Nat. Volatiles & Essent. Oils, 2021; 8(4): 13135-13144

# Improved Whale Optimization Algorithm For Clustering

**Pradeep Kumar D[1] , Sowmya B J[2] , Anita Kanavalli[3] , Neville Joseph Roy[4] , Karthik S.L[5] , Poorvi Goyal[6]**

Department of Computer Science and Engineering, Ramaiah Institute of Technology, MSRIT Post, M S Ramaiah Nagar, MSR Nagar, Bengaluru, Karnataka, India 560054

**Abstract:**

Clustering,inthefieldofdatamining,isdefinedastheprocessofgroupingsimilardatapoints.Nature-inspired algorithmsareusedinclusteringtoavoidprematureconvergenceintolocaloptima.Nature-inspiredalgorithms such as cuckoo search, firefly algorithm, bat algorithm, and flower pollination algorithm are defined as algorithms that emulate animals' behavior in nature under varied circumstances. One such algorithm is the Whale Optimization Algorithm (WOA), inspired by the humpback whales' bubble-net hunting strategy. Although WOA is observed to outperform several other nature-inspired algorithms, it suffers from exploration-exploitation imbalance and trapping in local optima. This paper proposes an improved Whale Optimization Algorithm with optimized hyperparameters determined using the Grid Search Algorithm to overcome the aforementioned. The proposed work is seen to outperform the existingWOA.

**Keywords:** Nature-based Algorithms, Clustering, Whale Optimization Algorithm, Grid Search,Optimization, Exploration,Exploitation.

## 1. INTRODUCTION

Nature-basedalgorithmsaredefinedasalgorithmsusedtosolveengineeringproblemsoroptimizethemby emulating events in nature. There are three types of nature-based algorithms [1] viz. Evolution-based such as Genetic Algorithms (GA) that emulate natural evolution laws, physics-based such as Black Hole (BH) algorithms that emulate the universe's physical rules, and swarm-based such as Particle Swarm Optimization (PSO) that emulate animal groups behavior innature

Thesearchingphaseconsistsoftwopartsinpopulation-basedoptimizationalgorithms,namelyexploration and exploitation [2]. Globally exploring operators are utilized by the optimizer to explore the given search space. The exploitation phase follows this exploration phase, and this involves further searching for solutions in promising areas. Exploitation refers to the local search capabilities in the promising areas of the search space found in the exploitation phase. Whale Optimization Algorithm (WOA) is a popular algorithm that emulates humpback whales' hunting methodology [3]. Whales are large mammals who live by themselves or in groups and prey on krill. Their preying methodology includes bubble-net feeding methods, which include upward spirals and double-loops. It has unique exploration and exploitation capabilities. However, finding an appropriate balance between the exploration and exploitation phases remains challenging due to the stochastic nature of the optimizationprocess.

This work proposes an improved version of the Whale Optimization Algorithm that uses the current state of the solution to determine the course of action taken by the search agent (whale), thereby providing an improvedexploration-exploitationbalanceoftheoriginalwhalealgorithm.Theefficiencyofthisalgorithm is tested using well-known benchmark functions to demonstrate its comparison with the original Whale Optimization Algorithm.

The rest of the paper is divided into the following sections: Section 2 details the related works, including variousnature-inspiredalgorithmsandtheircontributiontodataclustering.Section3focusesonthedesign and implementation of the proposed model of WOA. The observed experimental results are discussed in Section 4.

## 1. RELATED WORKS

Data mining refers to the process of recognizing patterns in databases using required methods. Clustering is defined as the process of grouping together similar kinds of data. It is divided into hierarchical and partitional clustering. It includes summation, association, and clustering. Among these, clustering is the most practical methodology, with a wide range of applications. Clustering aims to achieve optimum conditions by increasing similarity among data points in the same cluster and decreasing similarity among data points of different clusters.

Nature-inspiredalgorithmsintoday'sworldareincreasinglyusedtosolvedifficultcomputerproblemsand optimize the existing problems. Since optimization is becoming a difficult task in today's day and age due to the increasing complexity, these algorithms are used to solve difficult problems of multiple variables, dimensions, etc. Nature-based algorithms are divided into evolutionary algorithms (based on the survival of the fittest) and swarm intelligence-based algorithms (mimics the behavior of naturally occurring swarms). These algorithms [5] can be optimized to solve difficult engineering problems easily. The terminologyofthesealgorithmsisbasedontheirdomainofmetaheuristicandoptimizationratherthantheir domain of inspiration [6]. The redundancy of brute force approaches paves the way for more efficient algorithms. Although nature-inspired algorithms help improve efficiency in certain aspects, they are plagued with several challenges and difficulties. Few of these problems include analysis of algorithmic convergence and stability, the mathematical framework, parameter tuning, and the role of benchmarking and scalability[7].

PSO[8]isametaheuristicalgorithmthatimprovesagivenproblembyiterativelyoptimizingthecandidate solution with reference to a particular measure of quality. PSO is a metaheuristic in that it makes minimal assumptions about the problem being solved and can seek a large range of various solutions. A problem is solved using a population of candidate solutions, known as particles, and repositioning these particles in thesearchspaceusingwell-knownmathematicalformulas.Eachparticleinsearchspaceadjustsitsposition and velocity according to its own experience in addition to the experience of other particles The particle's actionsdependonthelocalbestpositionthatisupdatedasbetterpositionsarediscoveredbyotherparticles. This attracts the swarm to the bestsolutions.

CuckooSearch(CS)algorithm[9]is a popula rnature based algorithm based on the mannerisms of cuckoo birds laying eggs. The algorithm emulates the reproductive pattern of cuckoos, where a cuckoo selects a nest to lay an egg. Good nests with high-quality eggs are carried over to further generations. The eggs can be discovered with a particular probability and either the eggs are thrown or the nest is abandoned. The Cuckoo Search algorithm has three variations- one with linearly increasing switching parameter, one with exponentially increasing switching parameter, and one with an increasing power switching parameter. Itis observed that Cuckoo Search using exponentially increasing switching parameters [10] is more efficient than itscounterparts.

TheGreyWolfoptimizer(GWO)[11]isameta-heuristicalgorithminspiredbythebehaviorofgreywolves in nature. This emulates the hunting technique and leadership structure of grey wolves. This structure employsfourtypesofwolvesnamelyalpha,beta,delta,andomega.Thehuntingtechniqueconsistsofthree steps such as searching, encircling, and attacking prey. On comparing this optimizer with well-known algorithms, it is observed that the Grey Wolf Optimiser algorithm provides competitive results. This algorithm has been utilized to solve certain classical engineering design problems such as tension/compression spring, pressure vessel design, and welded beam providing promising results for the same.

ArtificialBeeColony[12]isanoptimizationmethodthatmimicsthebehaviorofhoneybees.Inthismodel, the colony is divided into three sets of bees: employed bees, onlooker bees, and scout bees. It is assumed thatthenumberoffoodsourcesisthesameasthatofthenumberofemployedbeesintheparticularcolony. The employed bees travel to the various food sources, return to the hive and perform a particular dance. The employed bee whose food supply is no longer available becomes a scout and is on the lookout for a newfoodsource.Thefoodsourceischosenbytheonlookerbeesbasedontheperformanceoftheemployed bees.Thefoodsourcewiththelargestamountofnectarfoundthusfarisregistered.Thisprocessiscarried outuntilvariousrequirementsaresatisfied.Inthisalgorithm,thequalityofthefoodsourcedeterminesthe quality or fitness of the correspondingsolution.

## 2. PROPOSEDMETHODOLOGY

TheoriginalWhaleOptimizationAlgorithm[13]actstosearch,encircleorattackbasedonrandomchoices. Continuousiterationsofthethreeactionsareperformeduntilthemostoptimalfitnessvalueisattained.The attack coefficient is arbitrarily set to 1.0. However, it is observed through experimental analysis that the chosen attack coefficient value is notoptimal.

**Choosing the optimal value of the attack coefficient**

Using the Grid Search Algorithm [14] and analyzing the results, an appropriate attack coefficient value is set.Thevaluesthataretestedtoattaintheoptimalvaluerangefrom0.1to1.2inincrementsof0.05against fourobjectivefunctionswithvaryingdimensionsviz.theBooth[15],Levi[16],Eggholder[17],andMatyas [18] objective functions. It is observed that, for each of the four testing functions, the updated attack coefficient value of 0.3 returns the highest fitness value.

The proposed model utilizes the current state of the solution to determine the search agent’s course of action: search, attack, or encircle. Using the methodology of the original WOA, the coefficients have been optimized to obtain a more optimum result. These modifications improve the overall runtime efficiency of the algorithm along with an improvement in clustering accuracy and efficiency. With the increase in the number of inputs, the produces more anticipated and consistent results across various inputs and test functions.

**Design of the proposed model**

Thewhaleoptimizationalgorithmconsistsof3integralcomponents-search,encircleandattack.Humpback whales initially search for their prey, encircle it while releasing bubbles and attack the prey by clustering them. In the original algorithm, the above operations were performed randomly and iteratively, where the solution with best fitness is found. The proposed model is designed to optimise the existing algorithm by determiningtheappropriateactiontobetakenbythesearchagent.Byalteringcoefficientsintheproposed algorithm,animprovementintheoptimumsolutionisobtained,resultinginincreasedconsistency,reduced runtime, improvements in clustering accuracy andefficiency.

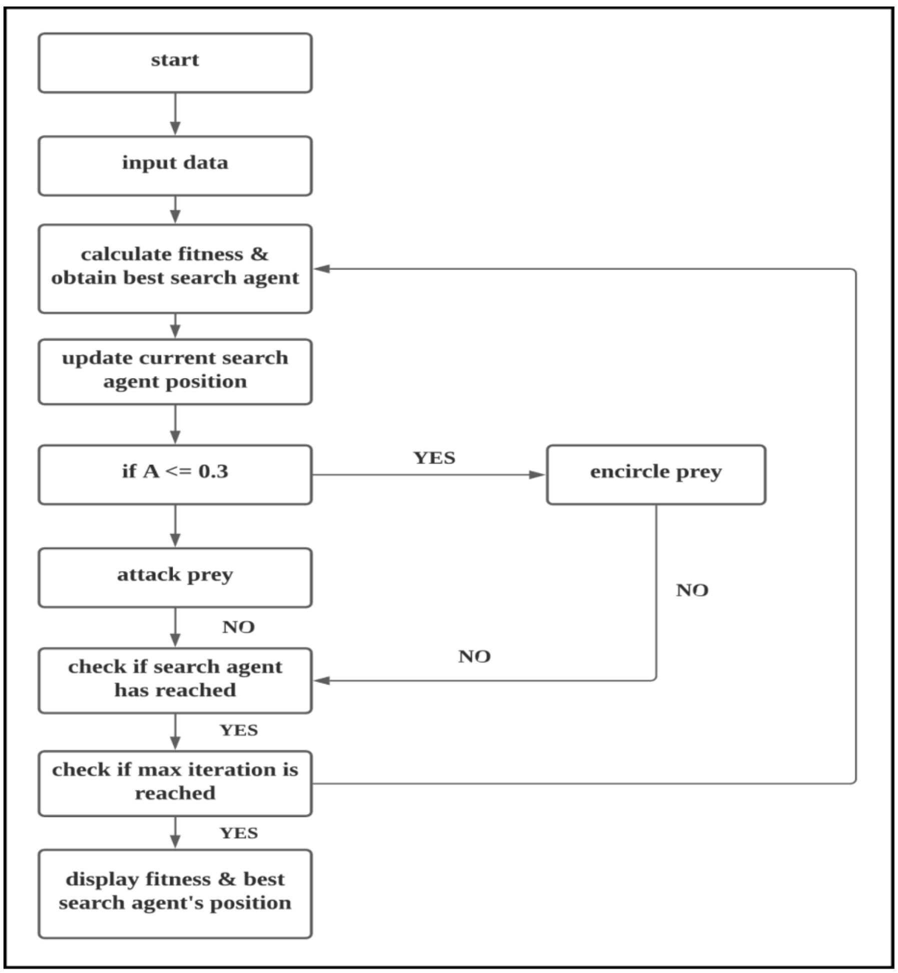


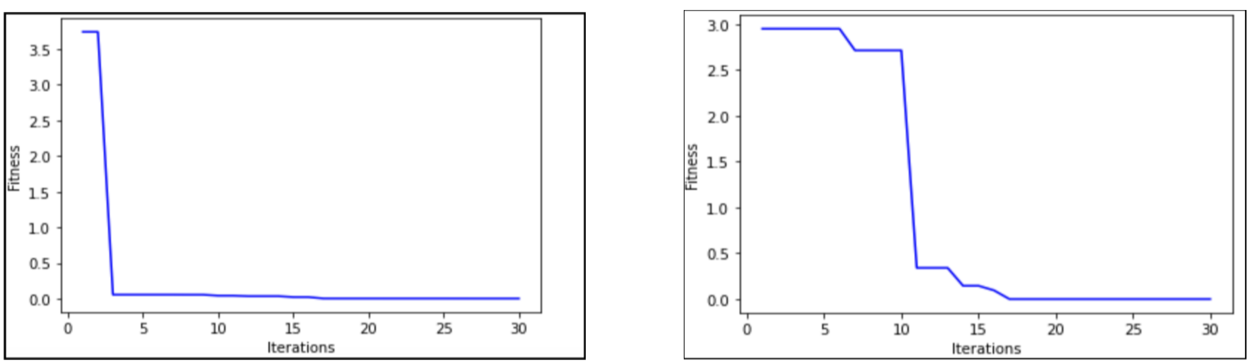

**Figure 1: Pipeline of Proposed Approach**

## 3. Results and PerformanceAnalysis

As seen in Table 1, on running both the original algorithm as well as the proposed improved methodology against four different benchmark functions of varying dimensions, it is observed that the proposed methodology outperforms the original Whale Optimization Algorithm for all test cases.

**Table 1.** Results returned by the original and proposed algorithms when run against four objective functions

| Sl. No | Objective Function | Original | Proposed |
|---|---|---|---|
| 1 | Booth | 0.000264 | 0.000622 |
| 2 | Matyas | 1.953538e-04 | 7.716957e-04 |
| 3 | Levy | 0.000225 | 0.000881 |
| 4 | Eggholder | -959.64066 | -958.95327 |



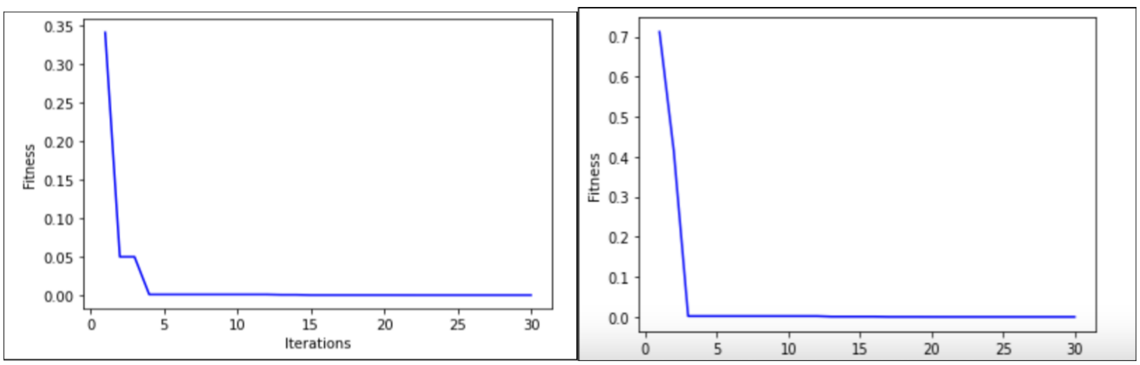

**(a)** **(b)**

**Figure 2. Iteration vs Fitness graph when run against the Booth objective function for (a) original algorithm**

**(b) proposed algorithm**

**Matyas Function**

As seen in Fig 3, the original algorithm is observed to return the optimal fitness value as 1.953538e-04, whiletheproposedalgorithmisobservedtoreturntheoptimalfitnessvalueas7.716957e-04.Sinceahigher fitnessvalueindicatesbetterperformanceofthemodel,theproposedmethodologyoutperformstheoriginal algorithm.

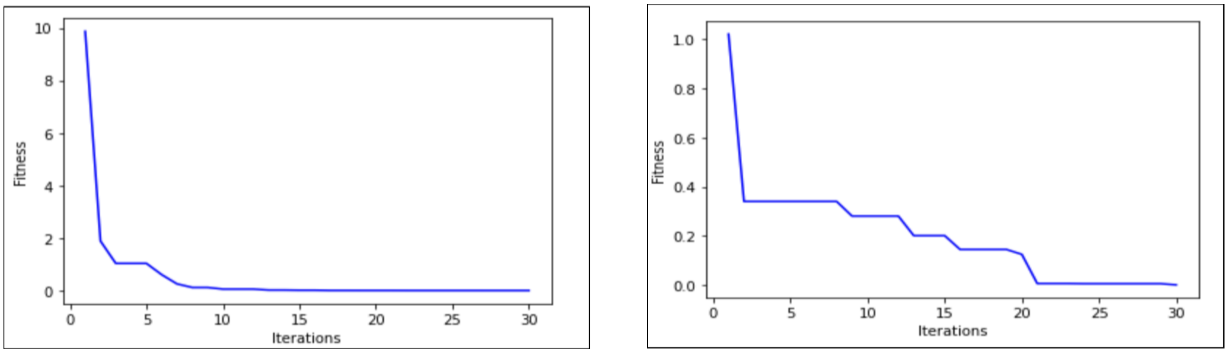



**(a)** **(b)**

**Figure 3. Iteration vs. Fitness graph when run against the Matyas objective function for (a) original algorithm; (b) proposed algorithm**

**Levy Function**

As seen in Fig 4., the original algorithm is observed to return the optimal fitness value as 0.000225, while the proposed algorithm is observed to return the optimal fitness value as 0.000881. Since a higher fitness value indicates better performance of the model, the proposed methodology outperforms the original algorithm.



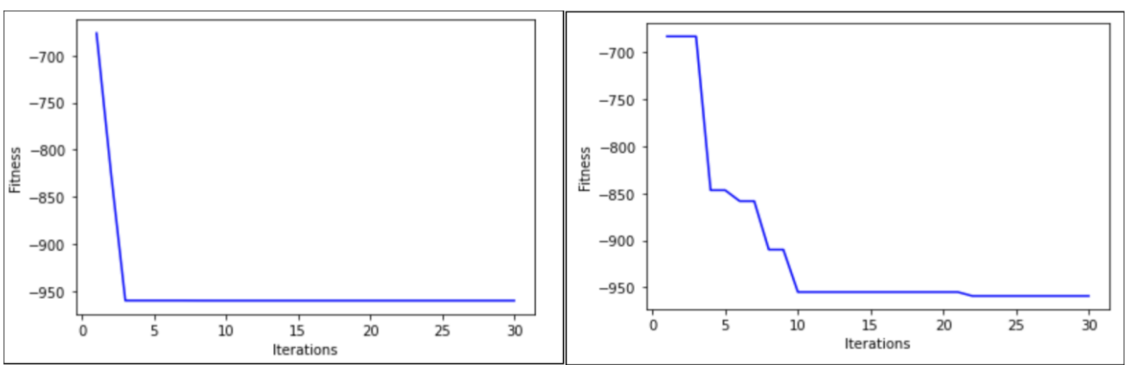

**(a)** **(b)**

**Figure 4. Iteration vs. Fitness graph when run against the Levy objective function for (a) original algorithm;**

**(b) proposedalgorithm**

**Eggholder Function**

AsseeninFig5.,theoriginalalgorithmisobservedtoreturntheoptimalfitnessvalueas-959.64066,while theproposedalgorithmisobservedtoreturntheoptimalfitnessvalueas-958.95327.Sinceahigherfitness value indicates better performance of the model, the proposed methodology outperforms the original algorithm.




**(a)** **(b)**

**Figure 5. Iteration vs. Fitness graph when run against the Eggholder objective function for (a) original algorithm; (b) proposed algorithm**

## CONCLUSION

Due to its simplicity, the WOA has found application in nearly every field of engineering and research. However, the basic algorithm does not allow for much flexibility and has limited performance. The algorithm also suffers from local optima stagnation and premature convergence. The proposed work helps provide an improved balance between the exploration and exploitation phases by using the current state of thesolutiontodetermineoperationperformedbythesearchagent(whale),whichimprovesthealgorithm’s overall efficiency and provides improved clustering accuracy. Future work could include further enhancement of the algorithm to achieve enhanced global convergence speed with better performance of thealgorithm.

## REFERENCES

[1] Agarwal, Parul, and Shikha Mehta. "Comparative analysis of nature inspired algorithms on data clustering." In 2015 IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), pp. 119-124. IEEE,2015.

[2] Olorunda O, Engelbrecht AP. Measuring exploration/exploitation in particle swarms using swarm diversity. In: Proceedings of the 2008 IEEE congress on evolutionary computation, CEC (IEEE world congress on computational intel- ligence); 2008. p.1128–34.

[3] Mirjalili,Seyedali,andAndrewLewis."TheWhaleOptimizationAlgorithm."Advancesinengineering software 95 (2016):51-67.

[4] Bindra, Kamalpreet, and Anuranjan Mishra. "A detailed study of clustering algorithms." In 2017 6th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO), pp. 371-376. IEEE,2017.

[5] Fister Jr, Iztok, Xin-She Yang, IztokFister, Janez Brest, and DušanFister. "A brief review of nature-inspired algorithms for optimization." arXiv preprint arXiv:1307.4186(2013).

[6] Lones, Michael A. "Mitigating metaphors: A comprehensible guide to recent nature-inspired algorithms." SN Computer Science 1, no. 1 (2020):1-12.

[7] Yang, Xin-She. "Nature-inspired optimization algorithms: Challenges and open problems." Journal of Computational Science 46 (2020):101104.

[8] Poli,Riccardo,JamesKennedy,andTimBlackwell."Particleswarmoptimization."Swarmintelligence 1, no. 1 (2007):33-57..

[9] Fister,Iztok,Xin-SheYang,andDušanFister."Cuckoosearch:abriefliteraturereview."Cuckoosearch and firefly algorithm (2014):49-62.

[10] Mareli, M., and B. Twala. "An adaptive Cuckoo search algorithm for optimization." Applied computing and informatics 14, no. 2 (2018):107-115.

[11] Mirjalili,Seyedali,SeyedMohammadMirjalili,andAndrewLewis."Greywolfoptimizer."Advances in engineering software 69 (2014):46-61.

[12] Karaboga, Dervis, and BahriyeAkay. "A comparative study of artificial bee colony algorithm." Applied mathematics and computation 214, no. 1 (2009):108-132.

[13] Trivedi, Indrajit N., Jangir Pradeep, JangirNarottam, Kumar Arvind, and LadumorDilip. "Novel adaptive Whale Optimization Algorithm for global optimization." Indian Journal of Science and Technology 9, no. 38 (2016):319-326.

[14] Bergstra, James, RémiBardenet, YoshuaBengio, and BalázsKégl. "Algorithms for hyper-parameter optimization." Advances in neural information processing systems 24(2011).

[15] Global Optimization Test Problems. Retrieved June 2013, from http://www-optima.amp.i.kyoto-u.ac.jp/member/student/hedar/Hedar_files/TestGO.htm

[16] Laguna, M., & Marti, R. Experimental Testing of Advanced Scatter Search Designs for Global Optimization of Multimodal Functions (2002). Retrieved June 2013, from http://www.uv.es/rmarti/paper/docs/global1.pdf.

[17] Global Optimization Test Functions Index. Retrieved June 2013, from http://infinity77.net/global_optimization/test_functions.html#test-functions-index.

[18] Global Optimization Test Problems. Retrieved June 2013, from http://www-optima.amp.i.kyoto-u.ac.jp/member/student/hedar/Hedar_files/TestGO.htm.