Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 7, July 2021: 3083- 3101

Novel Rider Hawk algorithm for feature selection in diagnosis of ASD using fMRI images

Arunekumar N B[#1], K Suresh Joseph[*2].

[#]Research scholar, Department of computer science , Pondicherry university, kalapet, Puducherry ,India

[2] Associate professor, Department of computer science , Pondicherry university, kalapet, Puducherry ,India

[1]arunekumarbala@gmail.com

[2] ksjoseph.csc@gmail.com

Abstract The autism spectrum disorder being a disease that affects the children and has a no proper medicinal cure till date the diagnosis and care is the only way to improve such children. The disease has a spectrum of disorders and therefore identifying the affected group is a challenging task. In this work a hybrid filter wrapper model for identifying the disease affected children using the fMRI is proposed. An isolation tree filter model is used along with the newly proposed rider hawk optimization algorithm for wrapper feature extraction. Feature selection is a great challenge in machine learning where the unwanted variables are found and averted to improve the accuracy of the algorithm. By combining the power of filter and wrapper feature selection algorithms the proposed pipeline is tested on ABIDE dataset that has rs-fMRI images of the patient's brain. An accuracy of 77.67% is obtained by using the proposed model for classification

Keywords— Rs-fMRI, ABIDE, Autism spectrum disorder, HHO, Artificial neural network.

I. Introduction

Autism spectrum disorder can be classified as a developmental disorder which may affect the cognition capability of children. This disease is termed to be a spectrum as it introduces various difficulties for communication, socializing, destruction of interests, repetition in the behaviour and so on in children. The disorder's prevalence rate is around 1% around the globe [1]. The probability of getting affected by ASD is higher for male children than female children. A huge challenge has been faced by the medical community to develop a proper drug that can cure ASD and till date there is no such proper drug for cure [2]. The drugs present can treat comorbidities but no direct drug for social communication enhancement is present. The early diagnosis will help in early intervention to decelerate the aggression, anxiety and to assist in the social communication. Once diagnosed early its essential to provide them with supportive environment. The paediatrician plays a key role in diagnosing autism with models like questionnaires which can't be so conclusive. Autism is an impairment of cognition which can be related to the functional activity of different parts of the brain and the correlation between those activities. The structural aspects of the brain are captured by sMRI i.e. the structural Magnetic resonance imaging which captures the spatial resolution in detail. From sMRI, the answer for where the concerned activity of a stimulus happens can be derived. The diagnosis which is the answer for question how an activity happens? Comprises both the temporal 'where ' and the spatial "when" bonded together. The temporal resolution gives the activity of the brain over a period of time which can be imaged by fMRI. The fMRI captures the BOLD blood oxygen level dependant signal that measures the change in paramagnetic deoxyhemoglobin from diamagnetic oxyhemoglobin.

The change occurs as a resulting process of activating the concerned part of brain. Such intensity changes in different regions of interests of the brain are recorded along the temporal resolution. The rs-fMRI which is a variant of fMRI modality records the activity in different parts of the brain while the patient is in resting state. The ABIDE [3] is one such rs-fMRI dataset which has been recorded among the people starting from the age of 7 to age of 64 with a median age of 14. The data set has equal representation for both classes. A total of 1112 individuals were 539 people with ASD and 573 people with typical control have been recorded for the study.

The dataset was subjected to C-PAC pre-processing pipeline that includes structural pre-processing such as stripping of skull, registration and segmentation based on the CC-200 atlas, normalisation and the functional pre-processing such as correction of slice time ,motion and band pass filtering between frequencies (0.01 and 0.1 Hz). After the pre-processing, extraction of connectome is done. The functional and structural connectomes are two different types of connectomes that are derived from the functional and structural MRI respectively. In this study the functional connectome is considered where each node depicts the grey matter volume of a ROI according to the CC200 atlas. For each individual the pre-processed ABIDE dataset comprises a 1-D activity signal for all the ROIs.

II. RELATED WORK

In recent years several meta-heuristic algorithms have been proposed for solving the parameters of MLP which one among the variants of ANN. The Meta heuristic algorithms can be classified into several types such as physics based, evolution based, Swarm intelligence based and so on. The swarm intelligence based algorithms mimic the swarming behaviour of animals for hunting. The PSO[4] swarm uses the global best and the personal best, GWO [5] uses hunting strategy of grey wolf, WOA [6] incorporates the spiral mechanism exhibited by humpback whales for hunting, BAT [7] algorithm mimics the echolocation behaviour of bats and ABC [8] uses the foraging technique of bees. The other predominantly used algorithms are FFA The authors [9] have proposed MLP-GSA with gravitational search algorithm to predict water level forecast for lake and compared with MLP-PSO, MLP-FFA, ARMA and ARIMA models. The MLP-LOA [10] uses lion optimization algorithm for choosing the count of hidden layers constrained with a maximum of two layers and the count neurons in each of it , momentum , learning rate , transfer function among the 3(Relu, thanh, sigmoid) and finally the weights and biases of the network.

The MLP-ALO was tested over classification task of 10 UCI datasets against MLP-DE and MLP-GA and other popular machine learning algorithms such as SVM, Naïve Bayes and decision tree. The stream water flow[11] is predicted using the MLP optimised by the IWD algorithm where and compared over different error functions of RMSE, MAE, RE and E. The EBat algorithm [12] is proposed by adding a mutation operator to enhance the exploratory behaviour. The proposed EBat is utilised in the MLP parameter optimisation for intrusion detection system over 10 different datasets with 10 existing optimization algorithms. Sigmoid for nonlinearity and MSE was used for error function.

The metrics used for comparison were accuracy, false alarm rate and detection rate. The oppositional grass bee (OGBEE) algorithm [13] is a fusion of opposition based algorithm, grasshopper algorithm and artificial bee colony algorithm fused together. Along with the complete characteristics of opposition based learning and GOA the employer bee characteristics of ABC is utilised. The OGBEE was tested over 13 benchmark functions of 30 dimensions and a multimodal sentiment analysis data set that comprises features from audio, video and text. The softmax is used as output function and the proposed algorithm is compared over other ANN and ELM models.

The other most popular algorithms are [14] cuckoo search algorithm, [11]intelligent water drop algorithm that works based on the flow of water drop, [15]grey wolf optimizer that works based on the encircling behaviour of wolf pack, [16] moth flame optimization, [17] JAYA algorithm , [18] Multi verse optimizer , [19] sine cosine algorithm and Salp swarm algorithm[20]. The HHO algorithm is based on the collaborative hunting behavior of the Harris hawk. As the HHO bifurcates the exploration and exploitation phase into strict halves the authors [21] have proposed an interchangeable phase management using an information exchange and modifying the linear escape energy parameter with chaotic behaviours.

The paper [22] fuses the behaviour of Salp Swarm algorithm and HHO , where first half of the Swarm uses the SSA and the second half uses the HHO. The proposed algorithm was tested over 36 benchmark functions and multi-level threshold image segmentation. The proposed GCHHO Algorithm [23] uses the Gaussian mutation and cuckoo search dimension logic operator to mutate selective dimensions along with the HHO algorithm. The proposed algorithm is tested over the CEC 2017 benchmark functions. (Dimension decided HHO with GM). (Modification of HHO with random distribution for optimum power flow) instead of the uniform distribution used in the levy flight, other distribution with boundary value was introduced. The proposed algorithm was being tested over 13 CEC benchmark functions and IEEE-30 bus test functions.

In the boosted binary HHO [24] the population is boosted with SSA mechanism after each round and greedy selection of agents is carried out for the next iteration of HHO update . The algorithm was tested over benchmark functions (23) and K-NN wrapper feature selection over 10 datasets The behaviour of GOA and HHO was compared [25] over training ANN for soil compression prediction both the algorithms were tested over RMSE , MAE and R^2 error functions , Where MAE was found best in that case .The hybrid HHO-SCA algorithm used the HHO and then sequentially the SCA for updating the agents. The algorithm is tested over CEC2017, 2018 functions and 10 other engineering design problems

III. METHODS

A. HHO algorithm

The rabbits are the preferred prey for the hawks, where these rabbits also perform actions to escape the several hunting strategies performed by hawks. This ecosystem of hunting by hawks and escaping by rabbits is being modeled by the naïve HHO algorithm to solve single objective optimization problems. The hawks are modeled as the agents while rabbit is modeled as the best solution found till current iteration. The overall hunting strategies of the hawks are bifurcated into exploration and exploitation phases. The exploration phase constitutes two operators while the exploitation phase happens in four sub-phases.

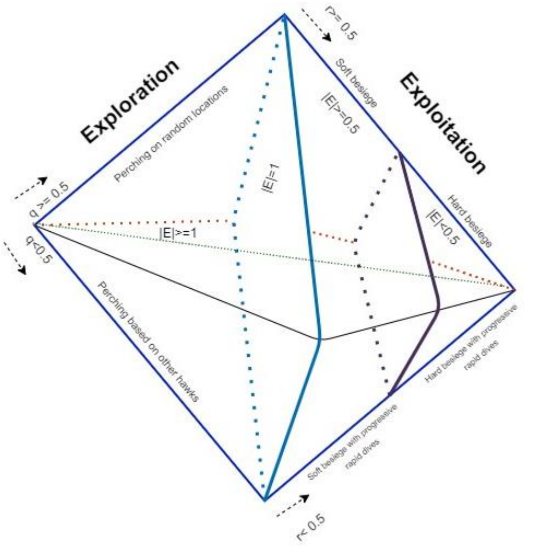

Figure 1 Harris Hawk Optimization Algorithm

1) The exploration phase

With equal chance of opportunity on exploratory iterations, two operators constituted with random parameters are utilized by agents as in Eqn (1)

$$x(i+1)\begin{cases} X_{rand}(i) - a_1 \, | X_{rand}(i) - 2a_2 \, X(i)|, & d < 0.5 \\ (X_{rab}(i) - X_{avg}(i)) - a_3 \, (LB + a_4 \, (UB - LB)) & d \geq 0.5 \end{cases} \quad (1) \qquad (1)$$

The X_{rand} location of specifies any random agent while X_{rab} and X_{avg} are location of rabbit and average value of all agents in swarm. The parameters $a_1,..,a_5$ and d are random values within range (0,1).

$$X_{avg}(i) = 1/N \sum_{i=1}^{N} X_n(i) \qquad (2)$$

The transition between exploration and exploitation phase is decided by the parameter E as in Eqn (3) which is influenced by parameter E_α that takes random values between (-1,1).

$$E = 2E_\alpha (1 - \frac{i}{1}) \qquad (3)$$

2) Exploitation phase:

In contrast to the exploration phase, the exploitation focuses on the target area to find more fine-tuned solutions. The HHO utilizes four different strategies for exploitation that are decided by the parameter E and random parameter a. As the exploitation is carried out after fixing the vague target area it mostly revolves around the location of rabbit.

3) Soft besiege

Current solution updates itself with respect to location of rabbit and its jumping strength as given in Eqns.(4), (5) and (6) where the variable 'a' is random parameter.

$$X(i+1) = dX(i) - E^{*} | j^{*} X_{rab}(i) - X(i) | \qquad (4)$$

$$dX(i) = X_{rab}(i) - X(i) \qquad (5)$$

$$J = 2(1 - a_5) \qquad (6)$$

4) Hard besiege

The attack is more targeted towards the rabbit location as given in Eqn.(7)

$$X(i+1) = X_{rab} - E^{*} | dX(i) | \qquad (7)$$

Soft besiege and progressive rapid dive: The update strategy is decided based on agent's fitness, where the agent will choose between a rabbit location based soft besiege using Equation (8) and rapid dive

$$G = X_{rab}(i) - E^{*} | J^{*} X_{rab}(i) - X(i) | \qquad (8)$$

The rapid dives are calculated using levy flight mechanism as given in Eqn.**(9**), where S is a D dimension vector and LF is the levy flight function as given in Eqn.

(10)

$$H = G + S^{*} LF(D) \qquad (9)$$

$$LF(x) = \frac{(u^{*}\sigma)}{|v|^{(1/\beta)}}$$

$$\sigma = \frac{\Gamma(1+\beta)^{*}\sin?(\frac{\pi\beta}{2})}{\Gamma(\frac{1+\beta}{2})^{*}\beta^{*}2^{(\beta-1)/2}} \qquad (10)$$

The u and v are random parameters and value of β is given as 1.5. The decision to choose in-between G or H from Equation (8) and Equation (9) respectively is given by Equation (11)

$$X(i+1) = \begin{cases} G \text{ if } F(G) < F(X(i)) \\ H \text{ if } F(H) < F(X(i)) \end{cases} \qquad (11)$$

Hard besiege and rapid dives: Similar to the early condition, here the agent chooses between hard besiege as given in Equation (12) and levy flight.

$$G' = X_{rab} - E \mid J * X_{rab}(i) - X_{avg}(i) \mid \quad (12)$$

$$H' = G' + S * LF(D) \quad (13)$$

$$X(i+1) = \begin{cases} G' \text{ if } F(G') < F(X(i)) \\ H' \text{ if } F(H') < F(X(i)) \end{cases} \quad (14)$$

B. Rider optimization algorithm

The rider optimization algorithm is based on an artificial scenario of four rider groups which chase the target based on the characteristics of their own. Each of the riders has separate characteristic which it follows to attain the target. Total population is divided into these four groups for searching in both the exploration and exploitation phase. The four types of riders are bypass rider, follower rider, overtaker rider and attacker rider. The bypass rider uses more random variables to bypass the others, follower rider follows the leader to navigate, the overtaker rider works on the basis of parameters like coordinate selector, direction indicator and success rate. Finally the attacker rider updates based on the current position and the leader position.

IV. PROPOSED METHODOLOGY

A. Pre-processing

The pre-processing is an important feature of the image processing pipeline where the unwanted features or noise present in the image must be removed. The fMRI data is a 3-D image where the unwanted layers of the image must be removed in the pre-processing step. The fMRI has spatial and temporal aspects embedded in it so it’s necessary to remove and clean the data in both the domain using structural pre-processing and functional pre-processing.

B. Structural pre-processing

The structural pre-processing is the process of removing unwanted segments in the spatial domain and retrieving the necessary parts alone. Some of the important steps in structural pre-processing are

1) Skull stripping:

The raw MRI image is recorded with the bone structure along with the brain which includes the skull. Therefore to analyse the brain skull is an unwanted artefact which can be removed by segmenting and striping it.

2) Segmentation:

The brain comprises of numerous regions. In order to analyse the functional specification of the brain individually it's essential to segment the brain into different parts therefore a segmentation process is carried out using the FSL tool. After the segmentation process of tissues is over, the brain is registered into a standard CC200 atlas. Once registration is completed now different regions can be easily identified.

C. Functional pre-processing

The functional pre-processing removes the unwanted features that are present in the temporal domain of the data. The functional pre-processing correct the anomalies present in the timeline of the data.

1) Slice time correction

The scanner equipment scans the image of the head in an interleaved manner. Therefore each layer of the head is recorded with some time delay than its strip above and below. In order to correct such problem the slice time correction is carried out.

2) Motion Correction

Even if the scan is acquired with caution there may be minute motion in the head of the patient during the process of scanning. This difference in scanning must be averted by motion correction where the brain is registered at a point and the variations are nullified by correction from that registered point.

3) Normalization

The brain image may have a vast scale of grey scale intensity values. It will be hard on computation while using the un normalized values. Therefore a normalization of mean intensity is carried out

4) Band pass filtering

The band pass filtering is the process of removing signals that are not into certain range. A band pass filtering of (0.01-0.1HZ) was utilized.

Apart from the important pre-processing step some other steps such as the removal of the nuisance signal, registration of functional images over anatomical space and standard space registrations were carried out. After all these pre-processing activities were carried out the regions of interest are extracted.

D. Feature extraction

The cleaned data is itself of very large dimension. Processing the entire data is humongous and needs extreme computation power for such activity so it's better to extract some meaning full features of low dimensions such that the classification of the data could be done easily. The feature extraction is where all possible features from the given data are being extracted.

1) Region of interest

The different regions of interest according to the Atlas which is being registered are being extracted from the fMRI of the brain. For this work around 200 regions are being considered. The Figure 2 shows 20 regions of the brain considered from the atlas.

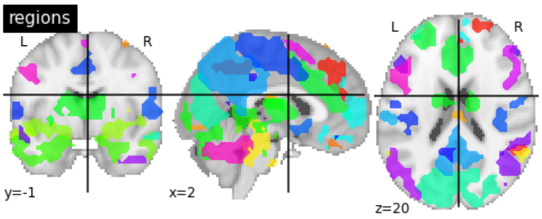

Figure 2 Regions of interest

2) BOLD signal extraction

From these extracted regions the intensity of the bold signals are being taken into consideration. This bold signal was observed as given in the figure. 3

These signals record the level of oxygen that has been present during the process of imaging the brain. As the recorded images are resting state fMRI, they are the ones which have been recorded for a stipulated amount of time when the patient is at resting state. The fMRI images were registered over the CC200 atlas and 200 time series signals for the oxygen level of regions were extracted for each patient.

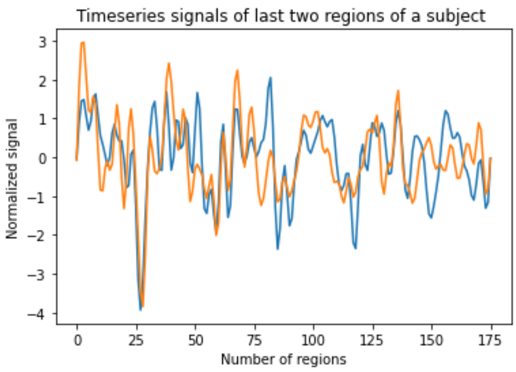

Figure 3 Bold Signals

3) Correlation matrix

After the extraction of this time series signal a partial correlation using the Karl Pearson correlation coefficient is being calculated among all the regions. The partial correlation is the measure of finding correlation between two variables by ignoring the other variables. Therefore a connectivity matrix is formed with the correlation coefficient between each pair of variables is constructed as in Figure 4. Once the correlation matrix is being constructed the upper triangular Matrix or the lower triangular matrix is being extracted as the correlation matrix is symmetric. The extracted matrix is flattened to form an array of features. The final vector with correlation pair between the regions is of 20000 dimensions in length which is a very huge dimension for a low complex model to classify.

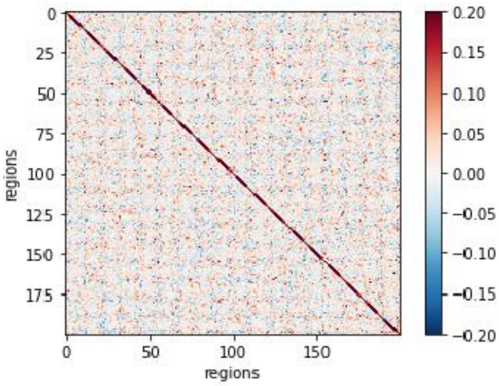

Figure 4 Correlation matrix

In order to reduce the complexity of algorithm and throw out unwanted features the isolation forest feature regression is used.

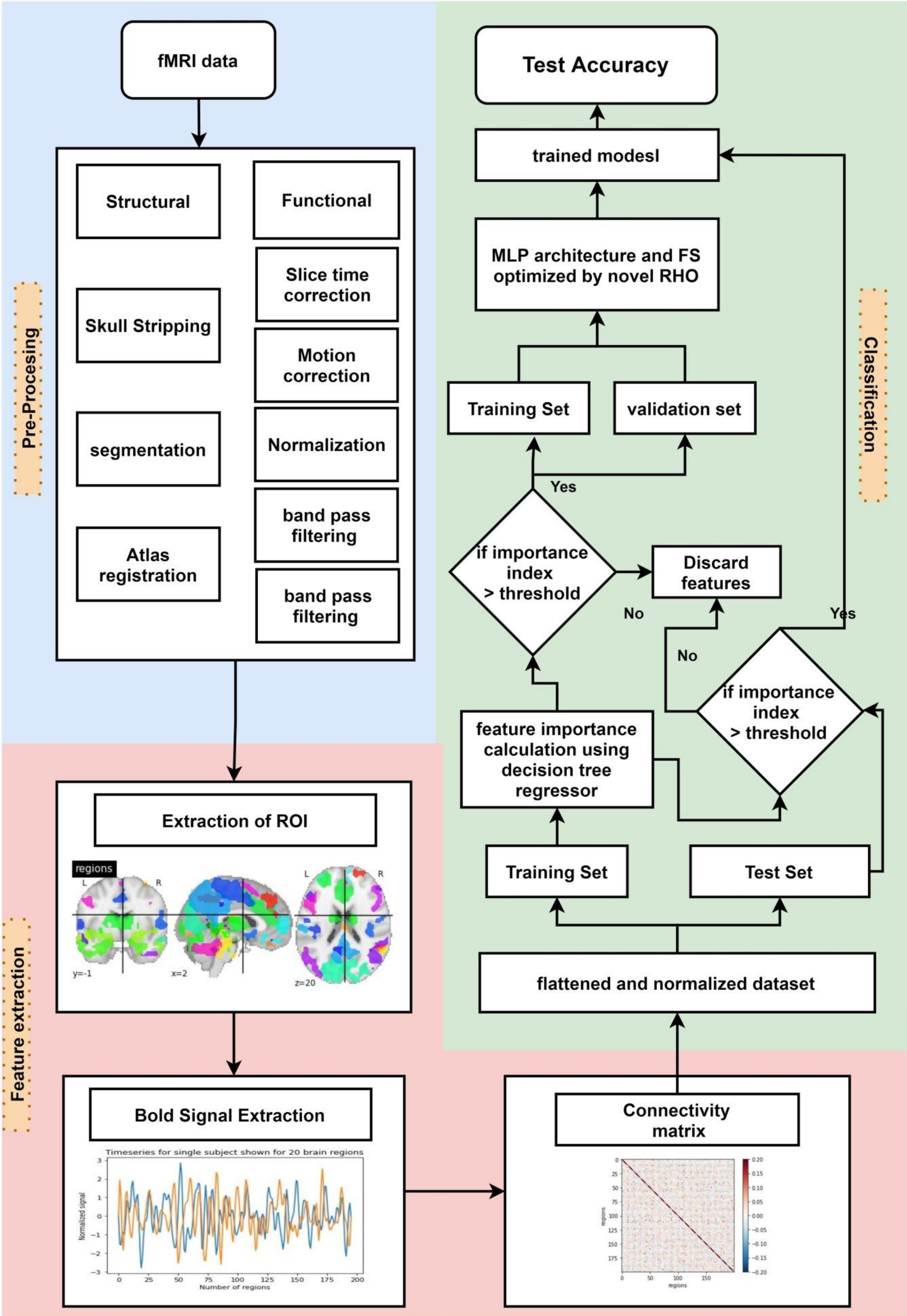

Figure 5 Flow diagram of the pipeline for ASD diagnosis using rs-fMRI dataset

The isolation forest regression is a decision tree based gini impurity algorithm for ranking the importance of each variable. Once the importance of those variables are being ranked, threshold is being fixed based upon the telescopic search. After the least important features are being discarded the most important features are being subjected to classification using the multilayer perceptron model whose architecture is being optimised by the newly proposed rider hawk optimisation algorithm.

E. Rider Hawk optimization algorithm

The features of the HHO algorithm and the ROA algorithm are fused together in order to form the RHO algorithm in the exploration phase. The bypass rider is introduced instead of one among the exploration equations of HHO algorithm and the follower rider methodology is used along with the Levy flight based on the choice of probability. The algorithm of the ROA is given in the Table 1

Table 1 Pseudo code RHO

Rider Hawk Optimization algorithm
Initialize the population with random parameters ***While*** (stopping criteria is not met) ***do***: Calculate fitness and get target rabbit ***for***(every solution) ***do***: Update aiding parameters E, J ***if1*** exploration state |E|≥1 then: ***if2*** q_1 >0.5 do: Update using bypass rider ***else if2*** q_1≤ 0.5 do: Update using Eqn.1 End ***if2*** ***else if1*** exploitation state |E|< 1 then: ***if3*** soft besiege (r ≥0.5 and |E|≥ 0.5): ***if4*** q_2 >0.5 do: Update using follower rider ***Elseif4*** q_2≤ 0.5 do: Update using Eqn.4 End ***if4*** ***if3*** hard besiege (r ≥0.5 and |E|< 0.5): ***if5*** q_3 >0.5 do: Update using follower rider ***else if5*** q_3≤ 0.5 do: Update using Eqn.7 End ***if5*** ***if3*** hard besiege & rapid dive (r < 0.5 and |E|≥ 0.5): Update using Eqn.11 ***if3*** hard besiege & rapid dive (r < 0.5 and |E|< 0.5): Update using Eqn.14

End ***if3***
End ***for***
End ***While***
Return X_{rabbit}

F. Artificial neural network

The artificial neural networks can easily find good approximates of nonlinear and complex functions. Therefore it can easily adapt to complex dataset and mine out the solutions according to it. There is no thumb rule to define the architecture for a problem and there are exhaustive numbers of choices in the number of hidden layers, no of neurons per hidden layer, nonlinearities and number of iterations needed.

Therefore to mine out the exact architecture that would give the best accuracy usually a grid search will be performed which is highly time consuming with huge number of choices therefore we use proposed Rider Hawk optimization algorithm for finding the best architecture .

G. Wrapper feature selection

The feature selection is the process of getting a subset of feature variables without any unwanted variables the wrapper feature selection uses a feedback from the classifier to judge the necessity of a feature. The feature selection being an NP hard problem and the complexity increases exponentially as the number of features increases an efficient optimizer is needed to solve the problem. The proposed rider hawk optimization algorithm is used along with the parameter selection for wrapper selection. The numbers of wrapper parameters are equal to the number of features in the pre-processed vectors

H. Objective encoding

The combined objective functions optimize the architecture of ANN as well as the feature selection. The combined encoding of a solution vector for the problem can be given as shown in the figure below

<table>
<tr><td colspan="6">ANN architecture parameters</td><td>Wrapper Feature Selection parameters</td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7…….n</td></tr>
</table>

Figure 6 Encoding of solution

In the above Figure 6 the allocation of each dimension is given as

- 1^{st} dim = non linearity index
- 2^{nd} dim = no of layers (max =3)
- 3^{rd} dim = no of iterations (max=1000)
- 4^{th}, 5th, 6th dims= no of hidden neurons in 1^{st}, 2^{nd} and 3^{rd} neurons respectively.
- From dimensions 7 to ‘n’ are the feature selection masks of wrapper feature selection

V. RESULTS

F n n o.	metri c	RHO	HHO	GWO	DE	MFO	BAT	JAYA	SCA	WOA	PSO	SSA

The raw rs-fMRI images from the dataset are being pre-processed and a flattened vector of 20 k dimension was obtained later. To decrease the size of the dimensions a feature importance using the isolation forest was used and the importance was

F1	**mean**	**2.78E-18**	5.96E-12	87.56521	46842.13	37109.62	72258.43	11024.1	24701.06	0.044083	2794.891	2459.882
	stddev	**9.47E-18**	3.08E-11	38.73525	11783.26	7257.83	22446.29	3737.689	16506.63	0.214587	1523.713	638.4491
	rank	**1**	**2**	**4**	**10**	**9**	**11**	**7**	**8**	**3**	**6**	**5**
F2	**mean**	**3.13E-08**	8.42E-08	3.230767	111.2081	157.6576	1.77E+20	71.02851	24.7275	0.00053	117.8179	29.71294
	stddev	**1.39E-07**	2.51E-07	0.844816	15.85695	32.4061	9.42E+20	34.30533	10.36341	0.000957	27.86629	3.053488
	rank	**1**	**2**	**4**	**8**	**10**	**11**	**7**	**5**	**3**	**9**	**6**
F3	**mean**	**3.67E-09**	0.599154	19981.72	96910.85	109746.7	434846.4	133525.5	150983.2	546101.8	33572.99	12466.31
	stddev	**1.54E-08**	3.219691	7963.74	31303.7	26625.36	330343.6	29920.19	48621.17	264018.1	14935.66	6271.523
	rank	**1**	**2**	**4**	**6**	**7**	**10**	**8**	**9**	**11**	**5**	**3**
F4	**mean**	**2.96E-10**	1.09E-06	28.57227	85.2262	83.09765	80.88897	87.77209	91.18032	85.70905	28.05118	19.96543
	stddev	**7.54E-10**	5.72E-06	8.166503	8.932328	2.997734	12.10654	5.483372	3.993993	10.18003	3.713969	2.794858
	rank	**1**	**2**	**5**	**8**	**7**	**6**	**10**	**11**	**9**	**4**	**3**
F5	**mean**	29.61284	**7.807199**	9183.093	1.06E+08	78503926	2.55E+08	32082011	1.40E+08	56.84693	670265.5	368263.6
	stddev	23.17171	**12.76071**	5667.574	45988828	29007783	1.22E+08	18083406	98745368	36.06359	491622.3	178974
	rank	**2**	**1**	**4**	**9**	**8**	**11**	**7**	**10**	**3**	**6**	**5**
F6	**mean**	**0.000485**	0.091062	134.6017	47968.37	35033.09	87776.41	11549.89	21571.71	7.553149	2498.807	2608.083
	stddev	**0.001652**	0.106444	74.19952	11403.49	8589.018	26500.75	4227.144	13078.75	1.675557	1097.306	735.2882
	rank	**1**	**2**	**4**	**10**	**9**	**11**	**7**	**8**	**3**	**5**	**6**
F7	**mean**	**0.003383**	0.004126	0.272038	91.82258	73.92072	202.3861	24.10631	112.5413	0.099861	142.9495	0.952684
	stddev	**0.002477**	0.004153	0.107618	39.95281	34.37677	100.1203	11.03536	53.51604	0.141165	47.21665	0.263446
	rank	**1**	**2**	**4**	**8**	**7**	**11**	**6**	**9**	**3**	**10**	**5**
F8	**mean**	-11896.2	-1.67E+04	-6623.5	-9403.14	-11030.9	**-1.10E+34**	-5338.33	-3875.07	-13590	-3170.74	-9646
	stddev	3767.267	3796.093	1936.137	1480.483	873.2097	**5.12E+34**	791.9875	501.5971	2461.679	716.5121	1005.545
	rank	**4**	**2**	**8**	**7**	**5**	**1**	**9**	**10**	**3**	**11**	**6**
F9	**mean**	**5.69E-15**	2.27E-14	234.2423	418.263	440.4844	644.1051	436.9628	285.6052	33.69046	551.09	177.556
	stddev	**2.25E-14**	1.03E-13	86.81036	66.42143	51.27108	61.63364	61.30767	104.1711	102.106	49.32758	20.18075
	rank	**1**	**2**	**5**	**7**	**9**	**11**	**8**	**6**	**3**	**10**	**4**
F	**mean**	**1.41E-**	5.68E-	3.82435	18.6811	19.8497	19.9184	16.2158	19.2727	0.00720	9.65878	9.32555

Table 2 Fitness Comparison of proposed algorithms over benchmark functions

F 10												4
	stddev	**3.83E-09**	1.40E-07	0.415764	0.866683	0.525721	0.555312	1.76107	2.489267	0.01321	2.045464	0.781999
	rank	**1**	**2**	**4**	**8**	**10**	**11**	**7**	**9**	**3**	**6**	**5**
F 11	mean	**1.85E-17**	3.73E-13	2.133529	420.2201	331.3551	673.5804	113.9352	176.8499	0.219178	186.0259	25.74854
	stddev	**9.96E-17**	1.62E-12	0.654975	93.97967	80.02675	179.4719	43.06082	123.6165	0.390089	32.89508	5.677891
	rank	**1**	**2**	**4**	**10**	**9**	**11**	**6**	**7**	**3**	**8**	**5**
F 12	mean	**3.69E-05**	0.002118	12.98622	2.10E+08	1.36E+08	4.26E+08	56285635	3.60E+08	4.882803	2114.12	30.35768
	stddev	**9.96E-05**	0.004861	7.870987	1.34E+08	1.03E+08	2.89E+08	55606598	1.97E+08	7.37467	4060.876	15.53614
	rank	**1**	**2**	**4**	**9**	**8**	**11**	**7**	**10**	**3**	**6**	**5**
F 13	mean	0.105875	**0.039882**	67.03332	4.92E+08	3.68E+08	8.95E+08	1.09E+08	8.31E+08	9.166112	150869.4	40017.89
	stddev	0.171613	**0.059462**	35.10944	2.44E+08	2.19E+08	6.02E+08	72384949	3.26E+08	11.52432	195051	50210.04
	rank	**2**	**1**	**4**	**9**	**8**	**11**	**7**	**10**	**3**	**6**	**5**
	average rank	**1.3846154**	1.8461538	4.4615385	8.3846154	8.1538462	9.7692308	7.3846154	8.6153846	4.0769231	7.0769231	4.8461538
	final rank	**1**	**2**	**4**	**9**	**8**	**11**	**7**	**10**	**3**	**6**	**5**

threshold to 0.001. On threshold of importance the dimensions of features were reduced to 58 which is the size of the wrapper feature selection parameter vector. The feature selection along with the neural network architecture parameters with a total of 64 dimensions were given to the proposed optimizer for finding the best architecture and the best solution.

Table 3 **Comparison of results for ASD diagnosis**

The experimentation for the proposed algorithm has been conducted in two phases where the first phase is to test the proposed novel rider hawk algorithm against the other optimization algorithms. For such test the proposed algorithm is being tested over a set of 13 non-biased CEC2014 benchmark functions. The bench mark functions comprise of uni modal and multi modal functions. The proposed algorithm is tested with benchmark functions for 50 dimensions and 500 iterations. This experiment is conducted for 30 times and the average and standard deviation for the 30 runs is given in the table the proposed algorithm is also tested against the other existing optimization algorithms such as the HHO [24], GWO [26], DE[27], MFO[16],BAT [7],JAYA[28], SCA[29],CS [30],PSO [31] and SSA [32] algorithms

From the table it’s evident that the proposed RHO algorithm has performed in most of the functions where it could not find the best solutions in F1,F8 and F13.

In the second phase of the experimentation the objective function for identification of ASD is coined and used. The objective function consists of two objectives which are to find the highest accuracy and with the minimal number of element in the feature subset. The objective function can be given as in Equation (15) has two components combined but does not constraint anything on the architecture parameter of the network. The exponential function increases the importance of the accuracy. By this objective it’s constrained to reduce the number of features with a good accuracy.

$$\begin{aligned} Minimize f(x) &= 0.99 * \exp(5(1 - accuracy)) \\ &+ 0.01 * (selectedfeaturs / totalfeatures)) \end{aligned} \quad \mathbf{(15\ )}$$

The proposed algorithm is used to optimize both the parameters and find the best solution.

The proposed RHO has optimized the architecture and the features such that an accuracy of 77.67% is obtained. The architecture defined by the algorithm is given having 114 neurons for the first layer and 29 neurons in the second layer. The activation chosen was tanh and the total iterations were 263. The wrapper feature selection has also been done where the algorithm has chosen 30 dimensions over the given 58 dimensions.

	Accuracy	Iters	Activation
RHO	**77.6%**	263	tanh
HHO	73.3%	131	relu
GWO	71.5%	183	relu
DE	66.9	93	relu
MFO	64%	393	relu
BAT	53.9%	92	logistic
JAYA	68.3%	183	logistic
SCA	67.6%	150	relu
WOA	72.6%	289	logistic
PSO	70.5%	176	relu
SSA	69.7	216	tanh

VI. Discussion and Limitations

From the results in the table it’s clearly seen that the proposed algorithm has achieved the least fitness on 77% of the functions. The proposed algorithm has achieved a near zero solution that is, less than 0.01 fitness in nearly 69% of the functions and less than 0.2 fitness in one function. Even in the function which it has not obtained the best fitness the proposed algorithm has performed quite well. Even in the 0.23% of the functions which it has not performed well. it has obtained second lowest fitness values in 0.15% of functions and a considerable performance in the function 8. The RHO algorithm has out beaten the GWO, DE, MFO , JAYA, SCA, WOA, PSO and SSA algorithm by 100 % on benchmark functions. The RHO has out beaten the HO algorithm over 84% of the functions and BAT algorithm over 92% of the algorithms. Proposed algorithm has chosen an architecture that is deep and of two layers such that the encoding of can be done. The neurons in the first layer are less than the second layer which is 114 and 29 respectively. Therefore a good encoding will happen. Later the tan h activation is considered here and iterations of 236 is restricted where in grid search will take a huge number of time. And a reduction of 51.7% has been achieved by the algorithm. With the best accuracy of 77.67% the algorithm has out beaten the existing works where the authors [33] used Graph neural networks along with the phenotype data to obtain an accuracy of 70.4%. The authors in the work [34] used random forest with subsets of ABIDE dataset and obtained a accuracy of 73.5% . The authors of the work [35] used subsets of the dataset individually and obtained accuracies ranging from 75% to 95 %. But the work wasn’t using the entire dataset. In the presented work the complete data set is taken into consideration and a unified model for classification is proposed with accuracy of 77.67%.

A. Limitations

The proposed algorithm uses the samples only from CC200 atlas after registration. The other atlases have not been considered and therefore for an extensive work using the other atlases can be considered in the future for the classification of the ASD disease.

VII. CONCLUSION

A novel Rider hawk optimization algorithm is proposed based on the features of rider optimization algorithm and the HHO algorithm. By fusing the algorithm's characteristics it can be clearly seen from the table that the proposed algorithms has outperformed other existing algorithms over the 13 standard CEC 2014 benchmark functions. For diagnosing the Autism Spectrum Disorder a pipeline for feature extraction and is used and the vector of connectivity is derived. From this vector an algorithm a pipeline for feature importance filtering and wrapper feature selection using the RHO is proposed. Along with the feature selection the proposed algorithm is also used to estimate the parameters of the ANN and thus an accuracy of .. is obtained over the abide dataset for diagnosing ASD

VIII. REFERENCES

[1] F. R. Volkmar and D. Pauls, "Autism," in *Lancet*, 2003, vol. 362, no. 9390, pp. 1133–1141.

[2] J. T. Mccracken *et al.*, "Drug development for Autism Spectrum Disorder (ASD): Progress, challenges, and future directions-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)," *Eur. Neuropsychopharmacol.*, vol. 48, pp. 3–31, 2021.

[3] G. Deshpande, L. E. Libero, K. R. Sreenivasan, H. D. Deshpande, and R. K. Kana, "Identification of neural connectivity signatures of autism using machine learning," *Front. Hum. Neurosci.*, no. OCT, Oct. 2013.

[4] T. Stephan and K. Suresh Joseph, "Particle Swarm Optimization-Based Energy Efficient Channel Assignment Technique for Clustered Cognitive Radio Sensor Networks," *Comput. J.*, vol. 61, no. 6, pp. 926–936, Jun. 2018.

[5] H. Faris, I. Aljarah, M. A. Al-Betar, and S. Mirjalili, "Grey wolf optimizer: a review of recent variants and applications," *Neural Computing and Applications*, vol. 30, no. 2. Springer London, pp. 413–435, 01-Jul-2018.

[6] F. S. Gharehchopogh and H. Gholizadeh, "A comprehensive survey: Whale Optimization Algorithm and its applications," *Swarm Evol. Comput.*, vol. 48, pp. 1–24, Aug. 2019.

[7] N. B. Arunekumar, A. Kumar, and K. S. Joseph, "Hybrid bat inspired algorithm for multiprocessor real-time scheduling preparation," in *2016 International Conference on Communication and Signal Processing (ICCSP)*, 2016, pp. 2194–2198.

[8] X. Chu *et al.*, "An artificial bee colony algorithm with adaptive heterogeneous competition for global optimization problems," *Appl. Soft Comput. J.*, vol. 93, Aug. 2020.

[9] M. A. Ghorbani, R. C. Deo, V. Karimi, M. H. Kashani, and S. Ghorbani, "Design and implementation of a hybrid MLP-GSA model with multi-layer perceptron-gravitational search algorithm for monthly lake water level forecasting," *Stoch. Environ. Res. Risk Assess.*, vol. 33, no. 1, pp. 125–147, 2019.

[10] P. Bansal, S. Gupta, S. Kumar, S. Sharma, and S. Sharma, "MLP-LOA: a metaheuristic approach to design an optimal multilayer perceptron," *Soft Comput.*, vol. 23, no. 23, pp. 12331–12345, 2019.

[11] Q. B. Pham *et al.*, "Hybrid model to improve the river streamflow forecasting utilizing multi-layer perceptron-based intelligent water drop optimization algorithm," *Soft Comput.*, vol. 24, no. 23, pp. 18039–18056, 2020.

[12] W. A. H. M. Ghanem and A. Jantan, *A new approach for intrusion detection system based on training multilayer perceptron by using enhanced Bat algorithm*, vol. 32, no. 15. Springer London, 2020.

[13] S. Bairavel and M. Krishnamurthy, "Novel OGBEE-based feature selection and feature-level fusion with MLP neural network for social media multimodal sentiment analysis," *Soft Comput.*, vol. 24, no. 24, pp. 18431–18445, 2020.

[14] R. Rajabioun, "Cuckoo Optimization Algorithm," vol. 11, pp. 5508–5518, 2011.

[15] H. Joshi and S. Arora, "Enhanced grey Wolf optimization algorithm for global optimization," *Fundam. Informaticae*, vol. 153, no. 3, pp. 235–264, 2017.

[16] A. Darwish, "Bio-inspired computing: Algorithms review, deep analysis, and the scope of applications," *Futur. Comput. Informatics J.*, vol. 3, no. 2, pp. 231–246, Dec. 2018.

[17] S. Rekha, D. Debahuti, and M. Minakhi, "A hybridized ELM using self-adaptive multi-population-based Jaya algorithm for currency exchange prediction : an empirical assessment," *Neural Comput. Appl.*, vol. 8, 2018.

[18] G. Jain, G. Yadav, D. Prakash, A. Shukla, and R. Tiwari, "MVO-based path planning scheme with coordination of UAVs in 3-D environment," *J. Comput. Sci.*, vol. 37, p. 101016, Oct. 2019.

[19] Y. Li, Y. Zhao, and J. Liu, "Dimension by dimension dynamic sine cosine algorithm for global optimization problems," *Appl. Soft Comput.*, vol. 98, Jan. 2021.

[20] M. Tubishat *et al.*, "Dynamic Salp swarm algorithm for feature selection," *Expert Syst. Appl.*, vol. 164, p. 113873, 2021.

[21] C. Qu, W. He, X. Peng, and X. Peng, "Harris Hawks optimization with information exchange," *Appl. Math. Model.*, vol. 84, pp. 52–75, Aug. 2020.

[22] M. A. Elaziz, A. A. Heidari, H. Fujita, and H. Moayedi, "A competitive chain-based Harris Hawks Optimizer for global optimization and multi-level image thresholding problems," *Appl. Soft Comput. J.*, vol. 95, p. 106347, 2020.

[23] S. Gupta, K. Deep, A. A. Heidari, H. Moayedi, and M. Wang, "Opposition-based learning Harris hawks optimization with advanced transition rules: principles and analysis," *Expert Syst. Appl.*, vol. 158, p. 113510, 2020.

[24] Y. Zhang, R. Liu, X. Wang, H. Chen, and C. Li, “Boosted binary Harris hawks optimizer and feature selection,” *Eng. with Comput. 2020*, pp. 1–30, May 2020.

[25] H. Moayedi, M. Gör, Z. Lyu, and D. T. Bui, “Herding Behaviors of grasshopper and Harris hawk for hybridizing the neural network in predicting the soil compression coefficient,” *Measurement*, vol. 152, p. 107389, Feb. 2020.

[26] X. Zhang, X. Wang, H. Chen, D. Wang, and Z. Fu, “Improved GWO for large-scale function optimization and MLP optimization in cancer identification,” *Neural Computing and Applications*, vol. 32, no. 5. Springer, pp. 1305–1325, 01-Mar-2020.

[27] J. Liang *et al.*, “Multimodal multiobjective optimization with differential evolution,” *Swarm Evol. Comput.*, vol. 44, pp. 1028–1059, Feb. 2019.

[28] S. O. Degertekin, G. Yalcin Bayar, and L. Lamberti, “Parameter free Jaya algorithm for truss sizing-layout optimization under natural frequency constraints,” *Comput. Struct.*, vol. 245, Mar. 2021.

[29] N. Li and L. Wang, “Bare-Bones Based Sine Cosine Algorithm for global optimization,” *J. Comput. Sci.*, vol. 47, Nov. 2020.

[30] S. Mirjalili and A. Lewis, “The Whale Optimization Algorithm,” *Adv. Eng. Softw.*, vol. 95, pp. 51–67, May 2016.

[31] K. Mistry, L. Zhang, S. C. Neoh, C. P. Lim, and B. Fielding, “A Micro-GA Embedded PSO Feature Selection Approach to Intelligent Facial Emotion Recognition,” *IEEE Trans. Cybern.*, vol. 47, no. 6, pp. 1496–1509, Jun. 2017.

[32] H. Faris *et al.*, “An efficient binary Salp Swarm Algorithm with crossover scheme for feature selection problems,” *Knowledge-Based Syst.*, vol. 154, no. January, pp. 43–67, 2018.

[33] S. Parisot *et al.*, “Disease prediction using graph convolutional networks: Application to Autism Spectrum Disorder and Alzheimer’s disease,” *Med. Image Anal.*, vol. 48, pp. 117–130, 2018.

[34] M. A. Reiter, A. Jahedi, A. R. J. Fredo, I. Fishman, B. Bailey, and R. A. Müller, “Performance of machine learning classification models of autism using resting-state fMRI is contingent on sample heterogeneity,” *Neural Comput. Appl.*, vol. 33, no. 8, pp. 3299–3310, Apr. 2021.

[35] C. Wang, Z. Xiao, and J. Wu, “Functional connectivity-based classification of autism and control using SVM-RFECV on rs-fMRI data,” *Phys. Medica*, vol. 65, pp. 99–105, Sep. 2019.