

## Modeling of Explainable Artificial Intelligence for Biomedical Mental Disorder Diagnosis

Anwer Mustafa Hilal<sup>1</sup>, Imène ISSAOUI<sup>2</sup>, Marwa Obayya<sup>3</sup>, Fahd N. Al-Wesabi<sup>4</sup>, Nadhem NEMRI<sup>5</sup>, Manar Ahmed Hamza<sup>1,\*</sup>, Mesfer Al Duhayyim<sup>6</sup> and Abu Sarwar Zamani<sup>1</sup>

<sup>1</sup>Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam Bin Abdulaziz University, Alkharj, 16278, Saudi Arabia

<sup>2</sup>Department of Natural and Applied Sciences, Community College, Qassim University, Buraydah, Saudi Arabia

<sup>3</sup>Department of Biomedical Engineering, College of Engineering, Princess Nourah Bint Abdulrahman University, Riyadh, 11564, Saudi Arabia

<sup>4</sup>Department of Computer Science, King Khalid University, Muhayel Aseer, Saudi Arabia & Faculty of Computer and IT, Sana'a University, Sana'a, 31220, Yemen

<sup>5</sup>Department of Information Systems, King Khalid University, Muhayel Aseer, 62529, Saudi Arabia

<sup>6</sup>Department of Natural and Applied Sciences, College of Community-Aflaj, Prince Sattam Bin Abdulaziz University, Al-Kharj, 16278, Saudi Arabia

\*Corresponding Author: Manar Ahmed Hamza. Email: ma.hamza@psau.edu.sa

Received: 14 August 2021; Accepted: 05 October 2021

**Abstract:** The abundant existence of both structured and unstructured data and rapid advancement of statistical models stressed the importance of introducing Explainable Artificial Intelligence (XAI), a process that explains how prediction is done in AI models. Biomedical mental disorder, i.e., Autism Spectral Disorder (ASD) needs to be identified and classified at early stage itself in order to reduce health crisis. With this background, the current paper presents XAI-based ASD diagnosis (XAI-ASD) model to detect and classify ASD precisely. The proposed XAI-ASD technique involves the design of Bacterial Foraging Optimization (BFO)-based Feature Selection (FS) technique. In addition, Whale Optimization Algorithm (WOA) with Deep Belief Network (DBN) model is also applied for ASD classification process in which the hyperparameters of DBN model are optimally tuned with the help of WOA. In order to ensure a better ASD diagnostic outcome, a series of simulation process was conducted on ASD dataset.

**Keywords:** Explainable artificial intelligence; autism spectral disorder; feature selection; data classification; machine learning; metaheuristics

### 1 Introduction

The application of Artificial Intelligence (AI) methods is currently unstoppable and pervasive due to its incredible characteristics and high prevalence of adoption. But, it carries certain sorts of problems, opportunities, and risks which should be tackled in order to have an uncompromised and



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

efficient development [1]. Explainable Artificial Intelligence (XAI) is the answer to overcome this challenge since it brings machines nearby human beings. From a study point of view, the discussion on XAI started several years ago while the idea developed with renewed vigor at the end of 2019 only. Then, Google declared its “AI-first” approach in 2017 and lately published a novel XAI toolset for developers [2]. Currently, different Deep Learning (DL) and Machine Learning (ML) applications exist whereas one could not understand the logic behind these mechanisms or how the decisions are made. So, ‘BlackBox’ was proposed to understand the effects and is based on ML methods. Its features are deliberated as the main challenge in the application of AI technique; Further, it also helps in decision making process which is not apparent and remain frequently unintelligible to the developers/experts’ perspective [3].

XAI system is capable of explaining the logic behind decisions, describe the strengths and weaknesses of decision making, and offer insights about the upcoming behaviours. AI applications are otherwise observed as autonomous driving systems and are utilized in financial, healthcare and military/legal sectors. In this case, there is a need to have reasoning behind the artificial partners in order to trust the decision and the data attained. The most common AI framework is now provided by DL approaches, whereas a Neural Network (NN) of tens or hundreds of layers of ‘neurons’, or ‘fundamental processing unit’, is employed. The complication of DL frameworks makes them act like ‘black boxes’ due to which it becomes nearly impossible to find the appropriate method for systems to provide certain answers [4].

The application of AI in healthcare, especially in diagnostic imaging, is rapidly increasing [5]. However, the contribution of DL frameworks turns the attention towards “accountability” of the process. Since the DL solutions are being extensively used, these problems would gradually become a main attention. In healthcare industry, responsibility or accountability, undertaken by the professionals, plays a major role since it is directly associated with the patients’ health. Each clinical decision should be taken based on available scientific evidence backed with professional experience [6]. The results produced by AI processing units is expected to contribute to medical decision making whereas the ‘black box’ architecture is barely compatible with healthcare sector. Moreover, this software application must be authorized whereas the importance of this process must be understood, despite the fact that it is an unexplained approach. Physicians are satisfied towards the application of approaches that use neural networks in challenging or complex diagnoses. However, they should also comprehend the way, how they conclude and validate the reports.

Autism Spectrum Disorder (ASD) is a mental illness development disorder that limits specific social behaviour and communication from normal evolution [7]. Indeed, the causes of ASD have been associated with neurological and genetic factors. In spite of its genetic association, ASD is often detected using behavioral indicators like imaginative ability, repetitive behaviours, social interaction, and interactions with others [8]. Children with ASD face severe earlier development problems than other infantile groups. These behavioral difficulties differ and comprise of challenges in responding to sensory information (tasting, hearing, smelling, and so on), difficulties in communicating, and impact the earlier learning process. Further, it also creates a challenging time in communicating with others and such children lag behind in language acquisition. A research study establishes that 33% of the children, with challenges except ASD, have few ASD symptoms, when unable to meet the complete classification conditions.

There have been several non-clinical and clinical diagnostic approaches available for ASD. Two established clinical diagnoses methods are Autism Diagnostic Interview (ADI) and Autism Diagnostic Observation Schedule-Revised (ADOS-R). Additionally, some of the approaches that

are used in diagnostics are parent-based nonclinical or self-administered techniques namely, Social Communication Questionnaire (SCQ) and Autism Quotient Trait (AQ). It is important to note that the present mainstream ASD diagnostic tools consume significant time to conduct a comprehensive diagnoses. In order to develop the diagnostic procedure of ASD, scientists recently have begun to adopt ML approaches [9]. The key objectives of this research on ML approaches to diagnose ASD are to reduce the diagnoses time to provide quick access to healthcare service, to reduce the dimensionality of input dataset, and to improve diagnostic accuracy in such a way that the maximum ranked feature of ASD is detected. Machine Learning is a field of research that incorporates search methods, AI, and mathematics to derive precise prediction models from the dataset.

The current research article presents an XAI-based ASD diagnosis (XAI-ASD) model to detect and classify ASD in a precise manner. The proposed XAI-ASD technique involves the designing of Bacterial Foraging Optimization (BFO) for Feature Selection (FS) approach. In addition, Whale Optimization Algorithm (WOA) with Deep Belief Network (DBN) model is applied for ASD classification process, in which the hyperparameters of DBN model are optimally tuned with the help of WOA. In order to validate the proposed method in terms of achieving better ASD diagnostic outcomes, a series of experimental analysis was conducted. The experimental results highlight that the results achieved by the proposed method were better and the proposed XAI-ASD technique is superior to other state-of-the-art techniques under different measures. In short, the contributions of the research paper are listed as follows.

- A new XAI-ASD model is presented for detection and classification of ASD
- A new BFO-based FS technique is designed to choose an effective subset of features for ASD detection.
- A new WOA-DBN technique is derived for detection and classification of ASD.
- The performance of the proposed XAI-ASD technique was validated on benchmark dataset and the outcomes were examined under different evaluation parameters.

The research paper has the following sections. Section 2 provides a comprehensive review of existing ASD diagnosis techniques. Section 3 elaborates the proposed XAI-ASD technique and Section 4 validates the results attained from XAI-ASD technique. Finally, Section 5 draws the conclusion.

## 2 Literature Survey

In Pawar et al. [10], XAI is deliberated as a method that could be employed in both diagnosis and analysis of healthcare data through AI system. The aim of the presented method was to achieve result tracing, responsibility, model improvement, and transparency in healthcare domain. Lauritsen et al. [11] presented a new xAI-EWS method for early recognition of acute critical illnesses. The proposed xAI-EWS method potentially executes medical translation by attaining an estimation with EHR data and its description. Eslami et al. [12] summarized the current developments in ML model for the diagnoses of ASD and ADHD. The researchers described and outlined the ML approach, particularly DL methods that are relevant to the area of research in these domains, drawbacks of the access methods, and upcoming directions for the domain. Further, the researchers also predicted how ADHD, ASD, and other mental disorders would be diagnosed, quantified, and accomplished in future through imaging methods such as MRI and ML techniques.

Eslami et al. [13] presented a DL algorithm i.e., ASD DiagNet which exhibited relatively high accuracy in the classification of ASD from a typical neuroscan. Initially, the researchers incorporated classical ML and DL methods which permits the isolation of ASD biomarkers from MRI datasets.

This approach named Auto-ASD-Network exploits the integration of DL and SVM in the classification of ASD scans from neurotypical scans. These interpretable methods assist in explaining the decisions made by DL technique which in turn results in knowledge acquisition for neuroscientists and clear analyses for physicians. Abbas et al. [14] proposed a multi modular, ML-based evaluation of autism including two corresponding models for unified results of diagnostics grade.

In Liang et al. [15], the main development is done based on the temporal coherencies among nearby frames as free supervision and setting up a global discriminative margin. Based on a wide range of assessment of the extracted features, the efficacy of these features has been established. Initially, the extracted feature is categorized by K-means approach to show how self-stimulation behavior is sorted out in a completely unsupervised manner. Next, a conditional entropy approach is employed to evaluate the efficacy of these features. Then, advanced result is attained by integrating the unsupervised TCDN approach with optimized supervised learning models (like k-NN, Discriminant, SVM).

Lombardi et al. [16] presented an explicable DL architecture to predict the age of a healthier cohort of subjects from ABIDE I database. This was done based on the morphological features that were generated from MRI scan images. Also, the researchers embedded two local XAI techniques such as SHAP and LIME to explain the results of DL algorithms, establish the participation of all brain morphological descriptors to the concluding predictable age of all subjects and explore the consistency of two techniques. Zhang et al. [17] proposed an approach in which DNN is exploited to analyze the subject's MRI and calculate the efficiency for initial screening of ASD. Early analyses of patients with fMRI were also related to sMRI. The experimental result shows that fMRI is highly sensitive compared to sMRI. Also, the researchers described the classification performance of fMRI.

### 3 The Proposed Model

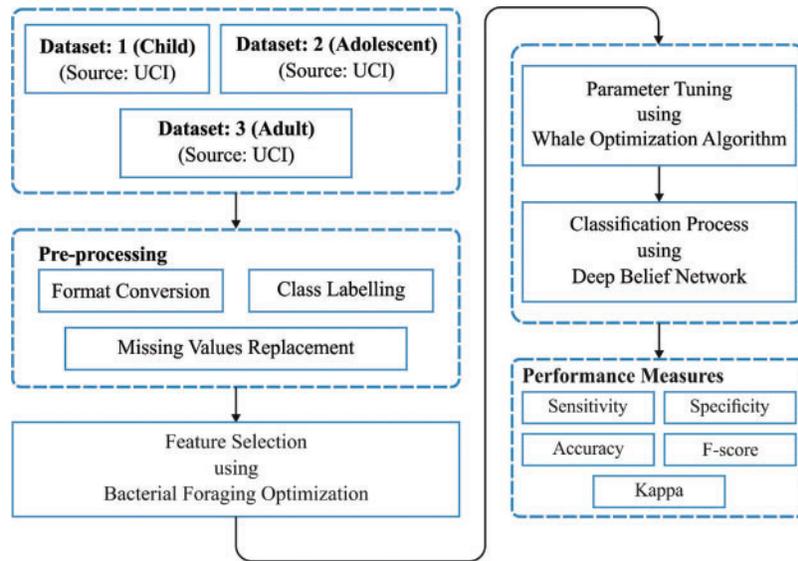
In this study, a new XAI-ASD technique is designed and validated to detect and classify different stages of ASD. The presented XAI-ASD technique encompasses different processes namely, pre-processing, WOA-based FS, DBN-based classification, and WOA-based parameter tuning. Fig. 1 illustrates the overall working process of XAI-ASD model. The complete functional processes of these modules are offered in the following sections.

#### 3.1 Pre-Processing

Initially, the patient data is pre-processed through three stages such as format conversion, missing value replacement, and class labeling. In the beginning, the input data in .arff format is modified into a companionable .csv format. Also, the missing values in the dataset are employed by median process. Eventually, class labeling model is incorporated to map the class labels of data against ASD.

#### 3.2 Algorithmic Design of BFO-FS Technique

After collecting and pre-processing the patient data to remove the undesirable data, FS procedure is executed in which BFO technique is utilized. A standard BFO approach has two important facts as given herewith.



**Figure 1:** Overall process of XAI-ASD method

- 1) Solution space initiation: the solution spatial dimension,  $D$  denotes the range based on which the mapping function  $f(x)$  is executed.
- 2) Bacterial initiation: the amount of bacteria is denoted by  $S$ . The place of  $i^{th}$  bacterium in an optimized space is specified by  $P_i(j, k, l)$  which is equivalent to optimal variable of the solution, viz.,  $P_i(j, k, l) = [m_1, m_2, \dots, m_D]$ .

Therefore, the fitness of  $i^{th}$  bacterium is identified as  $J_i(j, k, l)$ , in the optimized place and is determined, utilizing the function of bacterium place as follows.

$$J_i(j, k, l) = f(P_i(j, k, l)) = f_{i,j,k,l}(m_1, m_2, \dots, m_D). \tag{1}$$

In (1), the lesser values of function refer to maximum fitness.  $i$  refers to  $i^{th}$  bacterium, but  $j, k$ , and  $l$  transmit to central procedure of BFO technique: dispersal, reproduction, chemotaxis, and elimination.

**Chemotaxis**

It has massive amounts of swimming and flipping actions [18]. In  $j^{th}$  chemotaxis method, the motion of  $i^{th}$  bacterium is provided as follows.

$$P_i(j + 1, k, l) = P_i(j, k, l) + \frac{\Delta(i)}{\sqrt{\Delta^T(i)\Delta(i)}} C(i)n, \tag{2}$$

But the swimming step length of  $i^{th}$  bacterium is divided into a single swimming step size  $C(i)$ . The count of swimming  $n$  and  $(i)$  vector way of  $i^{th}$  bacterium from  $p$  dimension are provided in the optimized space. All the elements of  $(i)$  are numerical values in the range of  $-1$  and  $1$ , while a random value is fixed in the beginning within this range. When  $i^{th}$  bacterium detects the maximum fitness place which existed in the favorable atmosphere from  $j^{th}$  chemotaxis, it makes sure that it gets affected in the same way according to this time. Somewhat  $(i)$  a new random manner is elected.

### Swarming

The bacteria is considered for both repulsion and attraction. The numeric relation is given herewith.

$$J_{cc}(P_i) = \sum_{i=1}^s \left[ -d_{att} \exp \left( -\omega_{att} \sum_{m=1}^p (P_{i,m} - \bar{P}_m)^2 \right) \right] + \sum_{i=1}^s \left[ h_{rep} \exp \left( -\omega_{rep} \sum_{m=1}^p (P_{i,m} - \bar{P}_m) \right) \right], \quad (3)$$

where  $d_{att}$  represents the depth of the attracted material, when it is released by  $i^{th}$  bacterium, but  $\omega_{att}$  signifies the width of the same attracted material. However, as it is not possible for two bacteria to co-exist in the same place, the repulsion is modified into  $h_{rep}$  and  $\omega_{rep}$ . Next, in swarming method, the fitness of  $i^{th}$  bacterium is provided as follows.

$$J_i(j+1, k, l) = J_i(j, k, l) + J_{cc}(P_i(j, k, l)) \quad (4)$$

### Reproduction

The bacteria repeats the reproduction process, when it is exposed to a better atmosphere; otherwise, it tend to die. So, based on chemotaxis and swarming methods, the fitness of whole bacteria is fixed and computed. The fitness of  $i^{th}$  bacterium is defined as follows.

$$J_{i,health} = \sum_{j=1}^{N_c} J_i(j, k, l). \quad (5)$$

Half of the bacteria are at better state while,  $S_r = \left(\frac{s}{2}\right)$  is selected for survival and the other half tend to die. The survived bacteria then reproduces two colonies located at the same area and preserves the whole number of bacteria  $S$  set.

### Elimination and Dispersal

Then all the bacterium reproduce at the probabilities of  $P_{ed}$ . But the entire number of bacteria remains the same. If bacterium is distant, it can be randomly distributed to new places.

$$r = \text{random} [0, 1];$$

$$P_i(j, k, l) = \begin{cases} P_i(j, k, l) & r > P_{ed}, \\ m'_1, m'_2, \dots, m'_p & r < P_{ed}. \end{cases} \quad (6)$$

As demonstrated in (6), the elimination occurs but  $r_i < P_{ed}$ . The original place of  $i^{th}$  bacterium  $P_i$  is replaced utilizing the new one  $P'_i = (m'_1, m'_2, \dots, m'_p)$ . Therefore, a better variable  $m$  is upgraded for arbitrary variable,  $m'$  which is solved in the optimized space.

### Process Involved in BBFO-FS Algorithm

In FS issue, all the solutions are controlled towards the binary values of [0, 1]. In order to apply BFO technique in FS, BBFO technique is developed as a binary version. In BBFO technique, the solution is defined as 1D vector, but the length of vector depends on the amount of features in original dataset. All the cells in vector are valued at zero/one. The value 'one' represents the outcome feature that has been elected; otherwise, the value is defined as 0. Eq. (7) is utilized to map the continuous

values for binary ones.

$$Z_{mn} = \begin{cases} 1 & \text{if } X_{mn} > 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

whereas  $Z_{mn}$  stands for the creation of different solution vector,  $X$  and  $X_{mn}$  refers to the continuous place of search agent,  $m$  at dimension,  $n$ .

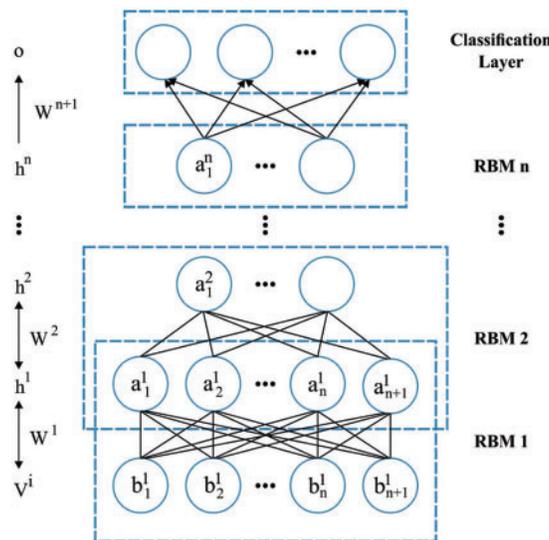
FS is modelled as a multi-objective optimization issue but two differing objectives are achieved i.e., less count of FS and superior classifier accuracy. Here, better result is the solution in terms of less number of FS and superior classifier accuracy. The classifier accuracy of KNN is employed as Fitness Function (FF) in order to assess the performance of entire search agents. In order to balance amongst the number of FS in all solutions (lesser) and the classifier accuracy (higher), the FF in Eq. (8) is utilized in ISSA and whole technique to assess the search agents.

$$Fitness = \rho Err(D) + \varphi \frac{|F|}{|T|} \tag{8}$$

While  $Err(D)$  represents the classifier error rate of the known subset,  $\rho$  and  $\varphi$  refer to constants that control the classifier accuracy and decrease feature,  $|F|$  implies the size of known feature subset and  $|T|$  indicates the entire count of features.

### 3.3 Design of WOA-DBN Technique

At last, the extracted features are then fed to DBN method for the classification of ASD. The framework of DBN method is shown in Fig. 2.



**Figure 2:** Structure of DBN

It includes a classification model and stacked RBM in deep network. The training procedure of DBN includes supervised as well as unsupervised learning stages. The stacked RBM region is located in classification layer for the creation of DBN method. DBN training procedure starts at the unsupervised training of RBM, with the help of greedy algorithm and the result of previous RBM

becomes the input of succeeding RBM. Finally, the supervised training procedure is performed in the classifier layer of DBN method.

In general, the main difference between DBN and MultiLayer Perceptron (MLP) layers is that the DBN layer is capable of performing the learning procedure through layerwise unsupervised training procedures. However, the MLP approach fails to perform feature learning tasks and is widely used for classification procedure. The resemblance between DBN and the MLP is that both are FC networks. Further, BP method is employed for supervised learning and is executed in the classifier layers of DBN. RBM is a likelihood-based graphical network, realized by the stochastic NN, in which two output states of the neuron are included. It depends upon BM approach whereas a FE method is considered based on energy model and optimization occurs through unsupervised training. Here, RBM method includes two sets of layers i.e., visible layer  $v = (v_1, v_2 \dots, v_n)$  which indicates the observation data, and hidden layer  $h = (h_1, h_2 \dots, h_m)$  which is determined as the FE layer. Primarily, the two topmost layers are denoted as the joint distribution of last hidden ( $h_i$ ) layer and output layer which are also called as the associated memory elements.

The learning process of the DBN models has two phases like supervised learning (fine-tuning) and unsupervised learning (pre-training) [19]. At first, the unsupervised learning is comprehended by contradictory deviations, to train the stacked RBM in a hierarchical manner. Then, the supervised learning is stimulated by BP method to fine-tune the early biases and weights of the whole network. The primary goal of the unsupervised training in DBN, is the optimization of RBM to extract the features from data. For a provided set of ( $v, h$ ), the energy function  $E(v, h|\theta)$  is determined as follows.

$$E(v, h|\theta) = - \sum_{i=1}^n a_i v_i - \sum_{j=1}^m b_j h_j - \sum_{i=1}^n \sum_{j=1}^m v_i w_{ij} h_j \quad (9)$$

Whereas  $\theta = \{W_{ij}, a_i, b_j\}$  represents the parameters in RBM,  $w$  represents the weight of layer connection, and  $a$  &  $b$  denote the bias of hidden and visible layers of the neuron, respectively. The joint likelihood distribution is determined as follows.

$$P(v, h|\theta) = \frac{e^{-E(v,h|\theta)}}{Z(\theta)}, \quad Z(\theta) = \sum_{v,h} e^{-E(v,h|\theta)} \quad (10)$$

During Gibbs sampling procedure, a likelihood distribution condition of the hidden neuron and visible neurons is represented herewith.

$$P(h_j = 1|v, \theta) = \text{sigmoid} \left( b_j + \sum_i^{v_i} w_{ij} \right) \quad (11)$$

$$P(v_i = 1|h, \theta) = \text{sigmoid} \left( a_i + \sum_j^{h_j} w_{ji} \right) \quad (12)$$

With Eq. (11), the likelihood  $h_j$  represents an active state. Since RBM has a symmetric feature for the determined hidden layer state  $h$ , the activation state likelihood of each neuron in the visible layer is shown in Eq. (12). This process is employed to obtain the corresponding weight  $w$  of RBM. The unsupervised learning of DBN trains the RBM in a hierarchical manner to obtain the early weight,  $W = \{w_1, w_2 \dots w_i\}$ . Supervised learning includes the fine-tuning of connections weight made from the unsupervised learning. BP method defines the gradient by labelling the training data set and altering the network variables, between the layers, to reduce the gradient. Finally, deep network framework is developed using the least predictive errors.

In order to achieve the optimum tuning of hyperparameter in DBN method, WOA is implemented. In the initial phase, an initialized procedure takes place. In the surrounding prey model, a humpback whale is noted around the place of prey. The whale starts surrounding the prey. To unclear the place in searching area, the current optimum result is regarded as the prey. If the optimum searching agent is determined, the other searching agent refreshes the condition in the way of optimum searching agent.

$$\vec{U} = |S \cdot \vec{K}^*(t) - \vec{K}(t)| \tag{13}$$

$$\vec{K}(t + 1) = \vec{K}^*(t_{best}) - Y \cdot \vec{U}.$$

where  $S = 2 \cdot r$  and  $Y = 2 \cdot I \cdot r - I$ . The optimum fitness result is attained at present which has an association with some of the parameters. It is not required to define the primary group of parameters and step sizes for the ideal solution.

Based on the fitness value of above iteration, the coefficient vector “Y” is attained by receiving the flexible probabilities’ function as follows.

$$y \Rightarrow Probability = \begin{cases} C_1(f_{max} - f_x)/(f_{max} - F_{avg}), & f_x \geq F_{avg} \\ C_3, & f_x \leq f_{avg}. \end{cases} \tag{14}$$

where,  $f_{min}$  and  $f_{max}$  denote the minimal and maximal values of Fitness Function (FF), but  $C_1$  and  $C_3$  lies in the range of [0, 1]. The place in the way of ideal solution gets dynamically altered by functional fitness. Then, to define the bubble-net nature of humpback whales, two enhanced methods are employed. The bubble-net process occurs with the help of exploitation as well as exploration phases.

Spiral procedure is employed amongst the place of whale and prey to mimic the helix-framed development of humpback whale which is given herewith [20]:

$$\vec{K}(t + 1) = e^{bt} \cdot \cos(2\pi \cdot y) \cdot \vec{U}' + \vec{K}^*(t). \tag{15}$$

It is observed that the humpback whale swims over the prey in constricting circle and also in a wind molded manner.

In order to illustrate this synchronous efficiency, 50% probabilities have been forecasted to have been optimized among the contracting surrounded and spiral schemes of whales. The mathematical expression as:

$$\vec{K}(t + 1) = \begin{cases} \vec{K}(t) - y \cdot \vec{U} & \text{if } p < 0.5 \\ \vec{U}' \cdot e^{bs} \cdot \cos(2\pi s) + \vec{K}(t) & \text{if } p \geq 0.5, \end{cases} \tag{16}$$

where  $y \rightarrow$  arbitrary value amongst [-1 to 1]. This illustrates the synchronous efficiency i.e., half probability that can be picked amongst the contract neighboring technique and twisted technique to revise the state of whales in the midst of optimization. The administrator makes a request by selecting a search expert aimlessly before an optimum search operator.

$$\vec{U} = |\vec{S} \cdot \vec{K}_{rand} - \vec{K}| \tag{17}$$

$$K(t + 1) = \vec{K}_{rand} - \vec{Y} \cdot \vec{U}. \tag{18}$$

So, the arbitrary values are employed mainly over 1 or under -1 to make the search agent reserved in reference to whale. Then during exploitation phase, the search place expert gets refreshed in the exploration phase as mentioned by arbitrarily-picked search operator. This is executed before an

optimum search operator shows up. This technique  $|\vec{Y}| > 1$  emphasizes the exploration and lets the WOA approach to progress towards a global optimal value and  $|\vec{Y}| > 1$  to stimulate the place of search agent. This process gets iterated until the maximal amount of iterations is achieved. A novel group of solutions is validated depending on the upgraded model.

#### 4 Performance Validation

The current section discusses the experimental validation of the proposed XAI-ASD technique on the applied benchmark dataset from UCI repository. The details related to the dataset are listed in [Tab. 1](#). Besides, the proposed XAI-ASD technique selected the following features from the dataset: 1, 3, 4, 6, 7, 8, 9, 12, 14, 16, and 19. Besides, [Tab. 2](#) lists out the features that exist in the ASD dataset.

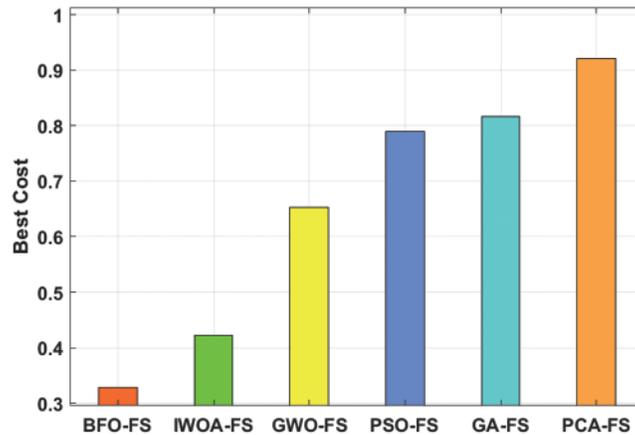
**Table 1:** Description of the dataset

No.	Dataset name	Sources	Number of attributes	Number of instances
1	ASD-children dataset	UCI	21	292
2	ASD-adolescent dataset	UCI	21	104
3	ASD-adult dataset	UCI	21	704

**Table 2:** Attributes in the applied dataset

Number	Attributes description
1	Patient age
2	Sex
3	Ethnicity
4	Born with jaundice
5	Family member with Pervasive Development Disorders (PDD)
6	Who is fulfilment the test
7	Country of residence
8	Used the screening app before or not
9	Screening test type
10–19	Based on the screening method answers of 10 questions
20	Screening score
21	Target class [Yes/No]

[Fig. 3](#) depicts the results achieved from best cost analysis of XAI-ASD technique against other methods. The figure showcases that PCA-FS and GA-FS techniques accomplished worse outcomes with the highest cost, whereas GWO-FS and PSO-FS techniques obtained slightly reduced ‘best cost’ values. At the same time, IWOA-FS technique achieved moderate best cost whereas the least best cost was achieved by XAI-ASD technique.



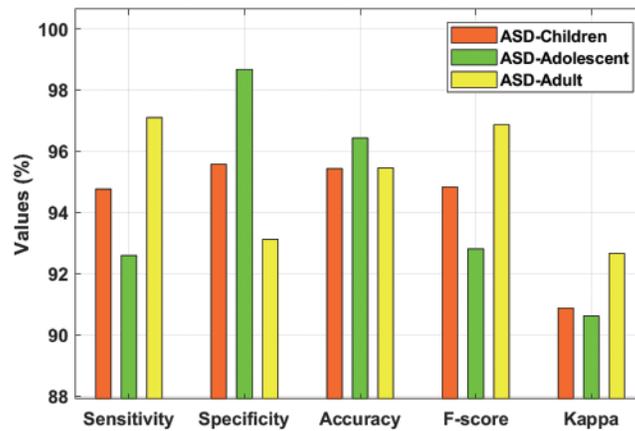
**Figure 3:** Best cost analysis of existing with proposed BFO-FS method

The ASD classification results of the proposed XAI-ASD technique on the applied dataset is given in [Tab. 3](#) and [Fig. 4](#). The experimental results demonstrate that XAI-ASD technique achieved the maximum ASD classification outcome. For instance, on ASD-Children dataset, XAI-ASD technique achieved a sensitivity of 94.76%, specificity of 95.57%, accuracy of 95.43%, F-score of 94.82%, and kappa of 90.87%. In addition, on ASD-Adolescent dataset, the proposed XAI-ASD approach obtained a sensitivity of 92.59%, specificity of 98.66%, accuracy of 96.43%, F-score of 92.80%, and kappa of 90.60%. Moreover, on ASD-Adult dataset, the proposed XAI-ASD method achieved a sensitivity of 97.09%, specificity of 93.12%, accuracy of 95.44%, F-score of 96.85%, and kappa of 92.67%.

**Table 3:** Results of the analysis of the proposed XAI-ASD method on applied dataset

Dataset	Sensitivity	Specificity	Accuracy	F-score	Kappa
ASD-children	94.76	95.57	95.43	94.82	90.87
ASD-adolescent	92.59	98.66	96.43	92.80	90.60
ASD-adult	97.09	93.12	95.44	96.85	92.67

In order to showcase the effective performance of XAI-ASD technique, a detailed comparative analysis was made against existing ASD diagnosis techniques and the results are shown in [Tab. 4](#) [20]. The results are inspected in terms of sensitivity, specificity, and accuracy.



**Figure 4:** Result analysis of XAI-ASD model with different measures

**Table 4:** Results of the analysis of existing and the proposed XAI-ASD method on applied dataset

Methods	Sensitivity	Specificity	Accuracy
XAI-ASD (children)	94.76	95.57	95.43
XAI-ASD (adolescent)	92.59	98.66	96.43
XAI-ASD (adult)	97.09	93.12	95.44
IWOA-FRBC (children)	93.57	93.42	93.49
IWOA-FRBC (adolescent)	91.09	97.08	95.45
IWOA-FRBC (adult)	95.24	92.68	94.23
Decision tree	53.30	54.90	54.70
Logistic regression	55.50	62.60	59.10
Neural network	53.30	71.20	62.00

Fig. 5 shows the results of sensitivity analysis of the proposed XAI-ASD technique against existing models. The figure portrays that DT, NN, and LR models produced low sensitivity values such as 53.3%, 53.3%, and 55.5% respectively. At the same time, IWOA-FRBC technique produced a somewhat moderate outcome. But the XAI-ASD (Children), XAI-ASD (Adolescent), and XAI-ASD (Adult) techniques accomplished maximum sensitivity values such as 94.76%, 92.59%, and 97.09% respectively.

Fig. 6 shows the results of specificity analysis of the proposed XAI-ASD approach against other recent algorithms. The figure exhibits that DT, LR, and NN methods achieved less specificity values such as 54.90%, 62.60%, and 71.20% correspondingly. Simultaneously, IWOA-FRBC technique produced a slightly moderate outcome. However, XAI-ASD (Children), XAI-ASD (Adolescent), and XAI-ASD (Adult) methodologies accomplished maximal specificity values such as 95.57%, 98.66%, and 93.12% correspondingly.

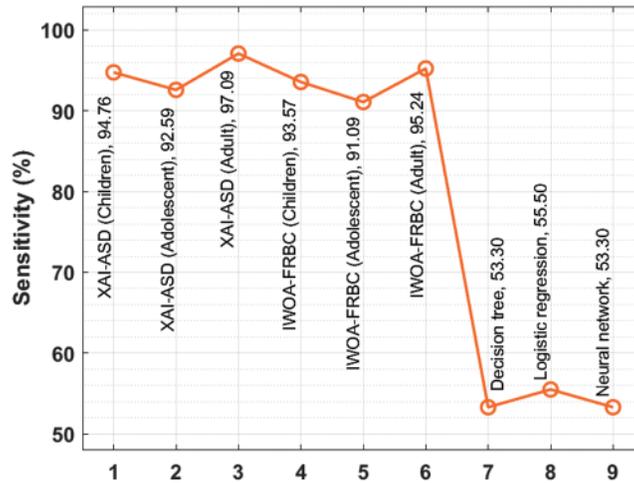


Figure 5: Sensitivity analysis of XAI-ASD model with recent techniques

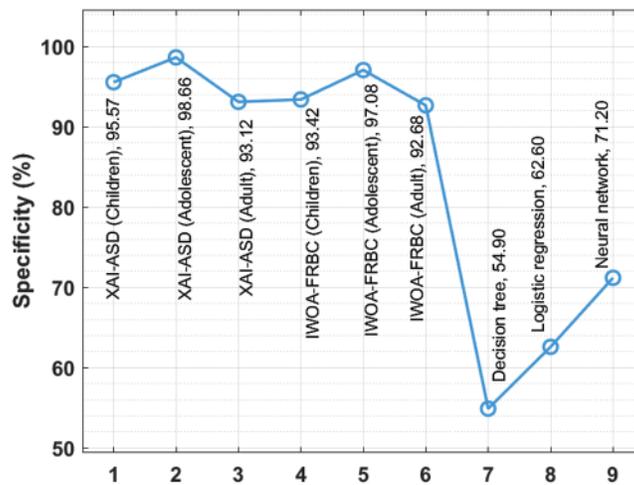
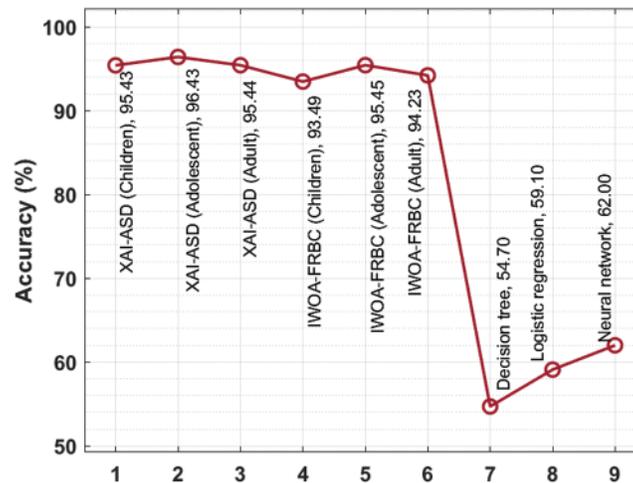


Figure 6: Specificity analysis of XAI-ASD model with recent techniques

Fig. 7 shows the results of accuracy analysis of the proposed XAI-ASD approach against other existing techniques. The figure shows that DT, LR, and NN methods achieved the least accuracy values such as 54.70%, 59.10%, and 62% correspondingly. Followed by, IWOA-FRBC method accomplished a somewhat moderate result. Finally, XAI-ASD (Children), XAI-ASD (Adolescent), and XAI-ASD (Adult) techniques accomplished superior accuracy values such as 95.43%, 96.43%, and 95.44% correspondingly. Based on the above results of the analysis, it is verified that the proposed XAI-ASD technique is an effective tool for diagnosis and classification of ASD.



**Figure 7:** Accuracy analysis of XAI-ASD model with recent techniques

## 5 Conclusion

In current study, a new XAI-ASD technique is designed and validated to detect and classify the different stages of ASD. The proposed XAI-ASD technique encompasses different processes namely pre-processing, WOA-based FS, DBN-based classification, and WOA-based parameter tuning. The use of WOA-FS technique helps in the selection of optimal feature subsets. Besides, the usage of WOA helps in the optimal selection of hyperparameters of DBN model that in turn helps in accomplishing an improved ASD diagnostic performance. In order to validate the ASD diagnostic outcomes by the proposed model, a series of simulations was conducted on benchmark dataset. The experimental results highlight the betterment of XAI-ASD technique over other recent state-of-the-art techniques under different measures. As a part of future work, the proposed XAI-ASD technique can be tested using real-time dataset collected from hospitals and IoT devices.

**Funding Statement:** The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work under grant number (RGP 2/209/42). [www.kku.edu.sa](http://www.kku.edu.sa). This research was funded by the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University through the Fast-Track Research Funding Program.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] F. K. Dosilovic, M. Brcic and N. Hlupic, "Explainable artificial intelligence: A survey," in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, Croatia, IEEE Explore, pp. 0210–0215, 2018.
- [2] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (xai)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.
- [3] W. Samek and K. R. Müller, "Towards explainable artificial intelligence," in *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning, Lecture Notes in Computer Science Book Series (LNCS)*, Springer, US, vol. 11700, pp. 5–22, 2019.

- [4] J. Uthayakumar, N. Metawa, K. Shankar and S. K. Lakshmanaprabu, "Intelligent hybrid model for financial crisis prediction using machine learning techniques," *Information Systems and e-Business Management*, vol. 18, no. 4, pp. 617–645, 2020.
- [5] W. Samek, T. Wiegand and K. R. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models," *ArXiv Preprint, ArXiv:1708.08296*, vol. 1, pp. 39–48, 2017.
- [6] M. G. Core, H. C. Lane, M. V. Lent, D. Gomboc, S. Solomon *et al.*, "Building explainable artificial intelligence systems," in *IAAI'06: Proc. of the 18th Conf. on Innovative Applications of Artificial Intelligence*, Boston, Massachusetts, pp. 1766–1773, 2006.
- [7] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (XAI): Toward medical XAI," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, pp. 1–21, 2020.
- [8] U. Masud, M. I. Baig, and A. Zeeshan, "Automatization analysis of the extremely sensitive laser-based dual-mode biomedical sensor," *Lasers Med Sci*, vol. 35, no. 7, pp. 1531–1542, 2020.
- [9] P. Koti, P. Dhavachelvan, T. Kalaipriyan, S. Arjunan, J. Uthayakumar *et al.*, "An efficient healthcare framework for kidney disease using hybrid harmony search algorithm," *Electronic Government, an International Journal*, vol. 16, no. 1, pp. 1, 2020.
- [10] U. Pawar, D. O. Shea, S. Rea and R. O. Reilly, "Explainable ai in healthcare," in *2020 Int. Conf. on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*, Dublin, Ireland, pp. 1–2, 2020.
- [11] S. M. Lauritsen, M. Kristensen, M. V. Olsen, M. S. Larsen, K. M. Lauritsen *et al.*, "Explainable artificial intelligence model to predict acute critical illness from electronic health records," *Nature Communications*, vol. 11, no. 1, pp. 3852, 2020.
- [12] T. Eslami, F. Almuqhim, J. S. Raiker and F. Saeed, "Machine learning methods for diagnosing autism spectrum disorder and attention- deficit/hyperactivity disorder using functional and structural mri: A survey," *Frontiers in Neuroinformatics*, vol. 14, pp. 575999, 2021.
- [13] T. Eslami, J. S. Raiker and F. Saeed, "Explainable and scalable machine learning algorithms for detection of autism spectrum disorder using fMRI data," *Neural Engineering Techniques for Autism Spectrum Disorder*, vol. 1, pp. 39–54, 2021.
- [14] H. Abbas, F. Garberson, S. L. Mayo, E. Glover and D. P. Wall, "Multi-modular ai approach to streamline autism diagnosis in young children," *Scientific Reports*, vol. 10, no. 1, pp. 5014, 2020.
- [15] S. Liang, A. Q. M. Sabri, F. Alnajjar and C. K. Loo, "Autism spectrum self-stimulatory behaviors classification using explainable temporal coherency deep features and svm classifier," *IEEE Access*, vol. 9, pp. 34264–34275, 2021.
- [16] A. Lombardi, D. Diacono, N. Amoroso, A. Monaco, J. M. R. S. Tavares *et al.*, "Explainable deep learning for personalized age prediction with brain morphology," *Frontiers in Neuroscience*, vol. 15, pp. 674055, 2021.
- [17] M. Zhang, Y. Ma, L. Zheng, Y. Wang, Z. Liu *et al.*, "An explainable diagnostic method for autism spectrum disorder using neural network," *Journal of Information Science & Engineering*, vol. 37, no. 2, pp. 1–16, 2021.
- [18] B. Niu, W. Yi, L. Tan, S. Geng and H. Wang, "A multi-objective feature selection method based on bacterial foraging optimization," *Natural Computing*, vol. 20, no. 1, pp. 63–76, 2021.
- [19] J. Yan, Y. Gao, Y. Yu, H. Xu and Z. Xu, "A prediction model based on deep belief network and least squares svr applied to cross-section water quality," *Water*, vol. 12, no. 7, pp. 1929, 2020.
- [20] K. Umamaheswari and L. Parthiban, "Improved whale optimization algorithm based feature selection with fuzzy rule base classifier for autism spectrum disorder diagnosis," *International Journal of Advanced Research in Engineering and Technology (IJARET)*, vol. 11, no. 11, pp. 41–55, 2020.