Indian Journal of Radio & Space Physics
Vol.50, June 2021, pp. 68-73

Detecting Autism spectrum disorder with sailfish optimisation

K Balakrishnan[a*], R Dhanalakshmi[a] & Utkarsh Mahadeo Khaire[b]

[a]Department of Computer Science and Engineering, Indian Institute of Information Technology Tiruchirappalli, Tiruchirappalli 620 012, India

[b]Department of Data Science and Intelligent Systems, Indian Institute of Information Technology Dharwad, Karnataka 580 009, India

Received: 12 February 2021; Accepted: 5 March 2021

Autism Spectrum Disorder (ASD), a neurodevelopmental disorder, has been a bottleneck to several clinical researchers due to data modularization, subjective analysis, and shifts in the accurate prediction of the disorder amongst the sample population. Subjective clinical research suffers from a lengthy procedure, which is a time-consuming process. In this paper, Sailfish Optimization (SFO), a recently developed nature-inspired meta-heuristics optimization algorithm, is being utilized to detect ASD. The hunting methodology of sailfish inspires SFO. Classical SFO has examined the search space in only one direction that affects its converging ability. The Random Opposition Based Learning (ROBL) strategy enhances the exploration capacity of SFO and successfully converges the predictive model to global optima. The proposed ROBL-based SFO (ROBL-SFO) selects relevant features from autism spectrum disorder (child and adult) datasets. According to the results obtained, the proposed model outperforms the convergence capability and reduces local-optimal stagnation compared to conventional SFOs.

Keywords: Autism, Random opposition-based learning, Sailfish optimization

1 Introduction

Autism is a neurodevelopment disorder with unique characteristics like social Interaction, improper behaviour, and communication with others. A study reveals that one child out of sixty-eight under the age of 8 and one adult out of 13 under sixty has autism in the United States of America[1]. Conventional clinical diagnosis involves a parent interview, a medical examination, a hearing test, observation, lead screening, speech and language evaluation, and early-stage sensory-motor assessment may reduce the chance of effect. Autism is a group of spectrum disorders with typical symptoms. Diagnosing autism is dissimilar in terms of autism in children and autism in adults[2]. Plenty of researchers provide better predicting algorithms in terms of accuracy with the ASD datasets and a real-world dataset[3].

The feature is an essential element in the machine learning classification problem. Selecting the best feature or relevant features will provide better accuracy in classification problems. Conventionally, most of the high dimensional datasets have more than 60000 features or attributes with fewer samples, not exceeding 100. The critical or relevant features can be identified using feature selection techniques viz., filter, wrapper, and embedded methods[4]. Specifically, the selection of features has an effect on obtaining the best outcomes with the precision of classification.

The Meta-Heuristic (MH) algorithm is more prominent in many engineering-related areas. Factors like uncomplicated and flexible implementation, local optima avoidance, and the independence of the gradient details makes MH most celebrated technique in the realm of Feature Selection (FS). It can be classified into four different groups, namely the Evolutionary algorithm (Genetic Algorithm[5] and Harmony Search[6]), Physics-based algorithm (Magnetic optimization algorithm[7] and Gravitational Search Algorithm[8]), Swarm-based algorithm (Particle Swarm Optimization[9] and Whale Optimization Algorithm[10]), and Human-based algorithm (Mine Blast Algorithm[11]). Out of these four classifications of meta-heuristic algorithms, Swarm optimization algorithms have the advantage of preserving the search space, rejecting any data when a new population has generated, and using less memory[12,13]. SFO[14] is a meta-heuristic algorithm inspired by a hunting technique of group of sailfish. The attack alternation approach of sailfish has been the core inspiration of the SFO algorithm. The

*Corresponding author (E-mail: bala.k.btech@gmail.com)

classical SFO suffers from stagnation at local optima and slower convergence speed because of the non-linear motion in prey search. Two groups of prey and predator populations have been used to represent the group hunting approach for SFO. The alteration attack breaks down the collective behaviour of the grouping policy. The prey movements are updated using the elite matrix, which selects the best position. In this article, the converging ability strengthened and increased SFO performance by introducing Random Opposition Based Learning (ROBL) into classical SFO. ROBL technique is used to improve exploration of the search space, motivated by the opposition between entities.

The major drawback of the classical SFO is that it explores the search space only in one direction, affecting its optimal performance and finding difficulty in a convergence of the model in global optima. Therefore, the ROBL strategy is used to improve the efficiency of the optimization algorithm and to substantially support the population to get out of the local minima.

2 Materials and Methods

This section provides mathematical modelling and the detail description of the classical SFO and Random Opposition-Based Learning.

2.1 Sailfish optimization

The SFO is a population-based MH optimization method driven by an attack-alteration of a group of hunting sailfishes that hunts a school of sardines. The SFO can search for the prey in a multidimensional search space. The sailfishes generate the random population which were then treated as a candidate solution. This algorithm observes two populations: 1) sailfish population, 2) sardine population. The sardine positions, elite, and injured matrix were used to find the best population. The alteration attack method is used to update the best sailfish and sardine position in the matrix. The optimization approach of SFO is provided in the pseudocode-1[14].

2.2 Random opposition-based learning

The initial population of random search agents is generated by MH algorithms based on prior knowledge. Due to the optimization algorithm's movement in only one direction of the search space, the algorithm may fall into local optima, and the result may not find optimal. The OBL technique effectively enhances the convergence capacity of MH algorithms, which overcomes the problem of optimality[15]. The search space can explore in both directions of the error surface. Consider o is random search agent range throughout [ub, lb] and is an opposite search agent, which is determined by Eq. 1

$$o = lb + ub - o \quad \dots (1)$$

where, lb and ub depict the random search agent's lower and upper boundaries, the opposite search agents are generated in n dimension search spaces using the Eqs 2 & 3.

$$o = o_1, o_2, \dots o_n \quad \dots (2)$$

$$\bar{o} = [\overline{o_1}, \overline{o_2} \dots \dots \overline{o_n}] \quad \dots (3)$$

Pseudocode-1: Sailfish Optimization

Initialize the random search agents for sailfish and sardine
Initialize the parameters
Find the fitness and build an elite matrix and injured matrix
while (Check termination criteria)
for each sailfish, calculate μ_i using Eq. (18)
Update the position using Eq. (17)
end for
Find the attack power using Eq. (21)
if attack power < 0.5
Calculate α and β using Eq.(22) and Eq.(23), respectively
Update the selected sardine position using Eq. (20)
else
Update the all-sardine position using Eq.(20)
end if
Find the fitness of all sardine
if there is a better solution in the sardine population, Replace the sailfish with injured sailfish using Eq.(24)
Remove the hunted sardine from the population
Update the best sailfish and best sardine
end if
end while
Return Best sailfish

The opposite direction values of $\bar{\mathbf{o}}$ is generated using the Eq4

$$\overline{o_j} = lb_j + ub_j - o \quad \dots (4)$$

The random OBL concept was used to improve the convergence ability, which-enhanced the diversity of the population. The Eq 4 was modified with the random values in between [0, 1]. The ROBL is mathematically represented in the below Eq5

$$\overline{o_j} = lb_j + ub_j * rand \quad \dots (5)$$

2.3 Dataset description

The performance of the proposed algorithm was experimented with autism spectrum disorder dataset for child and adult those were downloaded from internet. This ASD Diagnosis dataset contained 23 attributes which included one class attribute for predicting the disease and 22 features or attributes. Table 1 represents detailed description of the child's autism spectrum disorder dataset[16]. Tables (2&3) represent the number of samples available in child and adult dataset[17], both ASD and Non ASD patients.

2.4 The proposed framework

Figure 1 represents the proposed framework of ROBL-SFO. The proposed framework was carried out in three phases-Initialization phases, updating phase, and classification phase which is explained as follows.

2.4. 1 Initialization phase

The initial population of random search agents were generated during the initialization phase.SFO contains two different populations: 1) sailfish and 2) sardine. The random search agent of sailfish and sardine initial population is represented in Eqs 6 & 7, respectively.

$$RF\_Population = \begin{bmatrix} RF_{1,1} & \cdots & RF_{1,d} \\ \vdots & \ddots & \vdots \\ RF_{50,1} & \cdots & RF_{50,d} \end{bmatrix}_{50*d} \quad \ldots (6)$$

Table 1 — Dataset description[16]

Attribute	Values
A1 to A10	Yes, indicates value 1; No indicates value 0
Age	Value ranges from 1 to 80
Sex	Value 1 indicates male; Value 0 indicates female
Ethnicity	Aboriginal, White, Black, Hispania, Latino middle Eastern, South Asia, etc.
Jaundice	Yes, indicates value 1; No indicates value 0
Family ASD	Yes, indicates value 1; No indicates value 0
Residence	Different states and countries in Asia, South Asia, etc.
Used App Before	Yes, indicates value 1; No indicates value 0
Score	The value ranges from 0 to 10
Screening Type	1-3,4-11,12-16, 17 and above
Language	English, Russian, Spanish, French
User	Self, Parent, Relative, Others
ASD Class	Yes indicates value 1; No indicates value 0

$$R\_Population = \begin{bmatrix} R_{1,1} & \cdots & R_{1,d} \\ \vdots & \ddots & \vdots \\ R_{50,1} & \cdots & R_{50,d} \end{bmatrix}_{50*d} \quad \ldots (7)$$

The search agent (n) fixed as 50, and dimensions (d) depends on the input dataset. The opposite populations of sailfish and sardine are generated using the Eq 5 and represented in Eqs 8 & 9 respectively

$$\overline{RF\_Population} = \begin{bmatrix} \overline{RF_{1,1}} & \cdots & \overline{RF_{1,d}} \\ \vdots & \ddots & \vdots \\ \overline{RF_{50,1}} & \cdots & \overline{RF_{50,d}} \end{bmatrix}_{50*d} \quad \ldots (8)$$

$$\overline{R\_Population} = \begin{bmatrix} \overline{R_{1,1}} & \cdots & \overline{R_{1,d}} \\ \vdots & \ddots & \vdots \\ \overline{R_{50,1}} & \cdots & \overline{R_{50,d}} \end{bmatrix}_{50*d} \quad \ldots (9)$$

Finally, by integrating both sailfish populations (RF_Population and $\overline{RF\_Population}$), sardine populations (R_Population and $\overline{R\_Population}$), the proposed model produces the position matrix, which is represented in Eqs 10 & 11 respectively

$$RF\_Position\_matrix = \begin{bmatrix} PRF_{1,1} & \cdots & RPF_{1,d} \\ \vdots & \ddots & \vdots \\ PRF_{50,1} & \cdots & PRF_{50,d} \end{bmatrix}_{50*d} \quad \ldots (10)$$

$$R\_Position\_matrix = \begin{bmatrix} PR_{1,1} & \cdots & RP_{1,d} \\ \vdots & \ddots & \vdots \\ PR_{50,1} & \cdots & PR_{50,d} \end{bmatrix}_{50*d} \quad \ldots (11)$$

To determine each sailfish's fitness value and sardine is computed in Eq 11 based on Eqs 12 & 13, respectively.

$$RF_{fitness} = \begin{bmatrix} f(RF_{1,1} & \cdots & RF_{1,d}) \\ \vdots & \ddots & \vdots \\ f(RF_{100,1} & \cdots & RF_{100,d)} \end{bmatrix}_{100*d} = \begin{bmatrix} FRF_{1} \\ \vdots \\ FRF_{100} \end{bmatrix} \quad \ldots (12)$$

Table 2 — Autism spectrum disorder -Child[16]

Description	Count
Total number of records	292
Total number of positive autism cases	141
Total number of negative autism cases	151

Table 3 — Autism spectrum disorder -Adult[17]

Description	Count
Total number of records	704
Total number of positive autism cases	188
Total number of negative autism cases	515

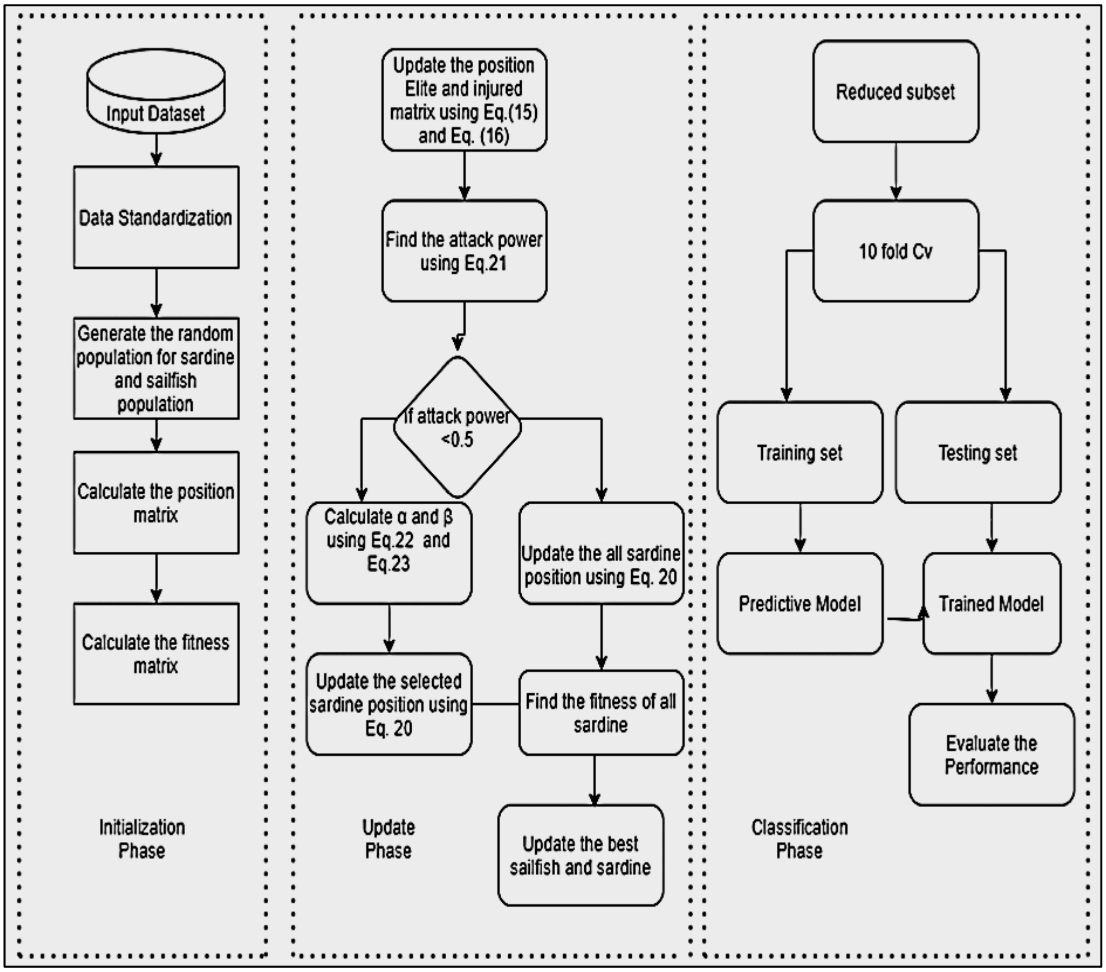

Fig 1 — The proposed framework of ROBL-SFO

$$R\_{fitness} = \begin{bmatrix} f(R_{1,1} & \cdots & R_{1,d}) \\ \vdots & \ddots & \vdots \\ f(R_{100,1} & \cdots & R_{100,d}) \end{bmatrix}_{100*d} = \begin{bmatrix} FR_1 \\ \vdots \\ FR_{100} \end{bmatrix} \quad \ldots (13)$$

where, f measures the fitness function, RF_fitness, and R_fitness saves each sailfish's fitness value and sardine and return the fitness or objective function value.

The proposed model maps the continuous values of the random search agents to binary by using the following equation

$$x_{i,j} = \begin{cases} 1, \frac{1}{1+e^{-x_{i,j}}} \\ 0, \text{otherwise} \end{cases} \quad \ldots (14)$$

2.4.2 Updating phase

The best position of sailfish in each iteration was saved as an elite matrix. Initially, the predator searched for prey and updated the elite matrix based on the best predator. For each iteration, the injured sardine location was also saved, and this sardine would be chosen as the best target for sailfish hunting. The elite matrix and injured position were represented in Eqs 15 & 16. Such positions would have a significant impact on SFO performance and avoid the wasting of time to re-discover previously discarded solutions.

$$RF\_elite = \begin{bmatrix} Y_{1,1} & \cdots & Y_{1,d} \\ \vdots & \ddots & \vdots \\ Y_{100,1} & \cdots & Y_{100,d} \end{bmatrix}_{100*d} \quad \ldots (15)$$

$$R\_injured = \begin{bmatrix} Z_{1,1} & \cdots & Z_{1,d} \\ \vdots & \ddots & \vdots \\ Z_{100,1} & \cdots & Z_{100,d} \end{bmatrix}_{100*d} \quad \ldots (16)$$

where, Y indicates the top predator. Here, both the predators and prey were considered a search agent because sometimes prey may also act as potential predators.

In general, the sailfish attack the prey when fewer contenders are available to attack in a particular region. The SFO algorithm illustrates an attack-alternating technique for sailfish when hunting in groups represented in Eq 17.

$$\mathrm{RF}^{i}_{new\_pos} = \mathrm{RF}^{i}_{elite} - \gamma_i \times \left(\mathrm{rand}(0,1) \times \left(\frac{\mathrm{RF}^{i}_{elite} + \mathrm{R}^{i}_{injured}}{2}\right) - \mathrm{RF}^{i}_{old}\right) \quad \dots (17)$$

where, RF^{i}_{old} is the current position. RF^{i}_{elite} is the best position of the sailfish $\mathrm{R}^{i}_{injured}$ is the best position of injured sailfish. $\mathrm{RF}^{i}_{new\_pos}$ is the newly updated position of sailfish. γ_i is the coefficient at an ith location which is generated below Eq 18

$$\gamma_i = 2 * \mathrm{rand}(0,1) * \mathrm{PD} - \mathrm{PD} \quad \dots (18)$$

where rand is the random value lies between [0,1]. The density of the prey in each iteration is calculated using the parameter PD. it is an important parameter to update the prey position of sailfish, which is calculated using the below formula

$$\mathrm{PD} = 1 - \left(\frac{N_{sf}}{N_s + N_{sf}}\right) \quad \dots (19)$$

where, N_{sf} represents a number of sailfish and N_S represents a number of sardines. The elite matrix and injured matrix position were updated on every iteration.

2.4.3 Hunting and catching prey

The best sailfish position was updated to a new best solution while hunting the sail and sardine fish at each iteration using Eqs 20 & 21, respectively.

$$\mathrm{R}^{i}_{new\_pos} = \left(\mathrm{rand}(0,1) \times \mathrm{RF}^{i}_{elite} + \mathrm{R}^{i}_{old} + \mathrm{pow}\right) \quad \dots (20)$$

$$\mathrm{pow} = C \times (1 - (2 * \mathrm{Iter} * \epsilon)) \quad \dots (21)$$

where, $\mathrm{R}^{i}_{new\_pos}$ is the newly updated position, RF^{i}_{elite} is the current best position, rand is an arbitrary number between 0 and 1, pow is an attack power of sailfish. To update the attack power, the parameters α and β are calculated using Eqs 22 & 23, respectively.

$$\alpha = No.of.sardine * pow \quad \dots (22)$$

$$\beta = v_i * \mathrm{Pow} \quad \dots (23)$$

Finally, the hunting of sailfish, a sardine, which is expressed as

$$U^{i}_{RF} = U^{i}_{R} \quad \dots (24)$$

where, U^{i}_{RF}, U^{i}_{R} represents the updated sailfish and sardine position, respectively.

2.4.4 Classification phase

In this step, the proposed model selects the top 7 features from the vector of selected features based on the feature occurrence using the Support Vector Machine (SVM). The autism dataset, both child and adult, is divided into a training part and testing part in addition to using the 10-fold cross-validation (CV) process.

3 Results and Discussion

The results obtained from ROBL-SFO were compared with the classical SFO and OBL-SFO. Figure 2 indicates the error rate of the prediction model throughout the epochs. The reduction of the

Fig. 2 — Comparison of converging ability of Classical SFO, OBL-SFO, and ROBL-SFO[14]

Table 4 — Classification accuracy of Classical SFO, OBL-SFO and ROBL-SFO[14]

Optimization method	Child	Adult
Classical -SFO	94.44	87.62
OBL-SFO	96.83	91.42
ROBL-SFO	97.30	94.20

error rate as the model progresses in each iteration demonstrated the convergence capacity of the proposed model towards the global minima. Compared to SFO and OBL-SFO, the proposed model converged faster towards the global minima for both the input datasets. The proposed model converged substantially towards the global optima after a few iterations that proved its efficiency in exploring and finding a better solution with different classifiers.

The performance analysis of the predictive models for the selected feature subset is summarized in Table 4 to validate the selected feature subsets. The performance of the algorithm for the classification is measured using the Eq 25

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} \quad \dots (25)$$

Where, TP, TN, FP, and FN represent the true positive, the true negative, the false positive and the false negative, respectively. Table 4 shows the ROBL-SFO result against conventional algorithms. The accuracy is more for the child than for the adult. While comparing the Classical SFO with ROBL, there was an increase in accuracy by 3.07%. The proposed ROBL-SFO worked efficiently with SVM since yielded higher accuracy for both the datasets.

From the results obtained, the identified significant contributions of the research are

(a) The incorporation of the ROBL in the searching strategy of SFO avoids the stagnation of the predictive model at local minima.

(b) In addition to the Random opposition based SFO, the SVM classifier is used to validate selected features based on the classification accuracy.

4 Conclusion

The proposed model outperforms classical SFO in the convergence rate and selecting the optimal subset of significant features. The 10-fold CV of the learning process guarantees that the model has not over fitted. From the results, it has been interpreted that the proposed model can effectively balance the trade-off between bias and variance. The proposed ROBL-SFO strategy enhances its exploration and exploitation capability that precludes local minima stagnation of the model.

Acknowledgement

This research was funded by the Department of Science and Technology, Government of India under the Interdisciplinary Cyber-Physical Systems (ICPS) scheme (Grant no. T-54).

References

1 Hull L, Petrides K V, & Mandy W, *Rev J Autism Dev Disord*, 7 (2020) 306.
2 Hassan M M, & Mohatar H M, *Inf Med Unlocked,* 16 (2019) 11.
3 Al-jawahiri R, & Miline E, *Peer J,* 5 (2017) 1.
4 Stańczyk U, & Jain L C, Feature Selection for Data and Pattern Recognition (Springer, Berlin, Heidelberg) 1st Edn, ISBN: 978-3-662-45619-4, 2015, p. 29.
5 Holland J H, Adaptation in Natural and Artificial Systems (MIT Press, United States) 1stEdn, ISBN:978-0-26208-213-6, 1975, p.317.
6 Lee K S, & Geem Z W, *Comput Methods Appl Mech Eng*, 194 (2005) 3902.
7 Zhu W Z, *Int J Adv Manuf Technol,* 15 (2020) 3881.
8 Zhu L, He S, Wang L, Zeng W, & Yang J, *IEEE Access,* 7 (2019) 114440.
9 Sahu B, & Mishra D, *Procedia Eng*, 38 (2012) 27.
10 Mirjalili S, & Lewis A, *Adv Eng Software*, 95 (2016) 51.
11 Yadav A, Sadollah A, Yadav N, & Kim J H, *Neural Comput Appl*, 32(7) (2020) 2423.
12 Mirjalili S, Gandomi A H, Mirjalili S Z, Saremi S, Faris H, & Mirjalili S M, *Adv Eng Software*, 114 (2017) 163.
13 Gharehchopogh F S, & Hojjat G, *Swarm Evol Comput*, 48 (2019) 1.
14 Tizhoosh H R, & Pantanowitz L, J Pathol Inform, 9 (38) (2018) 1.
15 Glover F, & Hao J K, *J Heuristics*, *25* (2019) 521.
16 https://archive.ics.uci.edu/ml/datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children.
17 https://archive.ics.uci.edu/ml/datasets/Autism+Screening+Adult#.