Enhanced Whale Optimization and Ensemble Learning Algorithm for Autism Spectrum Disorder Classification

S. Saravana Kumar[1], K. Selvakumar[2], V. Senthil Murugan[3]

[1]Research Scholar, Department of Information Technology, Faculty of Engineering and Technology, Annamalai University, Chidambaram, India, saravana.coolya@gmail.com
[2]Professor, Department of Information Technology, Faculty of Engineering and Technology, Annamalai University, Chidambaram, India, kskaucse@gmail.com
[3]Department of Networking and communications, Faculty of Engineering and Technology, SRM Institute of science and Technology, Kattankulathur, India, senthilv5@srmist.edu.in.

Autism spectrum disorder (ASD) is an array of neurodevelopmental illnesses characterized by difficulties with sociability and interaction and restricted repetitive activities. The amount of people with ASD has surged in recently, and the underlying etiology of the condition is still unknown. Although ASD cannot be cured, early discovery is ideal since it makes mitigation care more effective. Given what extent the signs are, an ASD diagnoses may give in the future, although it usually occurs around the age of two. Classification accuracy and feature selection are not substantially guaranteed in the current approach. To address these problems, Enhanced Whale Optimization Algorithm (EWOA) and Ensemble Modified Auto Encoder Convolution Neural Networks (EM-AECNN)-Artificial Neural Network (ANN) system are suggested for ASD classification n excellently. The suggested study contains pre-processing, feature selection, and classification. This work contains three main modules such as pre-processing, feature selection and categorization. Data pre-processing is a technique is executed through Synthetic Minority Oversampling Technique (SMOTE)which is utilized by eliminating the ASD dataset's unnecessary data. The EWOA method is utilized to choose the pertinent characteristics from the autism dataset during the feature selection process. By choosing the optimal fitness features, the objective function of EWOA is employed to increase classification accuracy. In order to categorize the accurate ASD outcomes, the EM-AECNN-ANN technique is ultimately employed. Given the testing outcome, it is determined that the suggested EM-AECNN-ANN algorithm provides greater precision, accuracy, recall and f-measure than the current approaches.

Keywords: Autism Spectrum Disorder (ASD), Enhanced Whale Optimization Algorithm (EWO) and Ensemble Modified Auto Encoder Convolution Neural Networks (EM-AECNN)-Artificial Neural Network (ANN).

1. Introduction

ASD is a neurological disease that has an impact on action and interaction. Although an autism assessment can be obtained at any age, autism is still considered a "developmental disorder". given that the initial two years of life are often when a significant number of signs emerge. There are significant differences in the standards and timelines utilized for the diagnosis of autism. Families have to wait up to 13 months [1] among the first screening and the diagnosis; if they belong to a minority group or have a lower socioeconomic level, this period may be much longer. The provision of speech and behavioural services, which have major beneficial effects on a child's development, particularly if provided swiftly, is directly impacted by these delays [2] [3]. Additionally, there can be significant differences in how different clinicians who execute formal diagnoses recognize the severity of the phenotypic. As a result, a significant portion of the population receives a diagnosis after the growth windows during which behavioural intervention probably had the biggest impact on life's quality and the development that proceeded.

According to WHO data, 0.63% of children have ASD [4] and it begins in childhood and progresses to teenage and adulthood. In general, neuro developmental illnesses manifest in their early years (5), and their treatments come at a high expense. ASD is categorized by stereotyped behaviour and a chronic absence of interpersonal communication, which typically ensues by a general deterioration in communication skills [5]. ASD is related to both genetic and neurological causes. The ability to think and imagine, socializing, repetitive behaviours, and problems with personal contact are all examples of ASD-related traits. It affects behaviour and education as well as how individuals communicate themselves and relate to others. The signs and warnings first appear in a very young child. There is currently no cure for this chronic condition. ASD has a significant financial impact due to the rise in the incidence of the disorder worldwide as well as the challenge and cost of patient identification.

A crucial step in identifying effective feature subsets (more discriminating) and raising dataset quality is feature selection. Utilizing a human brain-specific gene network as the foundation, [6] created a supplementary machine learning (ML) technique to present an estimate of autism risk factors across the entire genome. Utilizing the brain-specific network and these genome-wide estimates, it was shown that the majority of ASD genes converge on a limited amount of important brain phases of development. Probable mediators of ASD across numerous copy-number variations and detected probable pathogenic genes among frequently occurring copy-number variants related to autism were highlighted. The optimization-based algorithm is introduced for gene selection which is utilized to select the relevant and significant features from the original dataset.

The goal of the categorization approach is to accurately identify the target class for the provided input dataset. The training set and the values of a categorizing attribute are utilized to categorize the data. trained a Support Vector Machine (SVM) model for growing brain gene expression data in [7] utilizing certain features. To rank putative autism genes, a mean sensitivity of 74.4% and accuracy of 76.7% are found. In [8], build a single brain network for depicting features and classify ASD/TC using a Deep Neural Network (DNN) classifier.

Initially it builds a unique brain network for every participant and determines the connectivity among every two ROIs. Secondly, the F-score is utilized to rank the joining features in descending ranking, and the traits that rank highest are preferred. Then, DNN classifier is utilized to execute ASD categorization utilizing the chosen 3000 top features.

The objective of this study is the ASD categorization applying ensemble learning algorithms. ASD detection accuracy is not substantially guaranteed despite a plethora of study and approaches being provided. The current methods have shortcomings in terms of duration and imprecise outcomes for ASD detection. These issues can be resolved by Enhanced Whale Optimization Algorithm (EWOA) algorithm and Ensemble Modified Auto Encoder Convolution Neural Networks (EM-AECNN)-Artificial Neural Network (ANN) classifier is suggested to enhance prediction accuracy. Pre-processing, feature selection utilizing EWOA, and categorization utilizing EM-AECNN-ANN are the goal of this research. With the assistance of effective techniques and the provided autism dataset, the suggested strategy produces findings that are more accurate.

The balance of the work is arranged as follows: In Section 2, a succinct overview of the literature on ASD feature selection, pre-processing, and categorization methods is provided. In Section 3, the suggested EM-AECNN-ANN approach is described. Section 4 comprises the experimental data and a description of the outcome analysis. Section 5 provides a summary of the conclusions.

2. Related work

In [9], Yuan et al (2017) strive to create a solid ML technique for identifying autism by the use of natural language processing approaches derived from data taken from possible ASD medical records. This detection system consists of pre-processing, learning content representation, categorization, and the conversion of semi-structured and formless health records into digital format. Test outcomes are compared to the ground truth established by knowledgeable doctors and obtains a very promising 91.1% recall and 83.4% accuracy. The system for detecting ASD could greatly streamline and expedite the process of diagnosing ASD.

In [10], Sharawi et al (2017) introduced a feature selection using WOA where ideal feature subsets, were identified using wrapper-based approaches. Utilizing this method, the optimal feature subset that maintains the fewest features while optimizing the quality of classification was identified.

Kosmicki et al [11] developed ML methods achieved 98.27% and 97.66% accuracy, accordingly, for 9 out of 28 behaviors that were collected for component 2 items and 12 out of 28 behaviours that were collected for phase 3. With only a minor loss in accuracy, the number of activities decreased by more than 55% across both phases. This suggests that methods utilized to expedite the detection of ASD risk. Large-scale diagnoses can be facilitated by the application of mobile and preliminary risk evaluations.

Alzubi et al [12] recommended dependable hybrid feature choices to identify the best subsets of SNPs and gain important information. The CMIM approach and SVM-RFEs serve as the foundation for their suggested solution. They evaluated it with m RMR, CMIM, and Relief F feature choices using SVM, LDA, NB, and k-NN. The findings determine the effectiveness of the chosen feature selection strategy, which outperforms all other feature selection techniques tested and attains 89% classification accuracy for the utilized dataset.

Hossain et al. [13] developed a diagnosis technique utilizing the already available classification methods to achieve improved diagnosis goals. To find the optimal feature sets and categories, the researchers looked at ASD datasets for adults, teens, and toddlers. The results of their tests showed that categorization that utilized MLPs outperformed previously used techniques and attained 100% accuracy with fewer parameters. Furthermore, "relief F" chose superior features.

In [14], Konget al (2019) constructed a single brain network as a feature depiction, then classify ASD/TC using a DNN classifier. Initially for every patient, build a brain network and extract linking properties among every pair of ROIs. The interaction variables are sorted applying an F-score in descending order, selecting those attributes that have a superior ranking. A DNN is utilized during ASD/TC categorization utilizing the chosen 3000 top features. Ten-fold cross-validation was executed to assess the technique utilizing T1-weighted MRI images from the ABIDE I. The approach can classify ASD/TC data with an accuracy of 90.39% and an AUC of 0.9738, according to its findings. Comparing the trial findings shows that the suggested approach works better in ASD/TC categorization than certain cutting-edge techniques.

3. Proposed methodology

EWOA identifies important properties that can support effective categories in ASD datasets, and this study effort presents ensemble M-AECNN-ANN driven categories and dimensionality reductions for the identification efficiency of ASDs. The three primary stages of the suggested method are feature selection, categorization, and data pre-processing. Fig. 1 displays the suggested method's general block diagram.

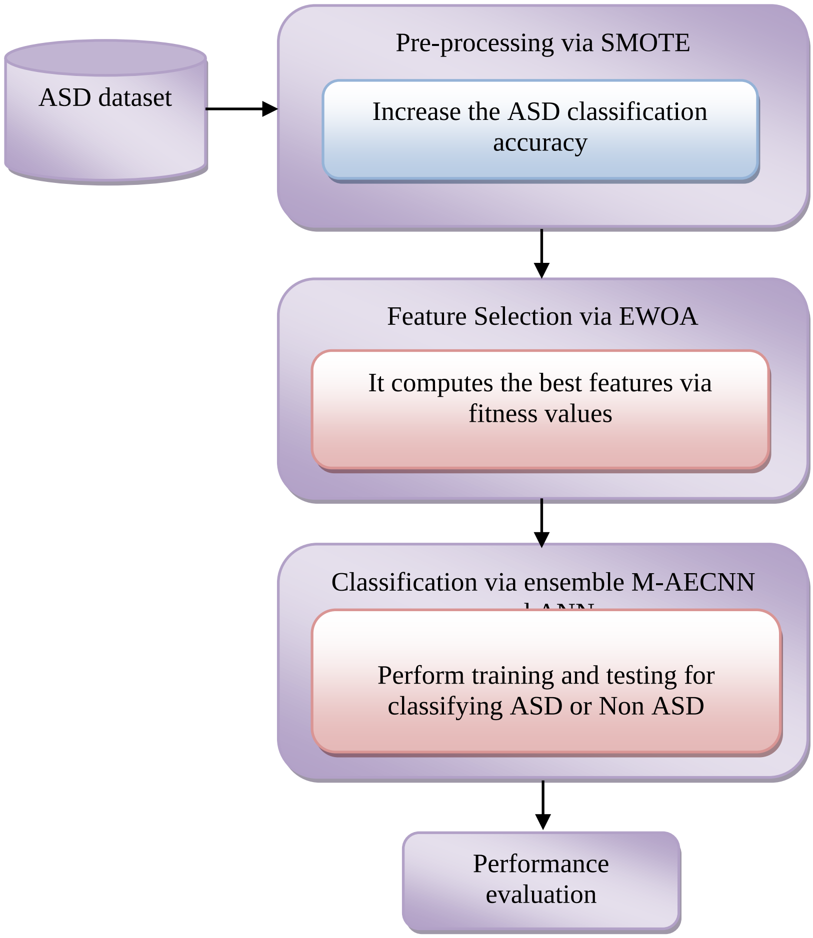


Fig 1 Overall block diagram of the proposed system

3.1 Pre-processing via Synthetic Minority Oversampling Technique (SMOTE)

Here, Pre-processing is carried out to improve the reliability of classification of the ASD dataset. Pre-processing the data is a common first step before training and analyzing the data utilizing ML techniques. A ML method is only as effective as the data it is provided. Data must be well formatted and contain key elements to have sufficient consistency to produce the best results [15]. Normalizing data are significant pre-processes for ML methods. They aim to remove some of the irrelevant features that act as distinguishing markers between different data sets. The majority of methods of classification attempt to improve prediction performance by obtaining clean samples to work with and by drawing as distinct a boundary as feasible between each class. When it comes to learning to classify synthetic cases most classifiers found that it is simpler to classify cases that are not close to the border. SMOTE is a new and innovative pre-processing technique that present for unbalanced training sets that utilizes these results. SMOTE generalization makes an effort to produce pure synthetic samples and accurately pinpoint the border. The two steps of our recommended procedure are as follows:

First stage, Artificial instances produced through the SMOTE technique and the method [16]:

$$N = 2 * (r - z) + z \qquad (1)$$

hereN signifies original synthetic counts, rimplies counts of majority classes, and z represents counts of minority class samples.

Second stage, artificial cases that are closest to the borderline and artificial samples generated by SMOTE that have a stronger link to a larger class than the minority are removed. Following the cleaning of the supplied data, the feature selection process explained in the following section is executed.

3.2 Feature selection using Enhanced Whale Optimization Algorithm (EWOA)

In this work, The EWO is employed to select features. Finding pertinent features from the ASD data to distinguish between ASD and non-ASD features is the purpose of feature selection. For many optimization issues, the WOA is utilized to choose pertinent attributes and determine the best solution. WOA method concentrated on choosing pertinent characteristics and demonstrated that its effectiveness may be modified to yield superior outcomes. WOA is utilized for selecting features in ASD datasets for accurate categorization.

Humpback whale hunting tactics and the development of an equation for those tactics served as the inspiration for the WOA approach. The whales hunt using a bubble-net approach, circling and feeding on their food, which is mostly tiny fish. Deep below the fish, whales emerge and begin to surface, forming a large ring of bubbles. The fish are compelled to surface by the bubbles acting as a trap. Whales pursue fish that are rising to the surface [17]. There are three steps to the hunting process in theory: exploitations, circling, and explorations where explorations dictate the quality of fishes consumed. Phase of encirclement: Whales locate the fish and encircle them. The ideal location's starting point is initially undefined and chosen at random. Other agents update their positions in response to the random initiations, and revised positions become best places to reach targets.

The variation of individuals is decreased during the WOA exploitation phase since it merely mimics and learns from the actions of the present best solution. The process of learning from random individuals during the WOA's exploration phase suffers from some blindness and ineffective information flow between groups, which has an impact on the method's pace of convergence. Therefore, the poor convergence speed of the conventional WOA is problematic. EWOA is suggested as a solution to the aforementioned issue in order to accelerate convergence.

In animal swarms, individuals gain knowledge from both neighbouring members which can significantly enhance personal qualities. By determining a person's social ranking, social impact, and social network, the adaptive social learning technique [18] forms the person's neighbourhood. This technique is utilized to build the whales' adaptive neighbourhood to enhance connections among groups, and a novel strategy utilizing neighbourhood updates is created to increase population diversity while preserving computation accuracy.

Equations (2) and (3) can be used to depict the location of the whale and the surrounding area. (3)

$$\vec{H} = |\vec{C}x\vec{X} * (t) - \vec{X}(t)| \quad (2)$$

$$\vec{X}(t+1) = \vec{X}(t) - \vec{A} \times \vec{H} \quad (3)$$

$\vec{A}$and $\vec{C}$imply vector coefficients obtained using equations (4) and (5), tsignifiescurrent iterations, phrase $\vec{X}$providespositional vectors and $\vec{X}*$ stands for arbitrary answers began arbitrarily.

$$\vec{A} = 2\vec{a} \times \vec{r} - \vec{a} \quad (4)$$

$$\vec{C} = 2x\vec{r} \quad (5)$$

Each time through an iteration, the elements of $\vec{a}$ are reduced linearly from 2 to 0, representing a random number from 0 to 1.

Bubble-net techniquesare used by humpback whales for hunting by circling their preys [19]. Whales enclose their prey, which includes fish, and then attempt to adjust their locations to choose the best course of action. Equations (6) and (7) reveal the WOA's primary math component.

$$X(t+1) = X^*(t) - A.|C.X^*(t) - X(t)| \text{ if } p < 0.5 \quad (6)$$

$$X(t+1)|C.X^*(t) - X(t)|.e^{bl} \cos(2\pi t) + X^*(t) \text{ if } p \geq 0.5 \quad (7)$$

here stands for time or iteration indices, X represents vectorsof whalepositions, and X* stands for obtained ideal solutions; A=2a. (r-a);C=2.r; r implies random vector with values ranging from 0 to 1; b stands for constant valuesbased on specific paths, determines form of logarithmic spirals where in this work, its value is 1; astands for coefficient vectors that reduce linearly from 2 to 0 on iterations.; The random number l is in the range of -1 to 1. When updating the positions of the whales, p, a random value from 0 to 1, is utilized to alternate among (6) and (7); in Eqs. (9) and (10), given that the probabilities are 50% and 50%, whales have an equal probability of randomly choosing either course through the optimization stage. The random value of vector A is [- 1, 1] when in the bubble-net stage, but it can be more or less than 1 within the searching stage. The method of searching is illustrated in Equation (8)

$$X(t+1) = X_{rand} - A\,.\,|C.X_{rand} - X(t)\ | \quad (8)$$

The WOA method is enforced to conduct a worldwide search by this random search system, which prioritizes the searching process when the value of |A| is bigger than one. The WOA searching method starts with the creation of random solutions. Subsequently, the method presented in Table 1 is employed to update these solutions repeatedly. Until a determined maximum number of iterations is achieved, the search will continue.

Exploitation phase: There are two steps in this stage. I surrounding, and ii) spirally updating the position. One way to define encircling behaviour is to decrease$\vec{a}$ linearly from 2 to 0 for each iteration. Revise the spiral's position: The whale's helical movement and relationship to the fish are determined by

$$\vec{X}(t+1) = \vec{D}xe^{b1}x\cos(2\pi l) + \vec{X}*(t) \quad (9)$$

here $\overrightarrow{D'} = |\vec{X}*(t)\_\vec{X}(t)|$ is the present distance among the whale and the fish, b an immovable element that symbolizes the whales' spiral motion, and b furthermore a random vector of $[-1, 1]$. Additionally, there is a possibility of choice: either the random vector of value is p [0,1] or the calculation of probability is provided by equation 6 for diving deep through circling and constructing a spiral.

Exploration phase: Whales look for fish that are rising to the top as part of a global fish exploration effort. $\vec{A}$, implies values of vectors in the interval [0,1] (0 denotes explorations and 1 denotes exploitations) serve as the basis for decisions to alternate between exploitations and explorations. And the equations (10) and (11) indicate the whale's new location.

$$\vec{H} = |\vec{C}x\vec{X}_{rand} - \vec{X}| \quad (10)$$

$$\vec{X}(t+1) = \vec{X}_{rand} - \vec{A}x\vec{H} \quad (11)$$

Where $\vec{X}_{rand}$provides the whale's new location, which is selected at random from among the other whales.

Adaptive social learning strategy

By building a neighbourhood membership relationship for every whale based on the social learning theory, it may enhance information sharing among groups, alter the behaviour of imitating the present best solution, and increase the method's capacity to deviate from the local optimal solution. Regarding the present-day populace

$$G(t) = \{x_1(t), x_2(t), \ldots., x_N(t)\} \quad (12)$$

here N implies size of the population. To create the sorted population, the fitness of each person is determined and ranked from small to large.

$$G_1(t) = \{x_{(1)}(t), x_{(2)}(t), \ldots., x_{(N)}(t)\} \quad (13)$$

and the social ranking of$x_{(i)}(t)$ is

$$I_{(i)}(t) = \frac{R_{(i)}(t)}{N} \, i = 1,2, \ldots, N \quad (14)$$

Where $R_{(i)}$ is random number and $I_{(i)}(t)$that a person's relationship with another person has developed.

Hence, the technique's exploitation phase primarily focuses on finding an ideal solution, while group cooperation allows for exploration. A unique whale search approach defined by adaptive social neighborhood approach is developed.

Algorithm 1: EWOA

Input: ASD dataset

Output: Optimized selections of features

Objective: Enhanced featureaccuracy features

```
Begin
establish the whale population's positions X (ASD dataset)
determine each whale's fitness (feature)
set a and r, compute A and C
set X* as the top hunter whale's position
set t= 1
while t ≤ max iterations do
for each hunting whale do
if p < 0.5
if |A| < 1
change locations of hunting whalesusing (6)
else if |A| ≥ 1
arbitrarily select other search agents
update current hunting whale positions with (7)
end if
else if p ≥ 0.5
update current hunting whale positions with(8)
end if
calculate local ideal solutionswith(12) & (13)
update exploitationswith(14)
end for
update X* if there is an improved solution
t = t +1
end while
output X*
end
```

The EWOA is employed by the suggested system to identify feature combinations that optimize classification accuracy while minimal chosen autistic features. Locating the ideal location in the search space that optimizes the specified fitness function needs intelligence because the feature space has each feature assigned a separate dimension, and the span of each dimension varies. Considering the autism training data, the EWOA's fitness function attempts to optimize its classification accuracy over the validation set.

3.3 Classification using ensemble M-AECNN-ANN algorithm

To categorize test data into yes (or) no classes, the ensemble-based M-AECNN-ANN method is suggested. In the present work, AE (Auto Encoder) - CNNs were employed to effectively diagnose citrus sickness. For instance, observe three hidden layers in two-dimensional CNNs utilizing AEs are shown in Fig. 2 [20]. In unsupervised learning, ML model F is trained to generate the same data as inputs x in a manner such as x F(x; w), here w represents ML system which are numerically optimized through training stages to lower errors E, producing w = argminw [E(x, F(x, w)].It is significant to observe that inputs or outputs R x are greater than the inner frameworks R, which are represented by green portions in Figure 2. If systems are taught to produce data that closely resemble inputs, then high-dimensional data was transformed into low-dimensional latent vectors. These connections can be described as,

$$\gamma = F_e(x), \qquad x \approx F_e(\gamma) \qquad (33)$$

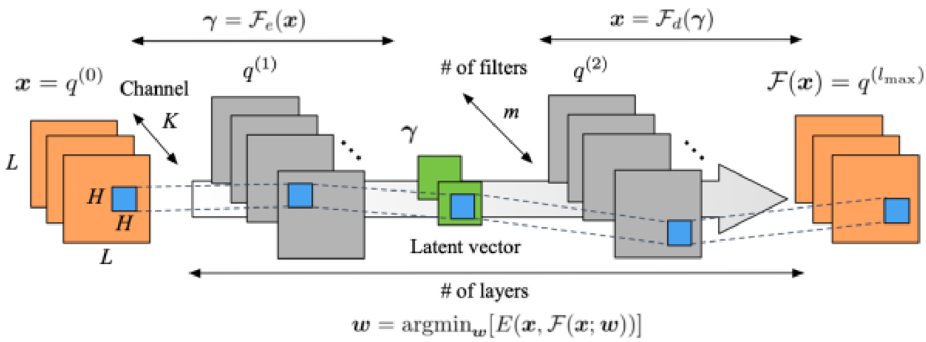
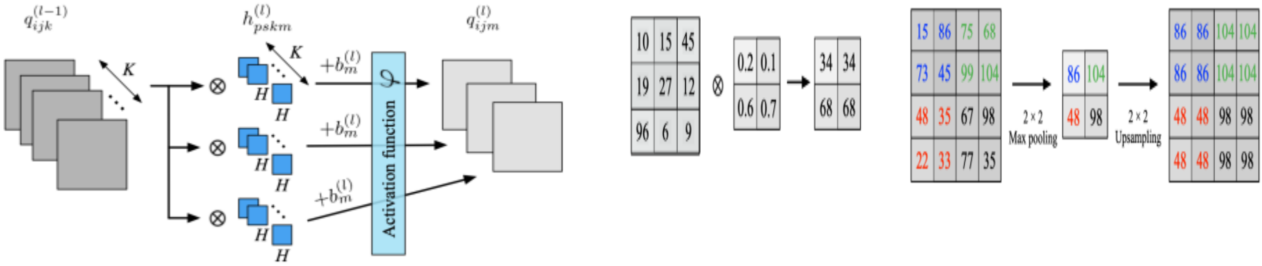


Fig 2 2D CNN based AE with three hidden layers.

Thus, as shown in Fig. 2, FdandFe represent the decoder and encoder elements of the AE, accordingly. As an aside, note that the extent to which may reduce the dimension depends critically on the kind of flow field working. In this study, the weights are updated iteratively during the training phase utilizing the Adam optimizer. The L2 norm error is the error function E. The idea of weight sharing is taught to CNNs. A CNN node at layer l, position (i, j), and filter index m produces the output q (l) by convoluting the filter h (l), which is represented in Fig. 5 examining the result of the upstream layer q (l1), as the blue H H squares.

$$q_{ijm}^{(l)} = \varphi(b_m^{(l)} + \sum_{k=0}^{K-1} \sum_{p=0}^{H-1} \sum_{s=0}^{H-1} h_{pskm}^{(l)} q_{i+p-C\, j+s-Ck}^{(l-1)}) \qquad (15)$$

here C = floor(H/2); K denotes filter counts in input/output layers' convolution layers and the corresponding flow variable counts for each point; b (l) m signifies bias, ϕ statesnonlinear activations functions or functions that increase monotonically. Hyperbolic tangent functions ϕ(s) = (e s − e −s) · (e s + e −s) −1 are used for activations. The hyperbolic tangent function was selected after a variety of nonlinear activation functions were examined. Additionally, it was shown that efficiency could be attained by substituting the Re LU for the sigmoid function. To produce the desired result, the filter coefficients h (l) pskm are optimized as weights through CNN training. The transmission of the enhanced filter coefficients within the same network layer illustrated in Fig 3. Be aware that CNNs function under the premise that there is no meaningful correlation between adjacent pixels in an image.



3(a) 3(b) 3(c)

Fig 3 Displays CNN's schematic singular convolution layerstructure (a) structures of singular layers of CNN; (b) convolution operations; (c) pooling and up sampling operations

Maximum pooling procedures As illustrated in the schemes of Fig. 3(c), encode and up sampling activities operate for decodes supplying dimension extensions and reductions required for structures of AE. By adopting the max pooling method, AEs can reduce the size of the input images while still being resilient over translation and rotation. Closest neighbour interpolation is employed in the decoder step of the up sampling process to transform low dimensional map data into high dimensional images. Despite having their roots in computer science, CNNs are employed for processing a variety of high-dimensional data, such as fluid dynamics. Reportedly, the employment of activations to bring up nonlinearities confers significant benefits to low dimensional maps over techniques based on linear principles. Nevertheless, interpretability is typically a problem with AE-based methods; especially, it can be challenging to comprehend physical significances of latent vectors produced by nonlinear filters. To comprehend the role of every latent vector in linear mode decomposition techniques, one must examine the energy containing ratio of AE-based options, as they do not share the similar idea as eigenvalues. This is because these modes according to AE are not perpendicular to one other. The Elman neural network that was suggested for handling this kind of energy-containing ratio.

3.3.1. EDO (Enhanced Dragonfly Optimization) Algorithm

DOA, population-based optimizers are popular and are based on the migration and hunting patterns of dragonflies provide the basis of the DOA method. When a swarm is static, all of its members fly over a limited region in close formation, searching for food sources. A notion states that dragonflies utilize dynamic swarming as part of their migrating strategy [21]. The swarm can depart because dragonflies like to fly in larger numbers. Figure 4 displays both static and moving dragonflies. Moreover, the creators of DA accomplish two fundamental ideas in other swarm-based frameworks: diversity, which is stimulated by static swarming actions, and expansion, which is recommended by dynamic swarming operations.

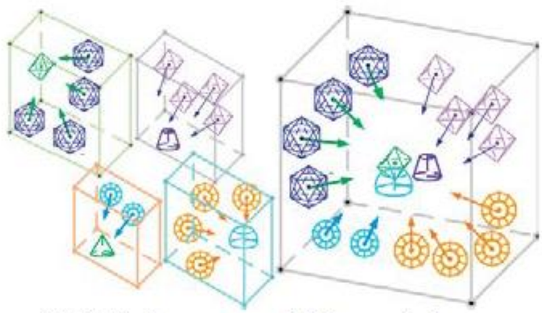


Fig 4 Dynamic and Static Dragonflies

DA, when X denotes position vectors, X j denotes X's neighbours at places j, N denotes neighbourhood sizes, and are distinguished by five behaviours:

Separations: This is a strategy dragonflies employ to differentiate among other agents. This method as stated in (16):

$$S_i = \sum_{j=1}^{N} X - X_i \qquad (16)$$

Alignments: Provide an example in which agents match the velocities of nearby dragonflies. This concept depends on (17), here Vj is the velocity vector of the j-th neighbour:

$$A_i = \frac{\sum_{j=1}^{N} V_i}{N} \qquad (17)$$

Cohesions: suggests that it is in members' nature to head for the closest bulk centre. This stage is described as (18):

$$C_i = \frac{\sum_{j=1}^{N} X_j}{N} - X \qquad (18)$$

Attractions: illustrates why the group's members are drawn to the food source. It serves to assess the food supply's and the i-th agent's potential for attraction.

$$F_i = F_{loc} - X \qquad (19)$$

Distractions: illustrates the dragonflies will frequently flee from a potential threat. The following describes how the opponent and the i-th dragonfly become preoccupied; here Eloc represents the opponent's location:

$$E_i = E_{loc} + X \qquad (20)$$

The fittest agents updatefood supplies and position vectors' fitnesses in DA. The enemies' placements and fitness levels are also approximations depending on the dragonfly with the lowest level of fitness. DA will be able to concentrate on more potential portions of the solution space and ignore less viable ones according to this insight. The positional and step vectors (X) are two criteria that are utilized to update positional vectors of dragonflies. The direction of motion for the dragonflies is indicated by the step vector in equation (21):

$X_{t+1} = (sS_i + aA_i + cC_i + fF_i + eE_i) + wX_t$ (21)

here s, w, a, c, f, and e calculate the represent weights of the locations and elements utilizing Eq. (22), and t iterations are:

$X_{t+1} = X_t + X_{t+1}$ (22)

A variety of local and global searches are frequently provided by the DA according to the separation, alignment, and cohesiveness optimizations. Two other factors that allow dragonflies to seize favourable circumstances while avoiding unfavourable ones are attraction and distraction.

The DA method is superior given these five swarming tendencies [22]. To determine the changing probability of the dragonfly position, DA uses a V-shaped transfer function. BDA does not necessitate dragonflies to select among 1 and 0, in contrast to other binary metaheuristics. Because of this, BDA had excellent exploration abilities, that enabled it discover the necessary search space.

a) Mutation Learning strategy (MLS) based Dragonfly Algorithm

The Modified Dragonfly (MD) updates its location by applying a mutation learning technique (MLS) that combines the ideas of personal finest and individual worst solutions (see Fig 5). In the traditional DA, the dragonflies direct the global worst solution—the enemy—and the global best solution—the food supply—for attraction and diversion. The personal best and personal worst dragonflies will be included in these activities to boost the likelihood of food seeking and opponent flight behaviours.

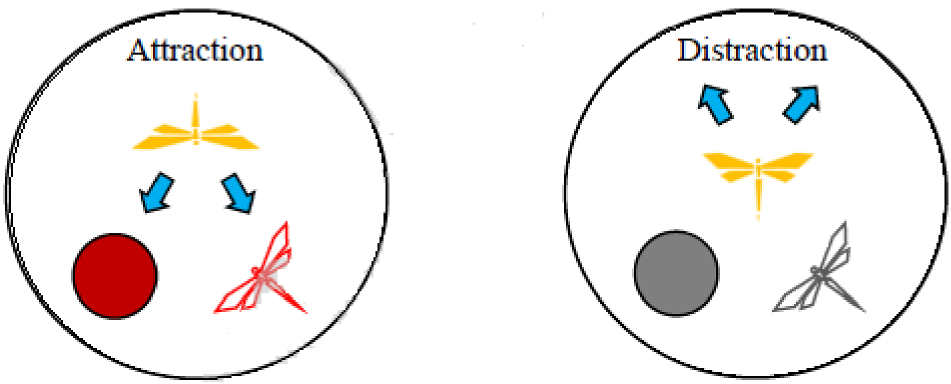


Fig 5 Attractions and distractions behaviours in MLS-DA.

In contrast to DA, MLS's distraction and attraction are determined utilizing the next methods.

$F_i = \frac{(Xpb_i - X_i) + (Xf - X_i)}{2}$ (23)

$E_i = \frac{(Xpw_i + X_i) + (Xe - X_i)}{2}$ (18)

where Xpb_idenote best dragonflies positions, Xpw_idenoteworst dragonflies, X denote dragonflys' places, Xf denote food sources, and Xe denote enemies. Additionally, during the search phase, the dragonflies could benefit from both their own best and the ideal solutions available globally because of the hyper learning method. The learning method's broad premise is shown in Fig. 6. The dragonflies attempts to replicate its best instances, both domestically and internationally, rather than moving in response to swarming.

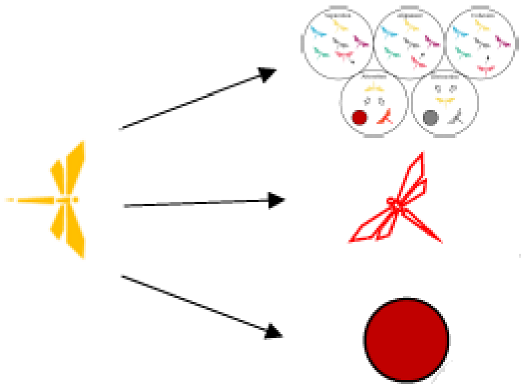


Fig 6General concept of the learning strategy.

The dragonfly locations of MLS-DA are updated utilizing:

$$X_i^d(t+1) = \begin{cases} \overline{X}_i^d & 0 \le r_1 < pl \\ Xpb_i^d(t) & pl \le r_1 < gl \\ Xf^d(t) & gl \le r_1 < 1 \end{cases} \quad (24)$$

$$\overline{X}_i^d = \begin{cases} 1 - X_i^d(t) r_2 < TF(\Delta X_i^d(t+1) \\ X_i^d(t) r_2 \ge TF(\Delta X_i^d(t+1) \end{cases} \quad (25)$$

where X stands for dragonfly locations, Xpb for optimal dragonfly roles, Xf for food sources, i for dragonfly orders, d for decision variable counts, t for the present iterations, and r1 and r2 for independent random values within 0 and 1. The fixed values of [0, 1] correspond to the global learning rate (gl) and the personal learning rate (pl).Equations (19-20) illustrate the importance of the pl and gl in the method of learning. When pl and gl are too low, the method becomes vulnerable to local optima trapping rather than looking for both individual and collective ideal solutions. If the values of pl and gl are too high, the position update mechanism will mimic DA. The choices made by pl and gl are therefore crucial. In Algorithm 2, the MLS-DOA pseudocode is displayed.

Algorithm 2: Mutation Learning Strategy based Dragonfly Algorithm for parameter tuning

Input: Enter X, S, A, and E.
Output: Optimized learning parameters
1) Place X and N dragonflies at arbitrary initial positions
2) Change the starting value of ΔX for the step vectors to zeros.
3) **while** (Not equal to Max iterations)
4) **for** *i*= 1 dragonfly counts, *N*
5) Compute fitnesses of (i-th) dragonflies.
6) Update X pbi, for fittest dragonflies.
7) Correct X pwi, n dividual worst dragonflies
8) **end for**
9) Update opponents (X e) and food sources (X f).
10) Update *s, α, c, f, e, and w*
11) **for** *i*= 1 to dragonfly counts, *N*
12) Use (10), (11) and (12) to calculate S, A, and C.
13) Use (13) and (14) to calculate F and E.
14) Use Mutation Learning Strategy to update step vectors in (17) and (18).
15) Refresh locationsof (*i*-th) dragonfliesusing (19 & 20)
16) **end for**
17) **end while**

ANN

ANNs are employed to gather information through learning. The input, output and hidden layer are the three stages of an ANN. The input layer gathers and analyzes the input images and outputs 'n' number of inputs. These procedures are executed utilizing certain weights. In neural networks, weights are the information that aids in problem solving. After some helpful hidden extraction, the data is transported from the input to output layer in the hidden layer. In this instance, AS and HCS are classified utilizing ANN. Utilizing ANN for training the adult, child, and adolescent dataset image, testing state images are classified, and ASD diagnosis is determined.

The backpropagation technique is employed by the network to train it [23]. The gradient descent technique is employed by the backpropagation process to find the error function's minimum. So, the set of weights that minimizes the error is a solution. The input data is first delivered to the network, where it is then transmitted until it reaches the output layer. Next, the error for each output neuron is computed by comparing the desired and actual outputs. The error for every neuron in every hidden layer is obtained by propagating this error backward across the network. A backpropagation method can change weights and biases employing these values.

The network's initial weights are assigned at random. The network propagates input xi when it arrives to it, resulting in an output ofo_i. Making each input's output $o_i$close to or similar to the intended output ti is the objective of the training procedure. The error function is minimized to achieve this:

$$E = \frac{1}{2}\sum_{i=1}^{n}(o_i - t_i)^2 \quad (26)$$

Initially, the output layer k error signal is computed:

$$\Delta_k = t_k - o_k \quad (27)$$

$$\delta_k = \Delta_k \alpha'_k \quad (28)$$

where$\alpha'_k$is activation functionderivative. This derivative for the output layer is equivalent to 1.

The output layer's weights are modified in accordance with:

$$\Delta w_{jk} = x_k \delta_k y \quad (29)$$

where $x_k$is the input from a neuron in the preceding layer (i.e. the corresponding neuron's output in the hidden layer), y is the learning rate.

To improve the speed of the algorithm, can written as

$$\Delta w_{jk}{}^{n} = x_k \delta_k y + \Delta w_{jk}{}^{n-1}\phi \quad (30)$$

whereϕ isa momentum factor. By remembering the prior modifications, the momentum component quickens the learning process and permits the method to operate in bigger steps.

The proposed ensemble based M-AECNN and ANN learning method achieves higher accuracy, precision, recall and f-measure metrics

4. Experimental result

ASD dataset is taken from the following link.https://www.kaggle.com/fabdelja/autism-screening-for+toddlers?select=Toddler+Autism+dataset+July+2018 .csv). Four datasets—ADULT, CHILD, ADOLESCENT, and TODDLER—are considered and one dataset is employed in testing. The classifier's efficiency is evaluated utilizing Sequential Minimal Optimisation with SVM (SMO-SVM), Conditional Mutual Information Maximisation (CMIM), Adaptive Grey Wolf Optimisation with SVM (AGWO-SVM), M-AECNN and EM-AECNN-ANN algorithms.

In the experiments, training establishes the optimized algorithm's variables, and validation sets the refined method’s and training hyperparameters. A range of criteria often utilized in binary categorization were employed to assess whether the different approaches predicted ASD data. Prior to calculating the various performance metrics, ascertain the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) rates. The first performance metric was precision, which is the percentage of relevant events restored out of all events improved. The second performance metric is recall, defined as the percentage of relevant events that are retrieved. Although accuracy and recall are incompatible metrics, they used to assess a prediction method's efficacy. Consequently, the F-measure, a single metric, can be produced by combining these two metrics and giving them equal weight. The final outcome metric was accuracy, which is the proportion of correctly predicted events to all projected incidences.

The precision ratio is the proportion of all correctly detected positive results to all expected positive data.

$$\mathrm{Precision} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP}) \qquad (31)$$

A recall is defined as the proportion of correctly identified positive findings to all data.

$$\mathrm{Recall} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}) \qquad (32)$$

True positives and false negatives are included while F-measures are weighted averages of precisions and recalls.

$$\mathrm{F} - \mathrm{measure} = 2 * (\mathrm{Recall} * \mathrm{Precision})/(\mathrm{Recall} + \mathrm{Precision}) \ (33)$$

Accuracy is calculated utilizing both positive and negative values.

$$\mathrm{Accuracy} = (\mathrm{TP} + \mathrm{FP})/(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}) \quad (34)$$

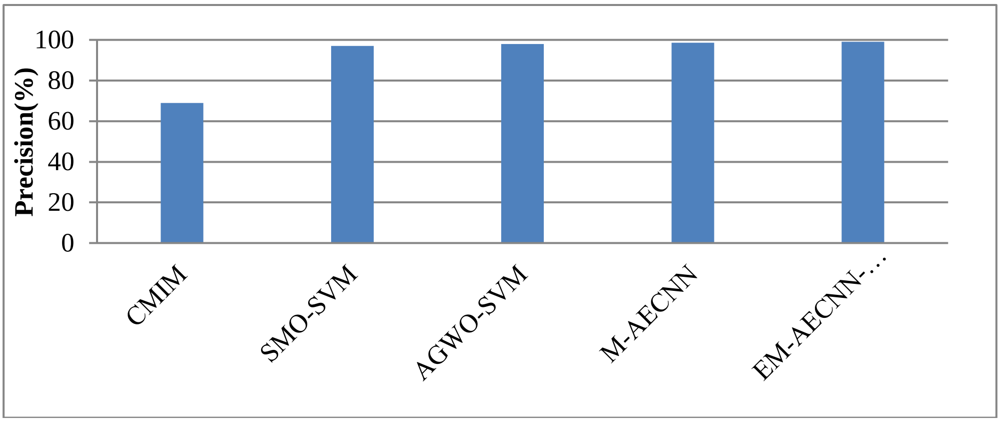


Fig7 Comparative precision values obtained for suggested and existing methods in ASD classification

Figure 7 shows comparative values obtained for current and suggested approaches (x-axis) with their precision values (y-axis) on the autism dataset. The suggested EM-AECNN-ANN method offers more precision than the current approaches, including CMIM, SMO-SVM, AGWO-SVM, and M-AECNN methods, which provide lower accuracy. So, the outcome indicates that by choosing features optimally, the suggested EM-AECNN-ANN method increases the ASD classification accuracy.

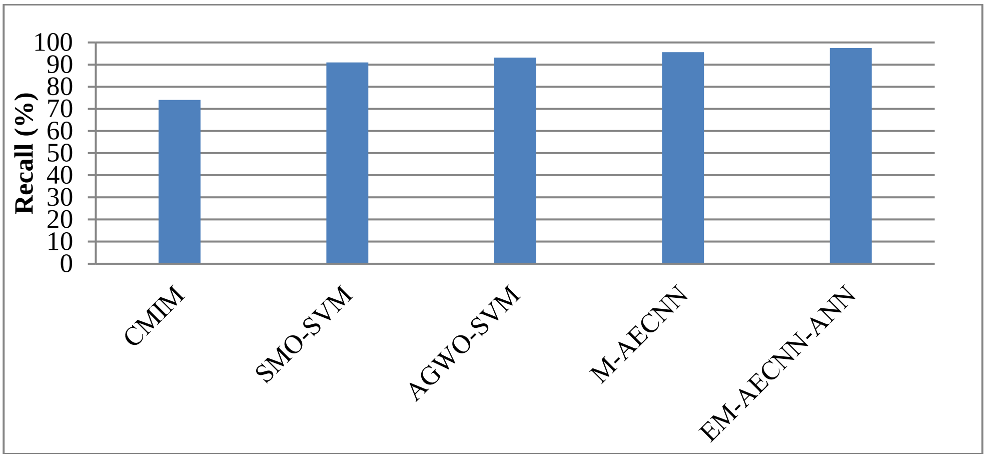


Fig 8 Comparative recall values obtained for suggested and existing methods in ASD classification

An examination of the suggested and present methods for identifying ASD data is shown in Fig 8. The primary goal of the data utilized to diagnose patients with signs of ASD by utilizing a range of characteristics that often have an effect on the diagnosis. Because of this, it is thought that the prediction model's categorization issue which is solved regardless of whether the individual has ASD. To validate the assessment results and the simulation's accuracy, the weak elements in the databases must be eliminated through the feature selection technique before deploying these ML algorithms. Thus, the outcome indicates that by choosing features optimally, the suggested EM-AECNN-ANN technique increases the ASD classification accuracy.

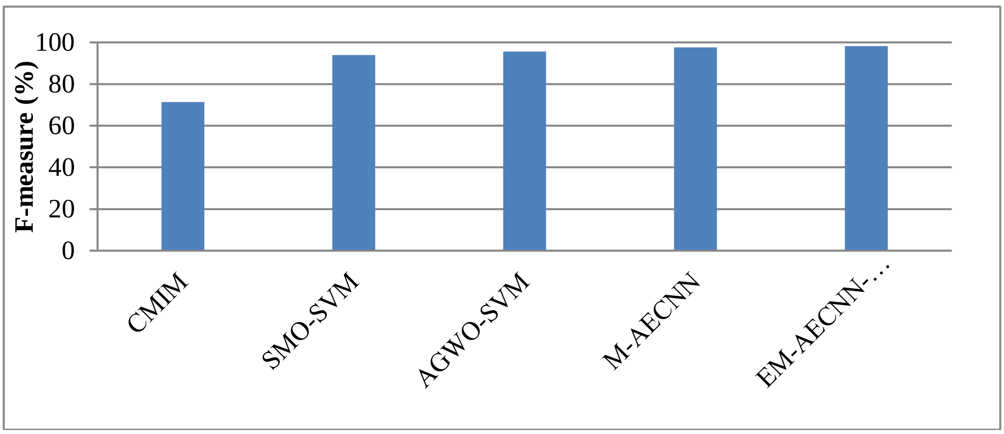


Fig 9 Comparative F-measure values obtained for suggested and existing methods in ASD classification

The F-measure analysis among the suggested and existing methods for classifying ASD data is shown in Fig. 9. Utilizing feature-selection and categorization methods, the toddler database variable derived from the ASD test showed the highest degree of association with the appropriate class. For the provided autism dataset, the suggested EM-AECNN-ANN method yields a greater F-measure than other methods, such as CMIM, SMO-SVM, AGWO-SVM, and M-AECNN algorithms. Thus, the outcome indicates that by choosing features optimally, the suggested EM-AECNN-ANN method increases the ASD classification accuracy.

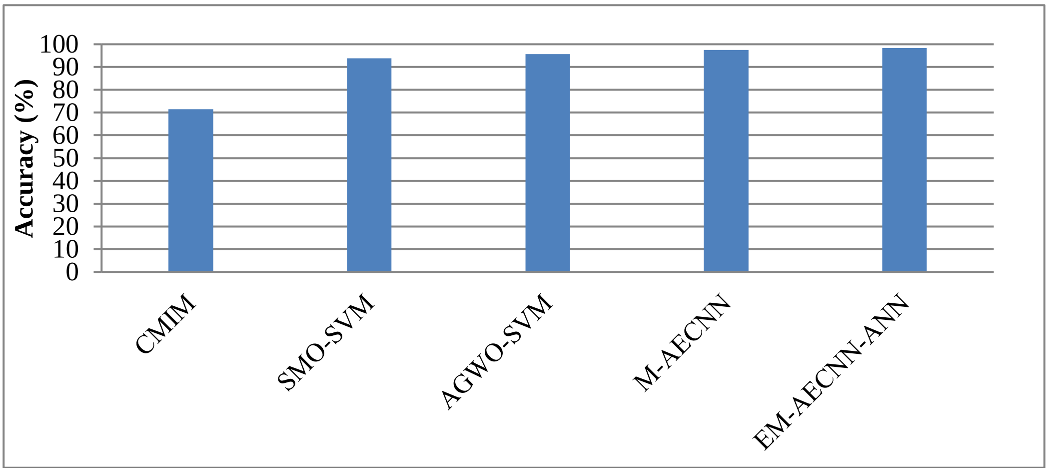


Fig 10 Comparative accuracies obtained for suggested and existing methods in ASD classification

An accuracy evaluation amongst the suggested and recent methods for classifying ASD data illustrated in Fig. 10. Based on the simulation findings, the EM-AECNN-ANN modelling outperforms the current CMIM, SMO-SVM, AGWO-SVM, and M-AECNN approaches with a high accuracy of 98.3%. The goal of this study with EWOA is to identify the significant and pertinent features for the provided autism dataset. So, the outcome indicates that by choosing features optimally, the suggested EM-AECNN-ANN method increases the ASD classification accuracy.

5. Conclusion

ASD is a complex, enduring neurodevelopmental disease that first emerges in children. It is linked to abnormal movements and abnormalities in gait. ASD can be accurately and automatically recognized, which helps with diagnosis and clinical decision-making. It also improves targeted therapy. To increase the accuracy of ASD categorization, this study presented the EM-AECNN-ANN method. The goal of the pre-processing is to eliminate unneeded features from the provided autism dataset. Next, the EWOA is employed for selecting features, and fitness values are utilized to calculate the more pertinent and informative autistic features. By using IWOA algorithm optimal local and global features are selected for the given autism dataset. The ASD classification is done by EM-AECNN-ANN technique and utilized to generate autism outcomes are yes/no classes that are more accurate. Following the completion of the training and testing phases, the error will decrease until it reaches a constant value. The test outcome shows that equated with recent systems, the EM-AECNN-ANN method performs superior based on its higher accuracy, precision, recall, and f-measure. In future work, feature extraction algorithm suggested to enhance the ASD classification efficiency.

References

1. Matson, Johnny L., Julie A. Hess, and Sara Mahan. "Moderating effects of challenging behaviors and communication deficits on social skills in children diagnosed with an autism spectrum disorder." Research in Autism Spectrum Disorders 7.1 (2013): 23-28.
2. Landa, Rebecca J., et al. "Intervention targeting development of socially synchronous engagement in toddlers with autism spectrum disorder: A randomized controlled trial." Journal of Child Psychology and Psychiatry 52.1 (2011): 13-21.
3. Christensen, Deborah L., et al. "Prevalence and characteristics of autism spectrum disorder among children aged 4 years—Early Autism and Developmental Disabilities Monitoring Network, seven sites, United States, 2010, 2012, and 2014." MMWR Surveillance Summaries 68.2 (2019): 1
4. Zwaigenbaum, Lonnie, and Melanie Penner. "Autism spectrum disorder: advances in diagnosis and evaluation." Bmj 361 (2018).
5. Mukherjee, Sharmila Banerjee. "Autism spectrum disorders—diagnosis and management." The Indian Journal of Pediatrics 84 (2017): 307-314.
6. Krishnan, Arjun, et al. "Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder." Nature neuroscience 19.11 (2016): 1454-1462
7. Cogill S, Wang L (2016) Support vector machine model of developmental brain gene expression data for prioritization of Autism risk gene candidates. Bioinformatics 32(23):3611–3618
8. Kong, Yazhou, et al. "Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier." Neuro computing 324 (2019): 63-68
9. Yuan, Jianbo, et al. "Autism spectrum disorder detection from semi-structured and unstructured medical data." EURASIP Journal on Bioinformatics and Systems Biology 2017.1 (2016): 3
10. Sharawi, Marwa, Hossam M. Zawbaa, and Eid Emary. "Feature selection approach based on whale optimization algorithm." 2017 Ninth international conference on advanced computational intelligence (ICACI). IEEE, 2017
11. Kosmicki, J. A., Sochat, V., Duda, M., & Wall, D. P. (2015). Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. Translational psychiatry, 5(2), e514-e514.
12. Alzubi, R., Ramzan, N., &Alzoubi, H. (2017, August). Hybrid feature selection method for autism spectrum disorder SNPs. In 2017 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB) (pp. 1-7). IEEE.
13. Hossain, M. D., Kabir, M. A., Anwar, A., & Islam, M. Z. (2021). Detecting autism spectrum disorder using machine learning techniques. Health Information Science and Systems, 9(1), 1-13
14. Kong, Yazhou, et al. "Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier." Neuro computing 324 (2019): 63-68.
15. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.
16. Han, H., Wang, W. Y., & Mao, B. H. (2005, August). Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In International conference on intelligent computing (pp. 878-887). Springer, Berlin, Heidelberg
17. Jin, Qibing, Zhonghua Xu, and Wu Cai. "An Improved Whale Optimization Algorithm with Random Evolution and Special Reinforcement Dual-Operation Strategy Collaboration." Symmetry 13.2 (2021): 238.

18. Fan, Q.; Chen, Z.; Zhang, W.; Fang, X. ESSAWOA: Enhanced Whale Optimization Algorithm integrated with Salp Swarm Algorithm for global optimization. Eng. Comput. 2020, 1–18
19. Sayed, G.I.; Darwish, A.; Hassanien, A.E. A New Chaotic Whale Optimization Algorithm for Features Selection. J. Classif. 2018, 35, 300–344
20. Masci, J., Meier, U., Cireşan, D., & Schmidhuber, J. (2011, June). Stacked convolutional auto-encoders for hierarchical feature extraction. In International conference on artificial neural networks (pp. 52-59). Springer, Berlin, Heidelberg.
21. Sairam, K., Naren, J., Vithya, G., & Srivathsan, S. (2019). Computer aided system for autism spectrum disorder using deep learning methods. Int. J. Psychosoc. Rehabil, 23(01).
22. Mafarja, M., Heidari, A. A., Faris, H., Mirjalili, S., &Aljarah, I. (2020). Dragonfly algorithm: theory, literature review, and application in feature selection. Nature-inspired optimizers, 47-67.
23. Mohamed, Nashaat El-Khamisy, and Ahmed Shawky Morsi El-Bhrawy. "Artificial Neural Networks in Data Mining." IOSR Journal of Computer Engineering 18 (2016): 55-59