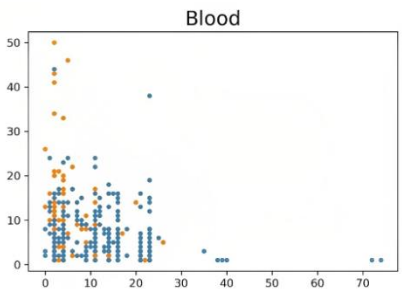
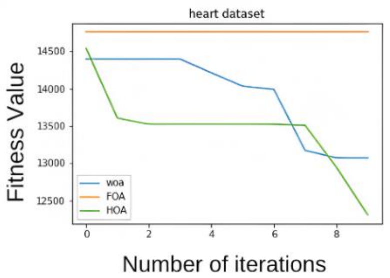
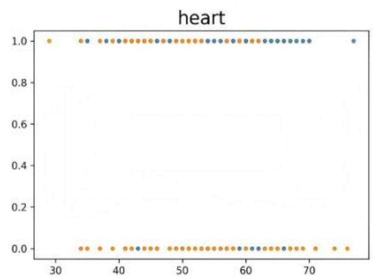
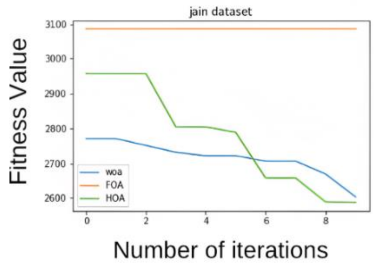
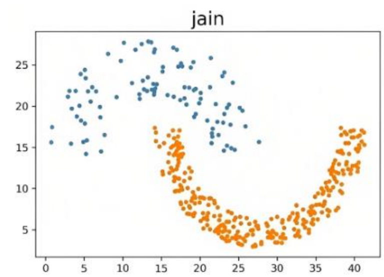
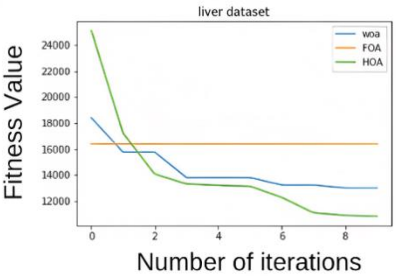
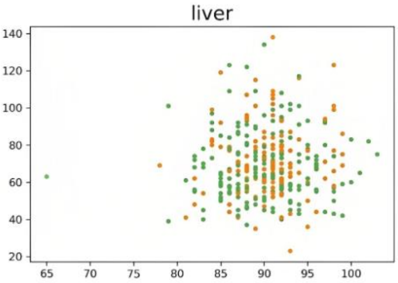
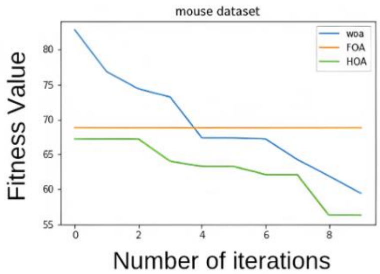
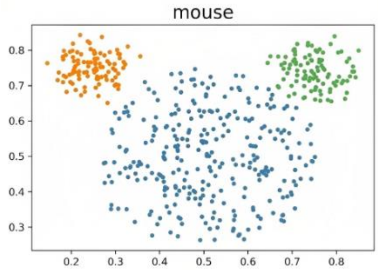
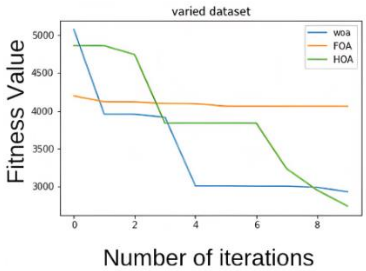
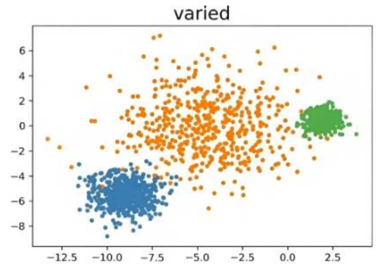
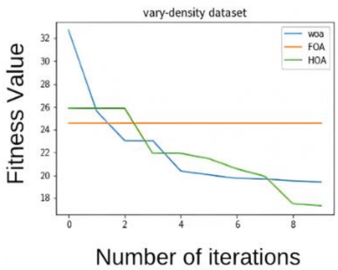
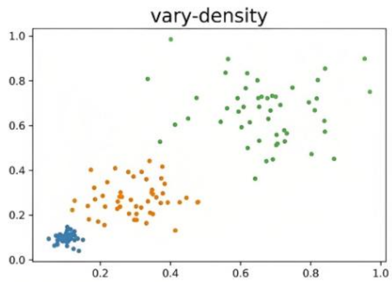
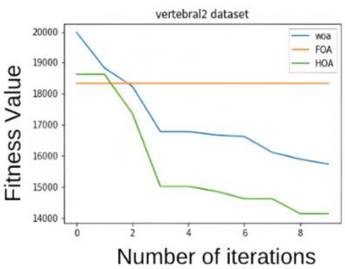
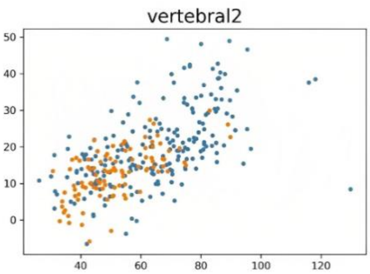
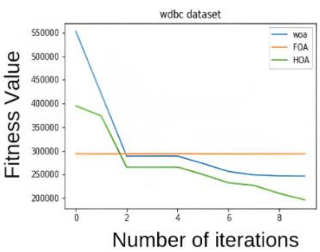
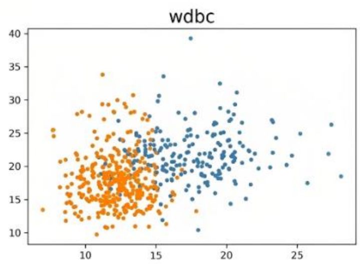
Cross-In-Tray Function		Hybrid converges to the minima faster than both whale and firefly No localization by any algorithm can be seen
Bukin N.6 Function		Hybrid converges to the minima faster than both whale and firefly The Firefly algorithm gets stuck at a local minima.
Schubert Function		Hybrid converges to the minima faster than both whale and firefly No localization by any algorithm can be seen

TABLE IV. EMPIRICAL ANALYSIS ON REAL TIME DATASETS

Name of The Dataset	Dataset	Cluster Diagram	Inferences
Aggregation Dataset			• The Hybrid has a lower fitness value than both the vanilla algorithms. • Firefly localizes because the number of iterations is very less.
Appendicitis Dataset			• The Hybrid has a lower fitness value than both the vanilla algorithms. • Firefly localizes at a local minima because the number of iterations is very less

Balance Dataset			• The Hybrid has a lower fitness value than both the vanilla algorithms.
Banknote Datasets			• The Hybrid has a lower fitness value than both the vanilla algorithms. • Firefly localizes because the number of iterations is very less
Seeds Dataset			• The whale has a lower fitness value than both the hybrid algorithms.
Blood Dataset			• The Hybrid has a comparable fitness value with the whale optimization algorithms.
Heart Dataset			• The Hybrid has a lower fitness value than both the vanilla algorithms.
Jain Dataset			• The Hybrid has a comparable fitness value than the whale optimization algorithms.

<table>
<tr><td>Liver Dataset</td><td></td><td></td><td>• The Hybrid has a lower fitness value than both the vanilla algorithms.</td></tr>
<tr><td>Mouse Dataset</td><td></td><td></td><td>• The Hybrid has a lower fitness value than both the vanilla algorithms.</td></tr>
<tr><td>Varied Dataset</td><td></td><td></td><td>• The Hybrid has a comparable fitness value than the whale optimization algorithms.</td></tr>
<tr><td>Varied Density Dataset</td><td></td><td></td><td>• The Hybrid has a lower fitness value than both the vanilla algorithms.</td></tr>
<tr><td>Vertebral 2 Dataset</td><td></td><td></td><td>• The Hybrid has a lower fitness value than both the vanilla algorithms.</td></tr>
<tr><td>Wdbc Dataset</td><td></td><td></td><td>• The Hybrid has a lower fitness value than both the vanilla algorithms.</td></tr>
</table>

C. Subspace Clustering

A development of conventional clustering, subspace clustering looks for clusters in several subspaces within a dataset. Many dimensions in high dimensional data are frequently of no meaning and can conceal pre-existing clusters in noisy data. Through an analysis of the full dataset, feature selection eliminates duplicate and unnecessary dimensions. Algorithms for subspace clustering localize the search for pertinent dimensions, enabling them to locate clusters that are distributed across numerous, potentially overlapping subspaces.

The empirical analysis we use subspace clustering to demonstrate the efficiency of our hybrid using a function and a following dataset as shown in Tables V and VI.

- The Holder Table Function.
- The UCI-Drivface dataset.

The analysis of hybrid approach, performs better with respect to its vanilla counterparts.

TABLE V. SUBSPACE CLUSTERING ON HOLDER TABLE FUNCTION

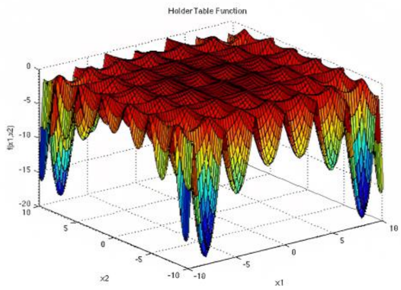
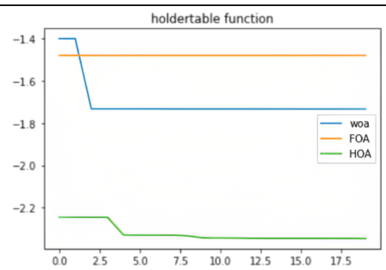
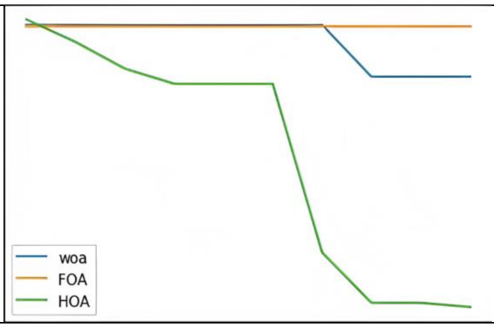
Holder Table Function-Subspace Clustering	
Function	$f(\mathbf{x}) = -\lvert\sin(x_1)\cos(x_2)\exp\left(\left\lvert 1-\frac{\sqrt{x_1^2+x_2^2}}{\pi}\right\rvert\right)$
Minimum	$f(x^*) = -19.2085$, $x^* = (8.05502, 9.66459)$
Visualization	
Empirical Result	
Inferences	• Hybrid converges to the minima faster than both whale and firefly • Some localization is seen by all algorithms.

TABLE VI. SUBSPACE CLUSTERING ON UCI-DRIVFACE DATASET

Drivface Dataset-Subspace Clustering	
Name	UCI-Drivface Dataset
Dimensions	6400 dimensions
Visualization	
Empirical Result	
Inferences	• Hybrid converges to the minima faster than both whale and firefly. • Some localization is seen throughout in the Firefly algorithm and initial localization in Whale algorithm but the hybrid has a more consistent convergence.

V. Conclusions

From the empirical analysis done on all the 14 CEC Benchmark functions and many instances of real life datasets and high dimensional datasets, the proposed hybrid does indeed comprise both the fast convergence of whale algorithm and at the same time non localization properties of the firefly algorithm is been witnessed. The implementation and architectural details can be inferred from the System design. The initial positions of the centroids are chosen at random, and when it comes time to update the positions of the centroids, a random probability function is used to determine the best course of action. The attractiveness parameter of the firefly is used as a "weight factor," essentially denoting the degree to which a move should be made in a specific direction depending on the other nodes. As a result, method was obtained that swiftly converges without hitting the local minima. From the charts it is observed that most of the times the hybrid manages to converge to the proposed minima within a small number of iterations, but also observed that more often than not in the speculated number of iterations the firefly algorithm doesn't even start converging as its computational requirement is of such a high margin. The algorithm which is proposed here has several noteworthy use cases in the various research domains where heuristic algorithms were used to explore in our future work. This Hybridized model can be applied for Design and Manufacturing problem.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Pradeep Kumar D, Sowmya B J were responsible for Initial problem identification, algorithm write-up, analysis, drafting of the manuscript, and simulation. Anita Kanavalli, Amaresh T N were responsible for the Literature survey and helped in the initial review process. Chinmay S Nadgir, Gagan A and Nischal were responsible for the Complexity analysis of the research, evaluation of the research work. Dr. Supreeth S and Prof. Shruthi G were responsible for the final formatting and applied for the journal. All authors worked together to implement and evaluate the integrated system, and approved the final version of the paper.

Acknowledgment

Authors acknowledge the support from MSRIT, REVA University for the facilities provided to carry out the research.

References

[1] E. Ezugwu *et al.*, "Metaheuristics: A comprehensive overview and classification along with bibliometric analysis," *Artificial Intelligence Review*, vol. 54, no. 6, pp. 4237–4316, Mar. 16, 2021. doi: 10.1007/s10462-020-09952-0

[2] A. José-García and W. Gómez-Flores, "Automatic clustering using nature-inspired metaheuristics: A survey," *Applied Soft Computing*, vol. 41, pp. 192–213, Apr. 2016. doi: 10.1016/j.asoc.2015.12.001

[3] C. J. A. Mitchell and M. Shannon, "Exploring cultural heritage tourism in rural newfoundland through the lens of the evolutionary economic geographer," *Journal of Rural Studies*, vol. 59, pp. 21–34, Apr. 2018. doi: 10.1016/j.jrurstud.2017.12.020

[4] Z. Wang, S. Gao, J. Wang, H. Yang, and Y. Todo, "A dendritic neuron model with adaptive synapses trained by differential evolution algorithm," *Computational Intelligence and Neuroscience*, vol. 2020, pp. 1–19, Jan. 17, 2020. doi: 10.1155/2020/2710561

[5] P. Agarwal, S. Mehta, and A. Abraham, "A meta-heuristic density-based subspace clustering algorithm for high-dimensional data," *Soft Computing*, vol. 25, no. 15, pp. 10237–10256, Jun. 21, 2021. doi: 0.1007/s00500-021-05973-1

[6] R. Soto, J. A. Gómez-Pulido, E. Rodriguez-Tello, and P. Isasi, "Advances in recent nature-inspired algorithms for neural engineering," *Computational Intelligence and Neuroscience*, vol. 2020, pp. 1–2, Oct. 14, 2020. doi: 10.1155/2020/7836239

[7] O. I. Oduntan and P. Thulasiraman, "Hybrid metaheuristic algorithm for clustering," in *Proc. 2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, Nov. 2018. doi: 10.1109/ssci.2018.8628863

[8] A. Tzanetos and G. Dounias, "A comprehensive survey on the applications of swarm intelligence and bio-inspired evolutionary strategies," *Learning and Analytics in Intelligent Systems*, pp. 337–378, 2020. doi: 10.1007/978-3-030-49724-8_15

[9] P. Agarwal and S. Mehta, "Nature-inspired algorithms: State-of-art, problems and prospects," *International Journal of Computer Applications*, vol. 100, no. 14, pp. 14–21, Aug. 20, 2014. doi: 10.5120/17593-8331

[10] G. Dhiman and V. Kumar, "Spotted hyena optimizer: A novel bio-inspired based metaheuristic technique for engineering applications," *Advances in Engineering Software*, vol. 114, pp. 48–70, Dec. 2017. doi: 10.1016/j.advengsoft.2017.05.014

[11] S. Dutta and A. Banerjee, "Optimal image fusion algorithm using modified whale optimization algorithm amalgamed with local search and BAT algorithm," in *Proc. 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, IEEE, Mar. 2020. doi: 10.1109/iccmc48092.2020.iccmc-000132

[12] A. Y. Zebari, S. M. Almufti, and C. M. Abdulrahman, "Bat Algorithm (BA): Review, applications and modifications," *International Journal of Scientific World*, vol. 8, no. 1, p. 1, Jan. 23, 2020. doi: 10.14419/ijsw.v8i1.30120

[13] P. Agarwal and S. Mehta, "Empirical analysis of five nature-inspired algorithms on real parameter optimization problems," *Artificial Intelligence Review*, vol. 50, no. 3, pp. 383–439, Mar. 04, 2017. doi: 10.1007/s10462-017-9547-5

[14] J. Chen, H. Rong, Z. Zhang, and R. Luo, "An adaptive evolutionary whale optimization algorithm," in *Proc. 2021 33rd Chinese Control and Decision Conference (CCDC)*, IEEE, May 22, 2021. doi: 10.1109/ccdc52312.2021.9601898

[15] S. R. Kumar and K. D. Singh, "Nature-inspired optimization algorithms: Research direction and survey," arXiv pre-print, 2021. doi: 10.48550/ARXIV.2102.04013

[16] S. Supreeth, K. Patil, S. D. Patil, and S. Rohith, "Comparative approach for VM scheduling using modified particle swarm optimization and genetic algorithm in cloud computing," in *Proc. 2022 IEEE International Conference on Data Science and Information System (ICDSIS)*, 2022. doi: 10.1109/ICDSIS55133.2022.9915907

[17] S. Supreeth and K. Patil, "Hybrid Genetic Algorithm and Modified-Particle Swarm Optimization Algorithm (GA-MPSO) for predicting scheduling virtual machines in educational cloud platforms," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 17, no. 07, pp. 208–225, Apr. 12, 2022. doi: 10.3991/ijet.v17i07.29223

[18] G. Shruthi, M. R. Mundada, B. J. Sowmya, and S. Supreeth, "Mayfly Taylor optimisation-based scheduling algorithm with deep reinforcement learning for dynamic scheduling in fog-cloud computing," *Applied Computational Intelligence and Soft Computing*, vol. 2022, pp. 1–17, Aug. 28, 2022. doi: 10.1155/2022/2131699

[19] S. Supreeth and K. Patil, "VM Scheduling for Efficient Dynamically Migrated Virtual Machines (VMS-EDMVM) in

cloud computing environment," *KSII Transactions on Internet and Information Systems*, vol. 16, no. 6, pp. 1892–1912, 2022. doi: 10.3837/tiis.2022.06.007

[20] P. Vasant, A. Banik, J. J. Thomas, J. A. Marmolejo-Saucedo, U. Fiore, and G.-W. Weber, "Bio-inspired approaches for a combined economic emission dispatch problem," *Human-Assisted Intelligent Computing*, pp. 3-1–3-38, Apr. 01, 2023. doi: 10.1088/978-0-7503-4801-0ch3

[21] G. Zhang, N. S. Hamzehkolaei, H. Rashnoozadeh, S. S. Band, and A. Mosavi, "Reliability assessment of compressive and splitting tensile strength prediction of roller compacted concrete pavement: introducing MARS-GOA-MCS," *International Journal of Pavement Engineering*, vol. 23, no. 14, pp. 5030–5047, Nov. 22, 2021. doi: 10.1080/10298436.2021.1990920

[22] N. S. Hamzehkolaei, M. Miri, and M. Rashki, "An improved binary bat flexible sampling algorithm for reliability-based design optimization of truss structures with discrete-continuous variables," *Engineering Computations*, vol. 35, no. 2, pp. 641–671, Apr. 16, 2018. doi: 10.1108/ec-06-2016-0207

[23] S. Yıldız *et al*., "A novel hybrid arithmetic optimization algorithm for solving constrained optimization problems," *Knowledge-Based Systems*, vol. 271, 110554, Jul. 2023. doi: 10.1016/j.knosys.2023.110554

[24] N. S. Hamzehkolaei and F. MiarNaeimi, "A new hybrid multi-level cross-entropy-based moth-flame optimization algorithm," *Soft Computing*, vol. 25, no. 22, pp. 14245–14279, Sep. 29, 2021. doi: 10.1007/s00500-021-06109-1

Copyright © 2023 by the authors. This is an open access article distributed under the Creative Commons Attribution License (CC BY-NC-ND 4.0), which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.