Deep Learning Approaches for Autism Spectrum Disorder: A Comprehensive Review

Sameeh A. Jassim [1]*, Aythem Kh. Kareem [2] and Ahmed A. Nafea [3]

[1]Department of Computer Sciences, College of Science, University of Al Maarif, Al Anbar, 31001, Iraq.
[2]Department of Heet Education, General Directorate of Education in Anbar, Ministry of Education, Heet, 31007 Anbar, Iraq.
[3]Department of Artificial Intelligence, College of Computer Science and IT, University of Anbar, Ramadi, Iraq.

***Corresponding author E-mail:** sameeh@uoa.edu.iq

https://doi.org/10.29072/basjs.20260108

ARTICLE INFO

Received: 22 August 2025

Accepted: 28 September 2025

Published: 30 April 2026

Keywords:

Autism Spectrum Disorder
Deep Learning
Vision Transformers
Neuroimaging (fMRI, EEG)
Explainable AI
Clinical Diagnosis

ABSTRACT

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental disorder with a complex etiology that presents obstacles to an early and reliable diagnosis. Recent innovations in deep learning (DL) have revolutionized ASD research by allowing the automatic identification of subtle, non-linear signatures from noisy and data-rich heterogeneous and multimodal sources. In this work, we provide the first systematic comparative study of 23 recent DL models for ASD detection, spanning multiple multimodal datasets from facial imaging, neuroimaging, functional magnetic resonance imaging (fMRI)/electroencephalogram (EEG), eye-tracking, kinematic profiles, and electronic health records (EHRs) via novel fusion architectures. Modern architectural designs such as Vision Transformers (ViTs), hybrid DL frameworks, attention-augmented models, and ensemble strategies achieve consistently high diagnostic accuracies that notably surpass the capabilities of classical machine learning (ML) approaches. Secondly, the improvement in transfer learning and multimodal fusion, together with interpretability methods (e.g., gradient-weighted class activation mapping (GradCAM) and attention heatmaps), makes features representable and strengthens the generalization of models. Although formidable progress has been made in this area, the adoption of real-world applications is still hampered by the heterogeneity of datasets, the lack of demographic representation and diversity among validation cohorts, and variations in clinical environments. This will necessitate the adoption of standardized pipelines, strict cross-population validation, and explanatory AI frameworks designed to ensure transparency and clinical utility. In sum, realizing the transformative capacity of DL for ASD Dx relies on ongoing interdisciplinary partnerships among computer scientists, clinicians, and neuroscientists to deliver new solutions that are both accurate and inclusive in nature (i.e., they need to be clinically relevant), as well as principled from an ethical point of view.

1. Introduction

Autism Spectrum Disorder is a neurodevelopmental disorder with impairment in social interaction, repetitive behavior, and communication. Global prevalence has increased exponentially from 1 in 160 children in 2012 to 1 in 100 in 2022, making ASD a significant public health issue [1-2]. The standard diagnosis relies on clinician-administered tools administered by clinicians, such as the Autism Diagnostic Observation Schedule and the revised Autism Diagnostic Interview, which are long, subjective, and frequently not accessible in low-resource settings[3-4]. Diagnostic delays average 4–5 years, missing the window of early intervention [5]. Artificial intelligence (AI), specifically deep learning, provides potential solutions through the automation of complex biomedical data analysis [6-8]. DL is highly capable of detecting subtle patterns from high-dimensional data -such as neuroimaging, electrophysiology, and behavioral data- that human observers might miss[9-10].

Early diagnosis is crucial to enable prompt intervention, yet traditional diagnostic methods are still time-consuming and subjective. Recent advances in DL have revolutionized the screening for ASD by enabling the automatic analysis of various biomarkers, facial appearance [11], neuroimages [12], eye-tracking behavior [13], and kinematic profiles [14]. Such approaches take advantage of multimodal data to address heterogeneity in ASD presentations, with studies demonstrating the success of convolutional neural networks (CNN) in the classification of facial images [11], transformers in health-record processing [15], and hybrid models (i.e., CNN- bidirectional long short-term memory network (BiLSTM) in the integration of functional magnetic resonance imaging (fMRI) and clinical data [16]. Non-invasive approaches, such as cry acoustic analysis [17] and upper limb movement assessment [14], further extend the range of available screening tools. This comparative survey consolidates 23 state-of-the-art DL frameworks and evaluates their performance, computational complexity, and clinical significance, highlighting advances in ensemble learning, attention mechanisms [18], and optimization strategies [19] that cumulatively pave the way towards large-scale early diagnosis of ASD. The objectives of this review are structured as the following research questions:

- What types of DL approaches have been applied to ASD diagnosis across different data modalities?
- How do these models perform in terms of accuracy and clinical applicability?
- What challenges limit their translation into clinical practice?
- What are the emerging trends and future directions in applying DL for ASD detection?

This review begins by establishing the foundational context in Section 2, covering the clinical aspects of ASD. Section 3 presents an overview of DL. Section 4 catalogs various data modalities (e.g., neuroimaging, behavioral videos, EHRs) and public datasets driving ASD research. Section 5 details DL applications for diagnosis, severity prediction, early screening, and multimodal fusion. Section 6 benchmarks the performance of the model through comparative analysis and tables. Critical challenges, including data scarcity, interpretability, and ethical concerns, are analyzed in Section 7, while Section 8 outlines future directions, including explainable AI and federated learning. The review concludes in Section 9 by synthesizing the transformative potential and emphasizing interdisciplinary collaboration to bridge research and clinical practice.

2. Background

This section introduces the background information relevant to understanding the complexity of ASD. The section begins with a definition of what constitutes ASD and its clinically heterogeneous presentation, and with the standardized diagnostic criteria as depicted in DSM-5. Additionally, this section discusses traditional evaluation methodologies, such as classical behavior tests in common use and clinician-based observation procedures, long a cornerstone of ASD diagnosis.

2.1 Autism Spectrum Disorder

ASD is one of the common neurodevelopmental disorders in children worldwide, and recent prevalence estimates are as high as 2.78% within the United States population. Symptoms include persistent issues with social interaction and social communication, accompanied by restricted, repetitive patterns of activities, behaviors, or interests. The diversity of ASD symptoms is so great that it is accompanied by the heterogeneity of symptoms observed in the patient population after evaluation. The diagnosis using the current methodology is based mainly on clinical observation, standardized behavioral assessment, and developmental history (which requires specialist personnel and can result in the diagnostic process, resulting in the fact that the median age of diagnosis is currently at about 60 months [20].

2.1.1 Clinical features and diagnosis criteria (e.g., DSM-5)

ASD is a broad, multifaceted care provision and neurodevelopmental disorder that affects social interaction, as well as repetitive behaviors, restricted interests, and activities. According to the Statistical Manual and Diagnostic of Mental Disorders, Fifth Edition, these core elements present during early adolescence frequently cause substantial daily functioning to materialize [21-22]. The DSM-5 has combined prior subcategories into a monoclonal spectrum while highlighting differences in the intensity and impact [23]. Clinical characteristics may include the following:

2.1.1.1 Social Communication Deficits

Difficulties in social-emotional reciprocity (eg, abnormal social approach, reduced sharing of interests), nonverbal communication (e.g., poor eye contact, limited gestures), and developing/maintaining relationships [21].

2.1.1.2 Restricted and Repetitive Behaviors (RRBs)

By their very nature, these are behaviors that are difficult to change. In this way, it can be seen that, regardless of its other ramifications, just like mother's grits, these are held on stubbornly and not allowed to rust away. Such characteristics are recorded in the International Classification of Diseases. Whether it is due to some biological reason or to man's genes, one thing is certain: ‘These traits are not easy to overcome’. The International Classification of Diseases, 11th revision (ICD-11) aligns with DSM-5, further classifying ASD based on co-occurring intellectual disability and functional language impairment[22].

2.1.2 Traditional assessment methods

The diagnosis is made using behavior and clinician-completed rating tools, typically more than one discipline professional (e.g., speech therapists, neurologists, psychologists) [14].

2.1.2.1 Behavioral Tests

- Autism Diagnostic Observation Schedule (ADOS): social interaction, Semi-structured communication, and play assessment with calibrated severity scores (ADOS-CSS) to quantify the extent of symptoms.
- Childhood Autism Rating Scale (CARS): Rates behavior from 1 through 4 based on direct observation, interviews, and collateral information.

- Social Communication Questionnaire (SCQ): Parent-rated screening of communication, social competence, and RRB.

2.1.2.2 Developmental and Adaptive tests [22]**:**

- Vineland Adaptive Behavior Scales (VABS): Adaptive measure of communication, socialization, daily living, and motor behavior.
- Mullen Scales of Early Learning (MSEL): Cognitive function is measured using subsets of visual reception, fine motor, and expressive/receptive language.

2.1.2.3 Screening Tools

Modified revised autism checklist for children with follow-up (M-CHAT-R/F): Parent report screening of toddlers for early detection.

These approaches are resource-consuming and subjective, and they usually result in diagnosis delays, particularly in underrepresented groups[21].

3. Deep Learning Overview

DL, a branch of ML, uses Neural Networks (NNs) with numerous hierarchical levels, "deep architectures" to learn in an automated manner complex representations from high-dimensional inputs. It represents a network that includes neurons with multiple layers and parameters between the input layer (IL) and output layer (OL). DL employs NN topologies as its base. Therefore, they are considered to be deep neural networks (DNNs). DL delivers independent characteristics of learning and its hierarchical model at multiple levels. In contrast to traditional ML techniques, this robustness is a development of the robust strategy; in brief, the fundamental architecture of DL is operated for the modification and extraction. The earlier layers perform essential processing to learn superficial characteristics or incoming data, and the result is transmitted to the upper layers, which are responsible for learning more complicated characteristics. Therefore, DL is well-suited to manage larger datasets and greater complexity [24-25]. Activation functions (AFs) recreate a decision-making function in the NNs, which is a facilitative to learning nonlinear complicated patterns. They are mainly used to present non-linear characteristics in NNs, thereby improving their capability to fit data. Table 1 presents the details of the typical types of AF in NNs, including tanh, sigmoid, ReLU, and its variants Randomized ReLU, Parametric ReLU, Leaky ReLU, Exponential linear unit, Maxout, and Switch. Clearly, compared to AFs such as tanh and sigmoid, ReLU does not saturate for both small and extensive inputs. Therefore, ReLU and its variants are the greatest in overcoming the

vanishing gradient issue. Concerning Maxout AF, it bypasses neuron death and other issues by taking advantage of the linearity and unsaturation advantages of ReLU functions [29].

Table 1. Summary of activation functions [26].

Activation Function	Definition
Tanh	$f(x) = \frac{2}{1+e^{-2x}} - 1$
Sigmoid	$f(x) = \frac{1}{1+e^{-x}}$
ReLU	$f(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$
PReLU	$f(x) = \begin{cases} x & \text{if } x \geq 0 \\ ax & \text{if } x \leq 0 \end{cases}$
LeakyReLU	$f(x) = \begin{cases} x & \text{if } x \geq 0 \\ ax & \text{if } x < 0 \end{cases}$
RReLU	$f(x) = \begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x \leq 0 \end{cases}$
ELU	$f(x) = \begin{cases} x & x \geq 0 \\ x & x < 0 \end{cases}$
Maxout	$f(x) = max_{j\in[1,k]} x^T W + b_{ij} W \in R^{d\times m\times k}$
Swish	$f(x) = x.sigmoid(\alpha x)$

3.1 Deep Learning Types

3.1.1 Artificial Neural Network

ANN is the fundamental method of DL and the prototype of other complicated DNNs. Its basic configuration is displayed in Figure 1, in which each neuron in a layer is fully connected to all neurons in the following layer, and each association is allocated a specific weight. Via continually modifying these weights, the network gradually improves its ability to map input to output accurately. However, with their impressive ability to infer patterns from inaccurate or complex data, ANNs can be employed to detect patterns and discover trends that are too complicated for individuals or other computer strategies to detect. A trained ANN can be assessed as an "expert" in the classification of information to which it has been trained to estimate [27-28].

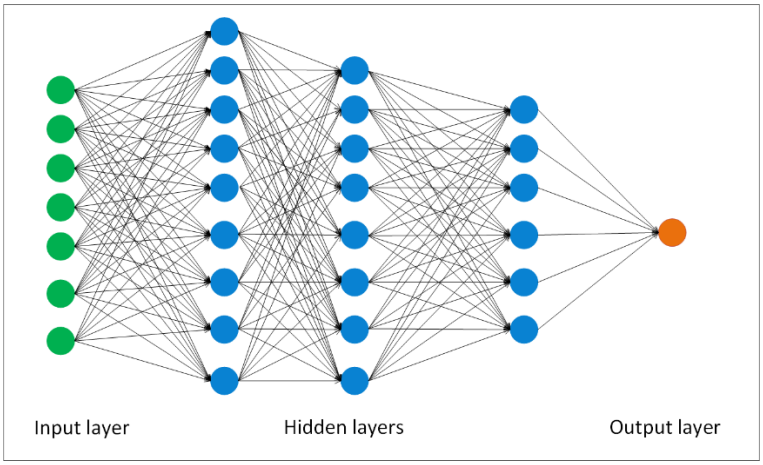

Figure 1. The ANN architecture

The main architecture of an ANN is made up of three types of layers, including an IL, one or more Hidden Layers (HLs), and an OL, which can be summarized in **Table 2** [27].

Table 2. The main architecture of the ANN

Layer	Description
Input Layer (IL)	The input layer (IL) is the ANN's entry point, matching the number of dataset features and passing them to the next layer without analysis.
Hidden Layers (HLs)	HLs are the core computational units of the ANN. HLs consist of neurons that involve weighted sums to their inputs, observed by a nonlinear activation function (AF) like Tanh or Sigmoid.
Output Layer (OL)	The OL is the final layer, responsible for creating the result of the analysis of the model.

3.1.2 Convolutional Neural Networks (CNNs)

The CNN is a type of feedforward NN that can cut features from data utilizing convolutional structures. Unlike traditional feature extraction (FE) techniques, CNN does not need manual FE. The principles of visual perception inspire CNN architecture. CNN kernels express various receptors that can react to different features. Activation Functions (AFs) simulate the function that only neural signals surpassing a specific threshold can be transmitted to the following neuron. Loss functions (LF) and optimizers are tools that people have developed to implement the entire CNN technique, allowing it to learn what to expect. CNN has three advantages that make it one of the most common techniques in the field of DL. These advantages are [29]:

- Local connections: Each neuron is connected to only a small number of neurons in the preceding layer, which is sufficient to accelerate convergence and reduce parameters.
- Weight sharing: A set of connections can communicate the exact weights, which further decreases the parameter number.

- Down-sampling dimension decrease: A Pooling Layer (PL) harnesses the principles of local correlation to down-sample, decreasing the amount of data while maintaining useful data. It can even decrease the number by drawing insignificant features.

Figure 2 illustrates the primary architecture of CNNs, which typically involves an IL, a Convolution Layer (CL), a Pooling layer (PL), a Fully Connected Layer (FCL), and an OL.

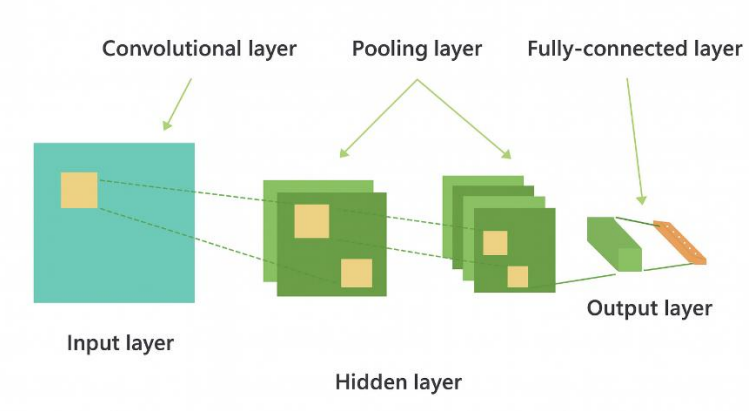

Figure 2. Basic architecture of CNN [28]**.**

CNN architecture plays a crucial role in configuring NN architectures, as a further suitable network architecture can improve the fitting development among layers or decrease redundant calculations within the network, resulting in excellent performance. CNN layers can be described as follows [26]:

- Input layer (IL): The IL of a CNN is the first layer that accepts the raw data. It does not serve any calculations, but rather serves as the basis for passing data to subsequent layers for FE. The IL confirms that the data are appropriately formatted and scaled for processing by the CLs that follow.
- Convolution Layer (CL): CL includes a convolution kernel set, whose working diagram is displayed in Figure 3.

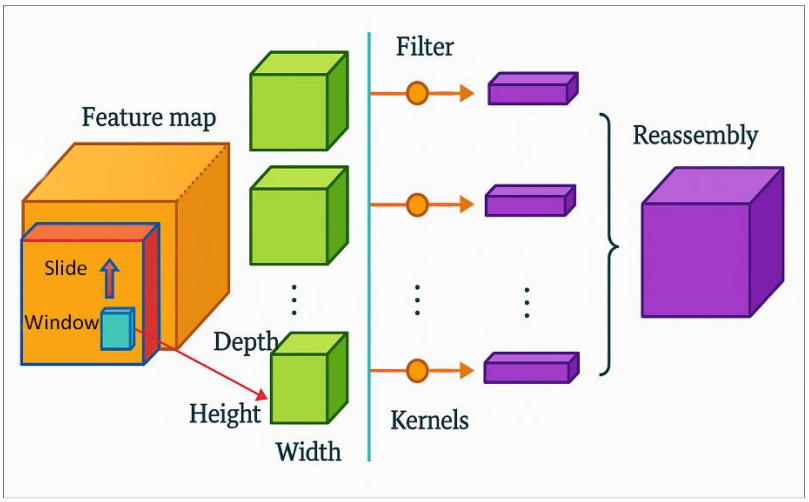

Figure 3. Working diagram of CL [26].

The method involves sliding a specific fixed-size window across the feature map, extracting neighboring feature tiles at different stages, and completing the tensor product of each feature tile with the comprehended convolution kernel of the weight matrix. The mathematical formula of the convolution process can be expressed in Equations (1) and (2):

$$x_i^k = f^k\left(u_i^k\right) \qquad \textbf{Eq. 1}$$

$$u_i^k = \sum_{j \in N_i} x_j^{k-1} * L_{ji}^k + b_i^k \qquad \textbf{Eq. 2}$$

Where: x_i^k describes the output of i^{th} channel in the CL k, $f^k(.)$ stands for the AF of CL, u_i^k describes the net activation of i^{th} channel in the CL k, N_i describes the feature map subset sampled via a sliding window in the feature map, and L_{ji}^k, b_i^k respectively represent the convolution kernel matrix of the CL k and the bias of the feature map.

- Pooling layer (PL) involves reducing the dimensions of input features in CNN, retaining the main structures, but reducing the computation time. It is often realized by adopting max pooling, average pooling, or stochastic pooling. In max pooling, the maximum value in a local receptive field is kept, while average pooling computes the average value in the local receptive field. Stochastic pooling, in contrast, samples a value from feature points included within the region. The pooling operation can be mathematically formulated as in Equations (3) and (4).

$$X_i^k = f^k\left(u_i^k\right) \qquad \textbf{Eq. 3}$$

$$u_i^k = \alpha_i^k . f_{pooling}^k\left(x_i^{k-1}\right) + b_i^k \qquad \textbf{Eq. 4}$$

Where: X_i^k and u_i^k respectively indicate the output and net activation of i^{th} channel in the PL k, $f^k(.)$ stands for the AF connected to PL k, α_i^k represents the weight coefficient of PL k, and $f_{pooling}^k$ describes the pooling function of PL k.

- Fully connected layer (FCL): FCL is typically located after the PL of a CNN, which is used to transform the 2-dimensional feature information output from the earlier layer into 1-dimensional. The output of FCL is received through a weighted mixture of FCL neurons in the earlier layer. The mathematical formulation of each neuron can be calculated using Equations (5) and (6):

$$x_i^k = f^k(u_i^k) \qquad \textbf{Eq. 5}$$

$$u_i^k = \max\left(0, \sum_j y_j^k . w_{ji}^k + b_i^k\right) \qquad \textbf{Eq. 6}$$

Where: x_i^k and u_i^k respectively describe the output and net activation of the i^{th} channel in the *FCL k,* $f^k(.)$ stands for the AF connected by the pooling layer k, w_{ji}^k and b_i^k are respectively, the bias and weight coefficient of *FCL k*, and *max ()* represents the most significant number in the selection range.

- Output Layer (OL): The OL is the layer that provides the model prediction. When multi-class or binary classification is the target problem, this is usually followed by an FCL. This layer converts the abstract representations learned in previous layers into a probability distribution or actual prediction calls, allowing the network to make decisions about the input data.

3.1.3 Recurrent Neural Networks

RNNs are a type of ANN specifically developed to model time-series or sequential data for sequence prediction and recognition. Time-series data include inherent temporal data that a simple prediction cannot capture. RNNs are developed with a state of memory that allows them to keep a hidden state qualified for storing data from earlier time phases in a sequence. A high-dimensional hidden condition and non-linear consequences enable the RNN hidden condition to possess excellent expressive capacity, allowing it to combine data over multiple timesteps and supply accurate predictions. However, standard RNNs have a limited memory capacity, which determines their ability to affect long-range networks. Due to their incapacity to create long-lasting memories, their ability to develop sequences is inconsistent. In such cases, it is

typical for forecasts to deviate from the manifold on which the training data are located [30]. Figure 4 displays the key architecture of RNN.

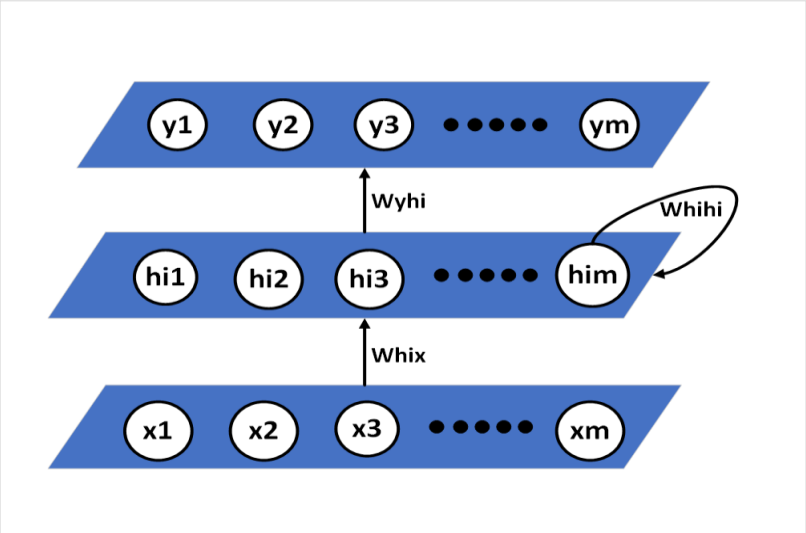

Figure 4. The key architecture of RNN

The RNN model receives an input vector sequence (x1,. . , xm). The RNN model computes the hidden states hi_m and outputs yhi_m through iteration of the corresponding Equations (7) and (8).

$$hi_m = f_{hi}(x_m w_{hix} + hi_{m-1} w_{hihi} + b_{hi}) \quad \textbf{Eq. 7}$$

$$y_m = f_y(hi_m w_{yhi} + b_y) \quad \textbf{Eq. 8}$$

3.1.4 long short-term memory (LSTM)

LSTM is a supervised DL method, and it is an extension of RNN. Evaporating gradients are a problem with RNNs that makes learning from extensive data challenging. When the gradient becomes smaller and lower, the parameter updates evolve insignificantly, indicating that no significant learning is occurring. The gradients include information utilized in the RNN parameter update. Due to the linear gradient issue, RNN and LSTM are employed. In comparison to RNN, it can place an extensive amount of data from earlier stages. The LSTM technique provides data from the existing time case, and CLs result from the earlier time point [31]. Figure 5 illustrates the main structure of an LSTM, including four NNs and different memory blocks comprehended as cells.

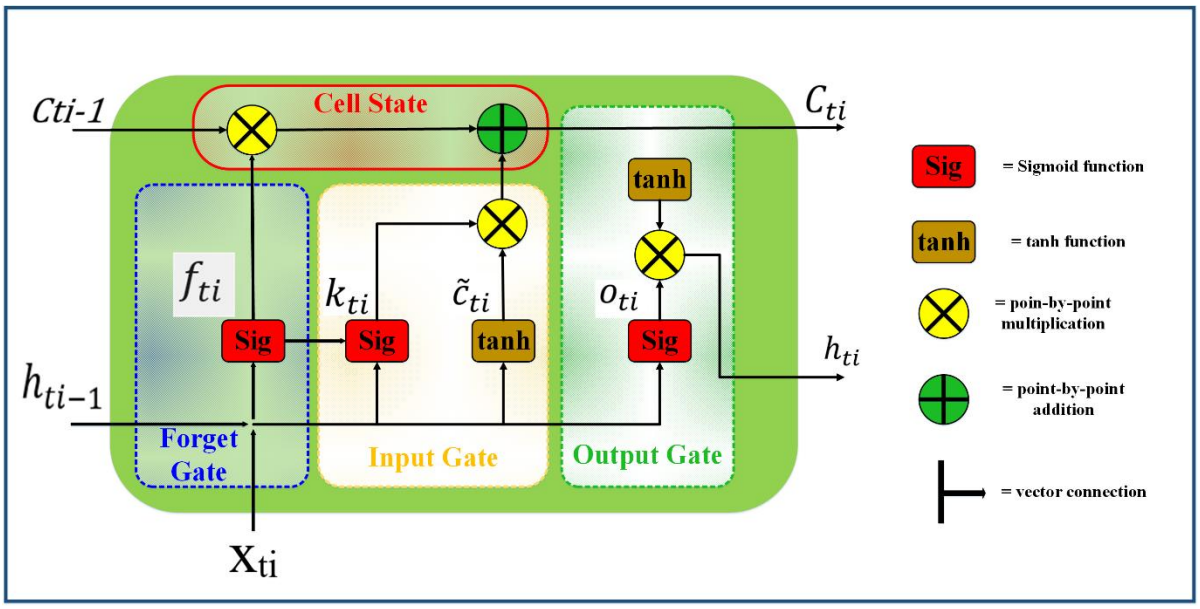

Figure 5. Structure of LSTM [32].

The LSTM networks are equipped with three essential gates to control the information flow: the forget gate (FG), the input gate (IG), and the output gate (OG). The forget gate decides which information in the cell state from the previous time step should be thrown away. At each timestep, the current input (x_{ti}) and the previous hidden state (h_{ti-1}) are passed through trainable matrices with bias, then activated using a sigmoid function (Sig), resulting in values between 0 and 1. Or using the probability interpretation where "0" means discard the information, and "1" means save it, as quantified in Eq. (9) [32].

$$f_{ti} = \sigma\left(W_f.[h_{ti-1}, x_{ti}] + b_f\right) \qquad \textbf{Eq. 9}$$

Where f_{ti} is forget gate activation at time step ti, W_f is the weight matrix for the forget gate, h_{ti-1} is the previous hidden state, x_{ti} is the current input, b_f is a bias term, and σ is a sigmoid function.

Following that, one must plan what data is required to be saved in the cells. This strategy consists of two components. First, the gate layer utilized for input, which is also a sigmoidal coating layer, determines the values to be updated. Second, a vector for a new character ti is developed by a hyperbolic tangent function ($tanh$) layer, planned for adding, in this state, as shown in Equations (10) and (11).

$$k_{ti} = \sigma(W_{ti}.[h_{ti-1}, x_{ti}] + b_{ti}) \qquad \textbf{Eq. 10}$$

$$\tilde{c}_{ti} = tanh(W_c[h_{ti-1}, x_{ti}] + b_c) \qquad \textbf{Eq. 11}$$

Where k_{ti} is input gate activation at time step ti, W_{ti} is the weight matrix for the input gate, h_{ti-1} is the previous hidden state, x_{ti} is the current input, b_{ti} is a bias vector for the input gate, $\tilde{c}_{ti}$ is the candidate cell state, and b_c It is a bias vector for candidate memory.

After all the decision-making and planning have been accomplished, the implementation takes place in this phase. The previous cell state *Cti-1* is updated to *Cti*. To directly forget the data as planned earlier, f_{ti} is multiplied by the earlier state, which is observed via the addition of $(f_{ti}\tilde{c}^*_{ti})$. This value, thus received, is the value of the new character, which is scaled via a limit in agreement with the decision that was made to update the cell state value, as shown in Equation (12).

$$C_{ti} = f_{ti}\tilde{c}^*_{ti-1} + k_{ti} * C_{ti} \qquad \textbf{Eq. 12}$$

Where C_{ti} is the cell state at time step ti, $\tilde{c}^*_{ti-1}$ is a modification of the previous candidate cell state.

It is now essential to plan the output, which is specified by the cell state, yet this output will be filtered. First, a sigmoid layer determines the part of the cell state that must be introduced as the output. Observing this, the cell state is passed through a *tanh* function (to determine the results to be 1 or -1), and then it can be boosted by multiplying it by the result of the sigmoid gate layer to acquire the exact output as decided, as shown in Equations (13) and (14).

$$o_{ti} = \sigma(W_o.[h_{ti-1}, x_{ti}] + b_o) \qquad \textbf{Eq. 13}$$

$$h_{ti} = o_{ti} * tanh(C_{ti}) \qquad \textbf{Eq. 14}$$

Where o_{ti} is the output gate at time step ti, W_o is the weight matrix for the output gate, b_o is the bias vector for the output gate, and h_{ti} is the hidden state at time ti.

3.1.5 Multilayer Perceptron (MLP)

MLP is the most effective and fundamental data-driven modeling mechanism in ANNs. An MLP with one IL, two HLs, and one OL can be illustrated in Figure 6. The IL is the initial phase of data input in the MLP. IL collects input variables from external data. These input variables enable the MLP to learn and comprehend the complex associations between the parameters. Consequently, it is crucial to organize and preprocess input variables before providing them to the MLP. The input variables are then handed over to the HL. HLs are vital for optimal MLP performance, especially when time complexity and accuracy are the primary considerations. Neurons in the HLs accept input values from the IL and compute the weighted

sum of these values. Then, an individual neuron uses an AF to calculate the weighted sum. An AF specifies whether a neuron should be activated based on its input value. Moreover, AF incorporates nonlinearity into the MLP model, enabling MLP to learn and recognize complicated data patterns. The activation results of all neurons are integrated to create an output, which is transferred to the model's OL; this output describes the MLP prediction output. Suppose that a four-layer MLP model (one IL, two HL, and one OL) is shown in Figure 6 [30].

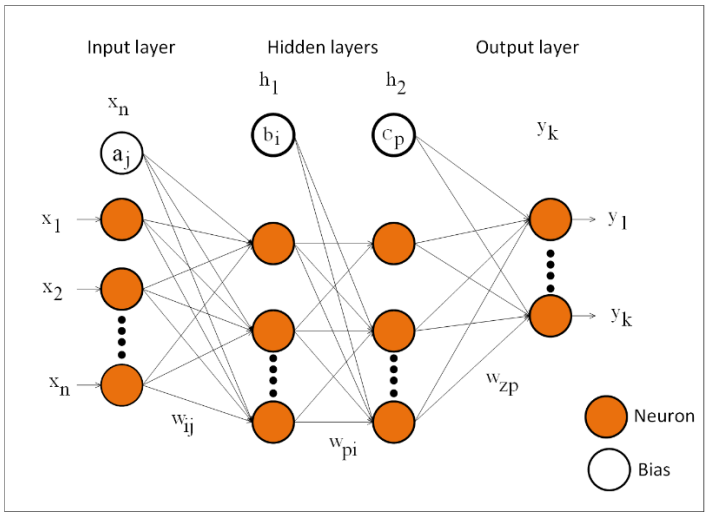

Figure 6. MLP architecture.

There are n input variables in the input node (xi, j = 1, 2, 3, etc.,n) and the output variables k in the output node (yl, l = 1, 2, 3, etc., k). The network consists of two HLs with n number of nodes in layer 1 and p number of nodes in layer 2 (such that hi, i = 1, 2, 3, etc., m and hp, p = 1, 2, 3, etc., q). Equations (15), (16), and (17) provide a mathematical description of the MLP model [30]:

$$h_n = f_h\left(\sum_{j=1}^{n} x_j w_{ij} + a_j\right) \qquad \textbf{Eq. 15}$$

$$h_o = f_h\left(\sum_{i=1}^{o} h_i w_{pi} + b_i\right) \qquad \textbf{Eq. 16}$$

$$h_k = f_y\left(\sum_{k=1}^{k} h_p w_{ik} + c_p\right) \qquad \textbf{Eq. 17}$$

Where w_{ij}, w_{pi} , and w_{ik} describe the weight parameters, a_j , b_i, and c_p describe the bias, and f_h and $f_y$describe the AFs.

3.1.6 Graph Neural Networks (GNNs)

Graph Neural Networks (GNNs) are a type of NNs that are designed to process graph-structured data. They are regularly classified into four categories: recurrent graph neural networks (RGNN), convolutional graph neural networks (CGNN), graph autoencoders (GAE), and spatial-temporal graph neural networks (STGNN), respectively. We define a graph by a skeleton Q = (E, V), where *E* corresponds to the set of edges, and *V* represents the set of nodes, and each edge (u,e) is a pair that connects node u to node e. GNNs learn to model whether two nodes are related or not and to generalize those relationships into useful representations. Even though there are diverse architectures, most GNNs are developed based on the neighborhood aggregation (or message-passing) framework that forms a basic idea to learn node and graph-level features, as shown in Figure 7 [33].

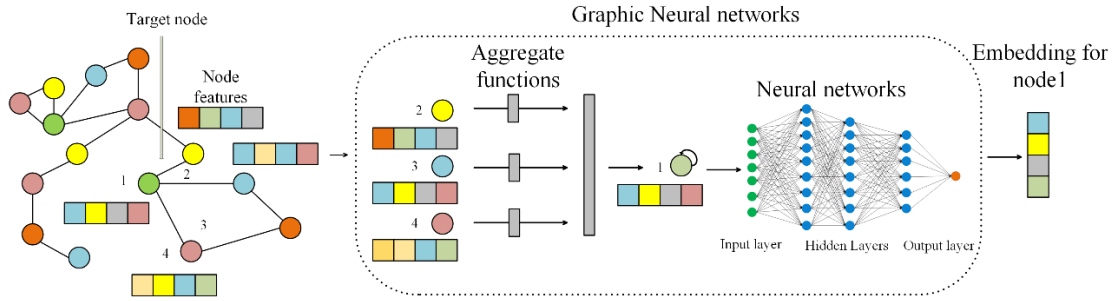

Figure 7. Functional architecture of a GNN

Mathematically, a GNN can be described in terms of its aggregation and update functions. Given a graph G = (E, V) with node features X, the GNN functions in layers. At each layer, every node aggregates data from its neighbors and perhaps itself. Afterward, the aggregated data is used to update the node's characteristics. The aggregation function can be calculated from Equation (18) [33].

$$a_k^{(v)} = AGGREGATE^{(v)}\left(\left\{h_u^{(v-1)} \middle| u \in Neighbors(e)\right\}\right) \quad \textbf{Eq. 18}$$

Where $a_k^{(v)}$ is the feature of node v from the initial features of node features for v = 1, and $AGGREGATE^{(v)}$ is a function that aggregates the features.

The update function can be calculated from Equation (19).

$$h_k^{(v)} = UPDATE^{(v)}\left(h_k^{(v-1)}, a_k^{(v-1)}\right) \quad \textbf{Eq. 19}$$

Where $h_k^{(v-1)}$ is the feature of node e from the earlier iteration, $a_k^{(v)}$ is the aggregated feature from the first phase, and $UPDATE^{(v)}$ is a function that updates the node's features.

In GNNs, aggregation aggregates information from neighboring nodes by functions such as mean, sum, or weighted sum. Then, the aggregated information is incorporated with a node's current features through deep neural network layers to obtain refined representations that encode both local and neighborhood information. It is this information propagation process, called message passing, which will be iteratively exploited to exchange and fuse information between nodes for learning local and global graph structures [33].

3.1.7 Autoencoders (AEs)

AEs are a specific kind of NN developed to learn efficient encodings of unlabeled data, generally for feature learning or dimensionality reduction. AEs seek to produce their input at the OL after decoding and encoding it via an HL, understood as the bottleneck. Figure 8 presents the general structure of AE [34].

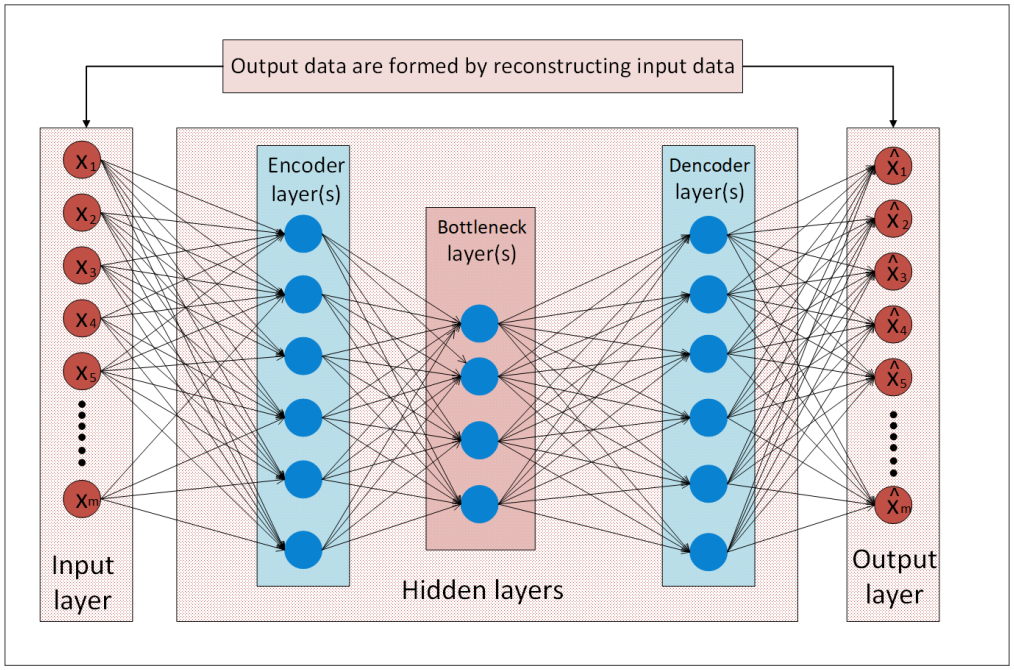

Figure 8. The general structure of AE [34].

This bottleneck is a vital feature of the AE architecture, as it forces the network to prioritize which characteristics of the input data are significant, thus learning to ignore 'noise' and less significant details. An AE is made up of two main components: the decoder and the encoder. The encoder function ：ℝd → ℝp maps the input x to a latent space model z, where typically p < d, representing the latent space, has fewer dimensions than the input space. The decoder

function: $\mathbb{R}p \rightarrow \mathbb{R}d$ then attempts to reconstruct the input, denoted x, from this condensed data. The design of these two functions can represent the total autoencoder, as shown in Equation (20) [35].

$$\hat{X} = \Psi(\phi(x)) \qquad \textbf{Eq. 20}$$

The objective of the AE is to underestimate the difference between the input x and its reconstruction y, a measure that is quantified utilizing a loss function. The loss function most commonly used in AE is the mean squared error (MSE), which is utilized to quantify the difference between the initial and reconstructed inputs. The loss function that is often more suitable for binary data is cross-entropy. For training data, X = x (1), x (2),..., x (m), the MSE and cross-entropy loss functions are represented in Equations (21) and (22) [35].

$$L(\phi,\psi) = \frac{1}{m}\sum_{j=1}^{m}\left\|x^{(j)} - \psi\left(\phi\left(x^{(j)}\right)\right)\right\|^2 \qquad \textbf{Eq. 21}$$

$$L(\phi,\psi) = -\frac{1}{m}\sum_{j=1}^{m}\left[x^{(j)}\log\left(\psi\left(\phi\left(x^{(j)}\right)\right)\right) + \left(1 - x^{(j)}\right)\log\left(1 - \psi\left(\phi\left(x^{(j)}\right)\right)\right)\right] \qquad \textbf{Eq. 22}$$

However, regularization representations, such as the L1 norm, may even be considered a loss function. Regularization prevents overfitting via penalizing large weights, and, when employed in AEs, it can even confirm that the learned models are sparse, which often enhances the usefulness and interpretability of the encoded features.

3.2 The Efficacy of DL in ASD Analysis

DL is particularly suited for research and diagnostic activities with respect to ASD because it inherently has strengths in resolving the core issues of the disorder:

3.2.1 Complex Feature Extraction from High-Dimensional Data

ASD manifests in subtle, non-linear, and in most cases, distributive forms in different data modalities (structural/functional magnetic resonance imaging, eye tracking, behavioral questionnaires, and gene data). DL models, particularly CNNs (for imagery), GCNs (for brain networks), and AEs/DBNs (for feature learning in general), learn hierarchical representations automatically from raw or pre-processed data. This avoids manual feature engineering, which is subjective in nature, labor-intensive, and potentially inadequate to represent the full richness of ASD biomarkers[36]. For example, GCNs can decode broadly impacted functional

subnetworks (eg, DMN, ECN, VIS, LIM) from different brain disorders from fMRI-derived brain functional networks (BFNs)[36].

3.2.2 Multimodal and Heterogeneous Data Pattern Recognition

Heterogeneity at the phenotypic and neurological levels in ASD calls for models with the ability to merge information from various sources. LSTMs can capture temporal patterns in behavior time series[37], and ensemble techniques combining AEs, LSTMs, and DBNs can capitalize on the complementarity of the models in capturing intrinsic differences and structures between data types (e.g., fusion of imaging and behavior scores) to enable improved detection of ASD[37]. The ability to unify representations from multimodal inputs forms the core of the development of integrated diagnostic tools. Modeling Spectrum and Relationships: One advantage of DL models lies in learning continuous representations of data in latent spaces. Deep-layer representations of fMRI data from a model trained on a spectrum of disorders by Liu et al. [36] indicated a continuous spectrum of patients with respect to ages and diagnoses in accordance with the clinician's concept of a "spectrum" of brain disorders. Such a built-in ability allows DL to model gradients and associations between subtypes of ASD or between ASD and associated conditions, such as ADHD, in a better way than discrete diagnostic groups[36].

3.2.3 Addressing Subjectivity and Enabling Automation

Traditional ASD diagnosis is based primarily on subjective evaluations and clinical observation and suffers from heterogeneity and delay. DL provides a path toward increasingly objective data-driven tools. By learning complex mappings from data to diagnostic labels directly (e.g., gaze behavior and facial expressions to features of the brain network and questionnaire responses), DL models can contribute to screening and diagnostic tasks with the potential for reducing waiting times and enhancing accessibility[37]. Methods like transfer learning, in which knowledge accumulated in a single disorder or data set is applied to another (e.g., fine-tuning a model previously trained in MCI/AD/VCI/ADHD data for detecting ASD)[36], enhance efficiency and leverage existing data.

4. Sources and Modalities of the Data

DL techniques for the identification of ASD use different data modalities that represent different pathophysiological or behavioral markers. The studies reviewed in this review depend mainly on the following sources of data:

4.1 Imaging data

Functional magnetic resonance imaging (fMRI) and structural magnetic resonance imaging (sMRI) are the most prominent neuroimaging modalities. The sMRI modality is used to provide high-resolution anatomical images of gray and white matter for structural differences associated with ASD, for example, variations in the cerebellum, frontal lobe, and limbic system. The BOLD signals measured with rs-fMRI define the current resting state condition of FC for the discovery of biomarkers that describe pathological connectivity among different areas of the brain. Typical preprocessing for fMRI data involves a series of analyses, including co-registration with sMRI, normalization to some standard space (e.g., MNI152), and smoothing, as well as the removal of trends, and some kind of filtering, denoising (e.g., nuisance regression), and then the correlation matrices (e.g., using Pearson correlation) among regions of interest defined by a brain atlas, e.g., AAL - Automated Anatomical Labeling (116 ROIs). The Autism Brain Imaging Data Exchange (ABIDE I) dataset forms the main resource consisting of sMRI and rs-fMRI data from 1112 participants (539 with ASD and 573 typically developing (TD) controls) collected from a host of global locations. Scanning protocol and participant state variations (e.g., eyes open/closed) in different sites pose challenges, which are overcome by site-specific selection of data (e.g., the NYU site with 172 participants) or by multisite learning approaches. The ADHD-200 dataset, consisting of rs-fMRI and sMRI from participants with Attention-Deficit/Hyperactivity Disorder (ADHD) and TD controls, is used for comparative analysis and identification of comorbidity (ASD+ADHD) [38-39].

In addition, electroencephalography (EEG) has been mentioned in studies as a method to capture the electrophysiological activity pattern in the brain that can be indicative of ASD. Longitudinal analysis of EEG power has been studied as a method of distinguishing different outcomes [39].

4.2 Behavioral and audio-visual data

Facial images, facial morphology, and expression are recognized as potential indicators of underlying neurodevelopmental differences in ASD. Studies use facial photographs, such as the Kaggle Autism Image Dataset (containing 2926 facial images labeled "Normal" or "Autism"), for automated detection. Preprocessing involves face detection (eg, using MTCNN), resizing, normalization (e.g., Z-score), noise reduction, and augmentation (e.g., rotation, flipping, cropping, brightness adjustment) to handle variations and prevent overfitting [18].

4.3 Auditory and linguistic information

Infant/Toddler Cry Analysis: Acoustic features of cries are investigated as noninvasive early vocal biomarkers. Studies use data sets that contain audio recordings of cries, such as the one described by Khozaei et al., which comprises cries from 62 children aged 18-54 months (31 ASD, 31 TD). Recordings are made in WAV format (16-bit, 44.1 kHz) in various environments (homes, autism centers, health centers) using high-quality recorders or smartphones, ensuring quiet conditions and excluding pain-associated cries. Preprocessing involves extracting pitch-based acoustic features like Jitter (pitch variability), Shimmer (amplitude variability), and Harmonics-to-Noise Ratio (HNR) (periodicity/voicing) using software like Praat. Statistical analysis reveals significant differences (increased Jitter/Shimmer, decreased HNR in ASD) [17].

4.4 Key Datasets and Benchmarks

ABIDE I: The cornerstone data set for neuroimaging-based ASD research, providing aggregated sMRI and rs-fMRI data from 1112 participants across multiple sites**[38-39]**. It is crucial for investigating neural mechanisms, but presents challenges due to site heterogeneity. ADHD-200: A global data set containing rs-fMRI and sMRI data used for the detection of ADHD and, combined with ABIDE, to differentiate ASD from ADHD or to detect comorbidity[39]. Kaggle Autism Image Dataset: A benchmark data set for facial image-based ASD detection, containing `299 prelabeled images [3]. Infant cry data set (Khozaei et al.): A specific data set used for the detection of ASD based on cry analysis, comprising recordings from 62 children (31 ASD, 31 TD) aged 18-54 months [17].

5. Applications of Deep Learning in ASD

This part talks about the different and strong uses of DL in all parts of the analysis of ASD. It starts by looking at how DL models, mainly CNN and RNN, are changing the diagnosis and grouping by analyzing hard neuroimaging data such as MRI and fMRI, as well as behavioral video analysis. After this, the discussion is about the use of DL in predicting the levels of ASD symptoms, mostly through smart regression models. Also, the part brings out the making of new early spotting and testing tools, involving mobile apps, wearable tools, and DL-powered real-time monitoring systems. Lastly, it looks at the new area of multimodal DL and how it can increase accuracy and strength by combining many data types.

5.1 Diagnosis and Classification of ASD

CNN and RNN have shown significant effectiveness in the classification of ASD using neuroimaging data. In fMRI-based classification, Liu et al. proposed MADE-for-ASD, a multi-atlas deep ensemble framework that fuses functional connectivity maps from three different brain atlases: the Automated Anatomical Labeling (AAL), Craddock 200, and Eickhoff-Zilles. MADE-for-ASD employed a stack sparse denoising autoencoder (SSDAE) coupled with an MLP that implemented weighted ensemble voting, reaching an accuracy rate of 75.2% on the large-scale ABIDE I dataset and 96.4% on the high-quality NYU subset. Furthermore, the introduction of demographic characteristics such as age, sex, handedness, and IQ increased the model's robustness[12]. In EEG-based classification, Xu et al. designed a hybrid CNN and LSTM network model to assess Time-Series Maps of Brain Functional Connectivity (TSM-BFC). This approach successfully encoded spatiotemporal dynamics in EEG signals, achieving accuracy rates of 81.08% for resting-state data and 74.55% for task-state data. Importantly, short-distance connectivity between the parietal and occipital regions, as well as long-distance interaction with the temporoparietal junctions, served as key discriminatory characteristics [40].

In addition, facial image analysis using CNNs has emerged as a non-invasive diagnostic tool. Farhat et al. compared the MobileNet and VGG16 classifiers to detect facial phenotypes. Using 5-fold cross-validation, VGG16 (with batch size 2 and Adam optimizer) achieved 99% validation accuracy and 87% test accuracy. The model identified structural features (for example, a broader upper face and wider eyes) with 85% precision and 90% recall, validated on a local Pakistani dataset (85% accuracy) [41].

5.2 Symptom Severity Prediction

Although severity prediction using regression models is less explicit in the provided studies, Xu et al. identified functional connectivity biomarkers correlated with the pathophysiology of ASD. Reduced long-distance connectivity from the right temporoparietal junction to the left posterior temporal lobe and altered short-range parietal-occipital connectivity were associated with the manifestation of symptoms. Although not a direct regression model, these characteristics could underpin future severity-prediction frameworks[40]. Liu et al. used F scores to rank critical brain regions (eg, precuneus, anterior cingulate) associated with ASD, providing a basis for quantitative symptom profiling [12].

5.3 Preliminary identification and evaluation tools

Mobile and Real-Time Applications: Farhat et al. demonstrated an assessment of the VGG16 model in small datasets, highlighting its potential for implementation on mobile devices. The lightweight version, with a batch size of 2, enabled fast inference, demonstrating its suitability for real-time screening tasks in low-resource settings like Pakistan [3]. In their EEG pipeline, Xu et al. leveraged deep convolutional generative adversarial networks (DCGANs) to augment small datasets, thus facilitating early detection despite limited clinical samples. Although wearable technology is not directly discussed, EEG-based approaches [40] imply some compatibility with portable neuroheadsets.

5.4 Multimodal Deep Learning

There are many studies that were integrating diverse data types significantly to enhance diagnostic accuracy, such as Liu et al., who combined fMRI connectivity matrices with demographic data (age, sex, and IQ), improving model generalizability across heterogeneous populations. Furthermore, Xu et al. [40] combined temporal dynamics (using LSTM) and spatial features (using CNN) derived from functional connectivity maps and thus captured complementary aspects of cerebral activity. Furthermore, the facial analysis study carried out by Farhat et al. [41] may be even richer by the addition of behavioral video or acoustic data, a factor that was missing in the present study. The ensemble method built by Liu et al. [12] demonstrated the strongest multimodal data integration by combining anatomical (AAL, Eickhoff-Zilles) and functional (Craddock 200) atlases with phenotypic data to achieve state-of-the-art results.

6. Comparison and Performance Analysis

As illustrated in Table 3, the comparative analysis reveals significant variations in performance across DL approaches for ASD detection, with non-invasive methods demonstrating exceptional results. CNNs remain a popular choice, with Sherkatghanad et al. [42]achieving 70.22% accuracy on ABIDE-I fMRI data, outperforming traditional classifiers such as SVM, KNN, and RF. Tamilvizhi et al. [13] demonstrated the power of multimodal learning by combining Inception v3 with LSTM, achieving 99.29% accuracy in eye tracking images. Other vision-based models, such as DenseNet201 and MobileNet variants, tested by Almars et al. [43], achieved solid accuracy (85–86.5%), with DenseNet201 slightly outperforming the rest. Studies such as those by Liu et al. [36] introduced novel architectures such as MAHGCN to analyze functional brain networks, though accuracy was lower (62.6%),

indicating challenges in graph-based modeling of ASD. Kareem et al. proposed a 1D-CNN that achieved a high accuracy of 99.45% with the Adults data set.

In contrast, structured data produced exceptional results; Rasul et al. [44] reported perfect classification (100%) for children using SVM and linear regression (LR) on UCI datasets, with ANN and ensemble models also performing robustly across children, adults, and combined datasets. Ensemble-based methods have been especially effective; Liu et al. [12] combined the Stacked Sparse Denoising Autoencoder (SSDAE) and MLP for 96.4% accuracy on the New York University (NYU) subset of ABIDE I, while Alotaibi et al. [37] integrated the AE, LSTM, and Deep Belief Network (DBN) to reach 97.79% in babies. Similarly, Jagadesh et al. [18] achieved 96.43% with a hybrid ensemble using LSTM, DBN, and Hybrid Kernel Extreme Learning Machine (HKELM), supported by MobileNetV2 features and a specialized optimizer. Multimodal and hybrid models are emerging as superior, as seen in Rajkumar et al. [45], where the combination of facial image and questionnaire data achieved 92% accuracy.

Image-based models also continue to evolve, with ViT enhanced by squeeze-and-excitation blocks outperforming traditional CNNs, reaching up to 97.77% [11]. Electroencephalogram EEG)- based approaches like those of Xu et al. [40] demonstrated strong results using CNN–LSTM (81.8%) for resting-state EEG, highlighting its ability to model spatial and temporal brain patterns. Göker [19] used the Walsh-Hadamard transform with 1D-CNN to achieve up to 98.77% accuracy depending on the size of the feature, suggesting that optimization in signal preprocessing significantly impacts the results. Meanwhile, Contreras et al. [42] boosted accuracy by 8% using ResNet-50 with SVM, while Chaitanya et al. [38] reached 98.95% with an attention-based symbolic-execution-CNN (SE-CNN). Transformer models, although powerful, showed limitations in large-scale registry data; Dick et al. [15] reported 61.9% accuracy, suggesting that further adaptation is needed for longitudinal health records.

Finally, demographic-sensitive models, as seen in Nogay & Adeli [46], highlighted the relevance of age and gender in the diagnosis of ASD, with CNN classification reaching 85.42%. Overall, while CNNs and MLPs remain foundational, ensemble and attention-based DL models, especially when applied to multimodal data, offer the most promising advances in ASD detection and diagnosis. Figure 9 summarizes the interplay between data modalities, benchmark datasets, and DL model families applied to ASD research from 2020 to 2025.

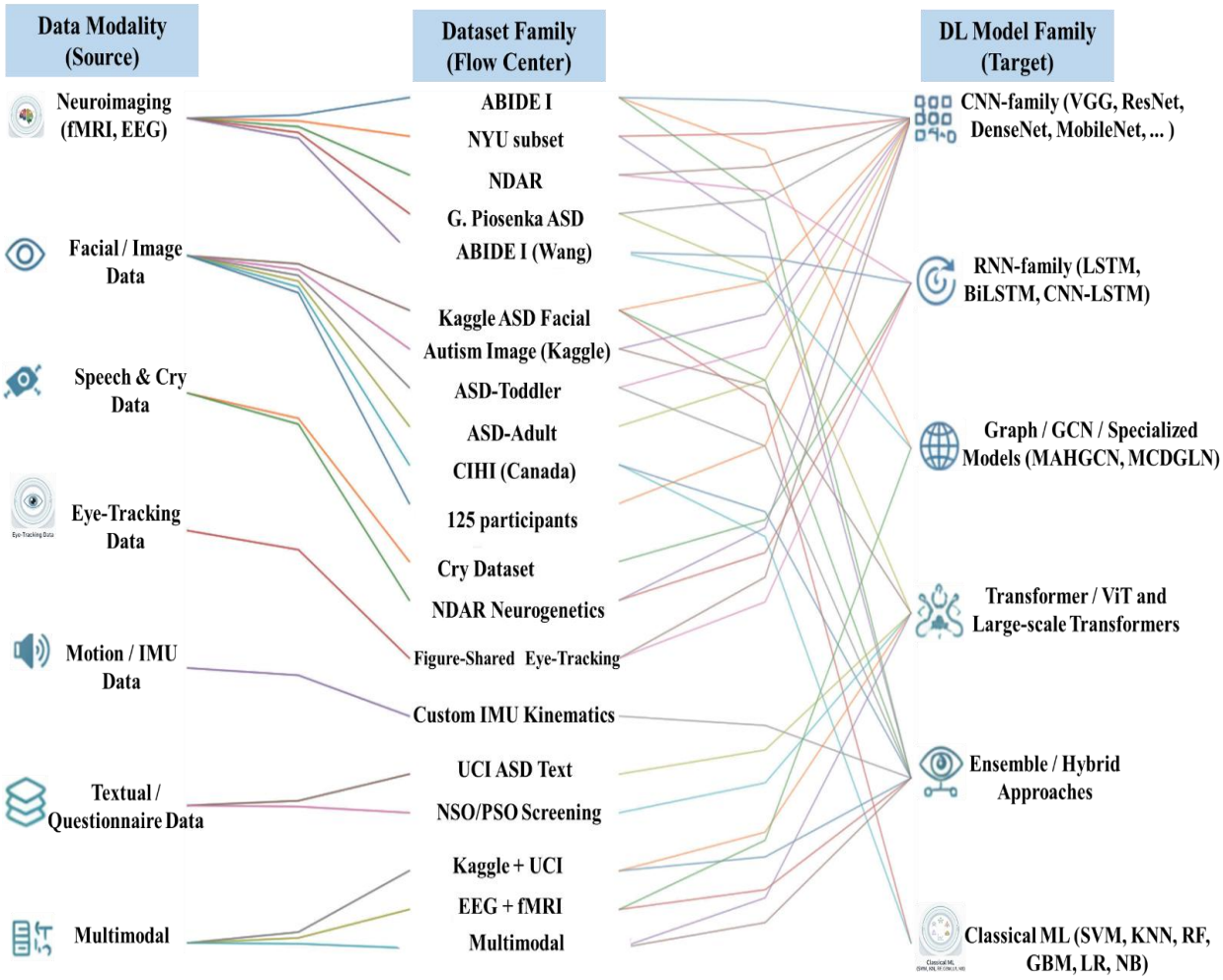

Figure 9. Comparison of Modalities, Datasets, and Deep Learning Model Families in ASD Research (2020–2025).

Table 3. Comparative Study of DL Approaches for Detection of ASD

Reference	Year	Dataset Name	DL Methods	Accuracy % (Best)
Sherkatghanad et al. [42]	2020	ABIDE-I fMRI data	CNN, SVM, KNN, RF	70.22
Tamilvizhi et al. [13]	2023	Figure-Shared Eye-Tracking Image Dataset	Inception v3 + LSTM	99.29
Almars et al. [43]	2023	Images: 3,374; Numerical: 1,100 records	DenseNet201, DensNet169, MobileNet, MobileNetV2, MobileNetV3Small, MobileNetV3Large	86.55
M. Liu et al. [36]	2023	4,410 fMRI scans (ABIDE-I: 499 ASD, 512 HC)	MAHGCN	62.6
Kareem et al. [47]	2023	ASD (Adults, Children, Adolescents)	1D0CNN	99.45
Rasul et al. [44]	2024	UCI: Children, Adults	KNN, SVM, RF, DT, EGBoot, LR, ANN, NB	100
X. Liu et al. [12]	2024	ABIDE I (CPAC pipeline), NYU subset of ABIDE I, proposed subset of ABIDE	SSDAE + MLP (ensemble model)	96.4
Farhat et al. [41]	2024	Kaggle ASD Facial Image Dataset	VGG16 (Batch=2), VGG16 (Batch=16), VGG16 (Batch=32), VGG16 (Batch=64), MobileNet (Batch=64), KNN, RF, Gradient Boost (GB), VGG19	88
Xu et al. [40]	2024	NDAR – Multimodal Developmental Neurogenetics of Females with ASD	CNN–LSTM (Resting EEG), CNN–LSTM (Task EEG), LSTM (Resting EEG), LSTM (Task EEG), VGG11 (Resting EEG), VGG11 (Task EEG)	81.8
Su et al. [14]	2024	Custom IMU-based upper limb kinematics data	MLP (4 layers)	98.62
Nogay & Adeli [48]	2024	ABIDE	CNN (Model 1 - Gender-based classification), CNN (Model 2 - Age-based classification), CNN (Model 3 - Age & Gender-based classification), (AlexNet, GoogLeNet, ResNet18, SqueezeNet (Transfer Learning))	85.42
Contreras et al. [49]	2025	2,536 images (1:1 ASD/TD) from Kaggle	ResNet-50 + SVM	92.67
Wang et al. [46]	2025	ABIDE-I: 1,035 subjects (505 ASD, 530 TD)	MCDGLN	73.3
Alotaibi et al. [37]	2025	ASD-Toddler (1,000) ASD-Adult (680)	Ensemble (AE, LSTM, Deep Belief Network (DBN)), Ensemble (AE, LSTM, DBN), DenseNet169, MobileNet, MobileNetV2, MobileNetV3Small, MobileNetV3Large	97.79

Reference	Year	Dataset Name	DL Methods	Accuracy % (Best)
Sheik Abdullah et al. [16]	2025	ABIDE I: 871 scans (505 ASD, 530 TD)	LSTM with Attention, BiLSTM with Attention, CNN -BiLSTM with Attention	93
Rajkumar et al. [45]	2025	125 participants (facial images + questionnaires)	RF, VGG16 CNN (fine-tuned), RF + VGG16 (combined), RF (severity prediction)	92
Nawghare & Prasad [50]	2025	ASD Image Dataset (Kaggle), ASD Text Dataset (UCI), Multimodal (Kaggle + UCI)	VGG16, ANN, Vision Transformer (ViT), RF, LR, Gradient Boosting Machines (GBM), VGG16 + RF (Hybrid)	95.74
Dick et al. [15]	2025	Newborn Screening Ontario (NSO), Prenatal Screening Ontario (PSO) and the Canadian Institute for Health Information (CIHI)	Extreme Gradient Boosting models and large-scale ensembled Transformer deep learning models	61.9
Ibadi & Lakizadeh [11]	2025	G. Piosenka ASD dataset, NYU subset of ABIDE I, proposed subset of ABIDE	VGG-19, ResNet50, MobileNetV3, EfficientNet-B4, Xception, EfficientNetB0, EfficientNetB1, EfficientB2, MobileNet, ASDvit (ViT + Squeeze-and-Excitation (SE)), CNN–LSTM (Task EEG), LSTM (Resting EEG), LSTM (Task EEG), VGG11 (Resting EEG), VGG11 (Task EEG), VGG16 (Batch=16), VGG16 (Batch=32), VGG16 (Batch=64), MobileNet (Batch=64), KNN, RF, GB, VGG19, SSDAE + MLP (with Ensemble), SSDAE + MLP (with Ensemble)	97.77
Göker [19]	2025	Resting-state EEG dataset (62 subjects, 31 ASD, 31 HC)	Walsh–Hadamard transform (FWHT) + 1D-CNN (16 features), FWHT + 1D-CNN (32 features), FWHT + 1D-CNN (64 features), FWHT + 1D-CNN (128 features)	98.77
Chaitanya R et al. [38]	2025	ABIDE I (NYU): 172 participants	LR, KNN, RF, SE-CNN Attention (Deep)	98.95
Laguna et al. [17]	2025	Cry data set (Khozaei et al.)	SVM (Recurrent CNN)	90.28
Jagadesh et al. [18]	2025	Autism Image Dataset (Kaggle)	Ensemble: LSTM, DBN, and HKELM, MobileNetV2, VGG16 + LR, VGG16 + RF, VGG16 + DT, VGG16 + GB, Multi Kernel SVM, KNN, RF, DT, EGBoot, LR, ANN, NB	100

7. Challenges and Limitations

The application of DL in the detection and intervention faces several significant challenges and limitations, as highlighted in the reviewed literature.

7.1 Data Scarcity and Class Unbalance

A major limitation is the lack of large, high-quality data sets required to train resilient models. Most of the research has to work with small sample sizes because it is hard to recruit subjects from a varied population of ASD and obtain complex multimodal data (e.g., neuroimagery and behavioral recordings). Inadequate data restricts the complexity of the model and increases the risk of overfitting [51-53]. Furthermore, class imbalance is widespread, with ASD cases greatly outweighed by neurotypical controls for most datasets. This imbalance skews the model performance scores and reduces the sensitivity to the correct identification of authentic ASD cases[54]. Techniques such as the Synthetic Minority Oversampling Technique (SMOTE) have been explored for the creation of artificial data[55]; however, its efficacy is largely determined by the quality of the data and can lead to the production of artificial patterns that do not represent true heterogeneity of ASD.

7.2 Lack of standardized datasets

There is no presence of standardized and accepted common ASD datasets for research. The methodology of data collection, diagnostic tools (DSM-5, ADOS, ADI-R), and the markers themselves come from very different studies [52-54], so that the comparison of the performances would be direct, the results reproducible, and algorithms generalizable. Other initiatives, such as ABIDE, are designed for different types of data pooling neuroimaging information[56], but no broader standardization involves behavioral, demographic, and multimodal information. Such a lack of consistency thwarts meta-analyses and clinical translation [51]**[55-56]**.

7.3 Model interpretability and explainability

Complex DL models are "black-box" in character, and this is a major barrier to clinical adoption. Clinicians need to have a comprehensible justification for intervention or diagnosis decisions. Although some research works utilize interpretability tools such as SHapley Additive exPlanations (SHAP) [51] to interpret the mapping of predictions to input features, it is hard to provide clinically interpretable justifications for models' intricate behavior (particularly deep neural networks). Limited interpretability erodes trust and makes it difficult to detect potential biases or incorrect patterns learned by the model [54][51][57]. Alignment of model decisions

with established ASD pathophysiology and clinical knowledge of ASD is essential but often unrealized.

7.4 Ethical and privacy concerns

The gathering and utilization of sensitive information (e.g., neuroimaging, video, physiological signals, extensive medical histories) raises significant ethical and privacy concerns. Obtaining informed consent, especially in children or individuals frequently comorbid with ASD, is challenging [51-55]. Compliance with ethical standards and data protection laws (e.g., HIPAA, GDPR) is crucial but rigorous, with extensive data needs for DL. Data security concerns also exist with breaches and misuse of prediction data. The use of technologies such as robots and wearables in therapy further aggravates privacy threats[58-61]. It calls for rigorous anonymization methods and protected data storage systems, but increases complexity.

7.5 Generalizability between populations.

One of the significant limitations is the lack of generalizability of models when applied to heterogeneous populations. Most models are developed and validated in narrow, often homogeneous, populations (e.g., narrow age ranges such as 5-17 years, various ethnicities, particular geographic locations, or patients with rigorous inclusion/exclusion criteria) [54][61-63]. This limits generalizability to individuals from other demographic groups (e.g., other ethnicities [54]), age brackets (e.g., adults or very young children are underrepresented [56]), genders (because of the male preponderance in the 4:1 prevalence ratio dataset representing populations), or those with differing comorbidities and symptom severities. Other concerns that influence generalizability are cultural diversity in the presentation [58], variation in diagnostic practice, and socioeconomic influences on data quality [52][54][58]. Models that are fine-tuned to smaller feature sets can oversimplify the multifaceted phenotype and fail to capture its entire spectrum[54].

8. Future Directions

Even with the breakthroughs achieved in AI-driven ASD identification, major obstacles remain in translating these into clinical use and scalability. Thus, this section sets forth critical avenues of progress, emphasizing larger multimodal datasets to reflect heterogeneity in ASD; explainable AI (XAI) that can engender trust clinically; real-time telehealth tools for accessible screening; personalized prediction systems that can also predict interventions; and federated learning capable of ensuring data safety by compliance with privacy.

8.1 Need for larger, diverse, multimodal datasets

Larger, more diverse, and multimodal datasets would overcome these stated limitations, typical straits encountered because of small, and often not very generalizable, homogeneous data sets. Although studies based on EEG biomarkers reveal the difficulties in identifying reliable biomarkers in a heterogeneous population of patients with ASD, their datasets are not rich in terms of age, ethnicity, severity, and comorbid conditions. The use of only one data modality (eg, only behavioral questionnaires or only EEG) will never be sufficient. Large-scale multimodal datasets should be prioritized in future efforts. Such data sets would integrate behavioral assessments (ADOS, ADI-R), neuroimaging (EEG, fMRI), genetic information, eye tracking data, and motor patterns in diverse populations (age, gender, ethnicity, geographic location), capturing the heterogeneity of ASD in its widest and most robust modeling possible [55]**[64-66]**.

8.2 Explainable AI (XAI) for clinician trust

Very complex AI/ML models-conceptually 'black boxes'-particularly DL and advanced fuzzy neural nets-have constrained clinical adoption. Models must be interpretable for clinicians to understand why a prediction is made. Research using fuzzy expert systems showed the advantage of interpretable fuzzy rules in severity assessment[64]. Future work should integrate XAI techniques directly into diagnostic tools[67]. This requires the development of inherently explainable models or the use of post hoc explanation methods to identify the factors (e.g., certain EEG band power, eye movement, and questionnaires) that significantly affect the diagnosis of ASD or its severity predictions, improving trust, and allowing auditing by clinicians [64-67].

8.3 Real-time diagnostic tools and telehealth integration

The application of technologies such as IoT sensors, wearables (e.g., devices for electroencephalogram monitoring or movement capture), and user-friendly platforms (such as tablet and smartphone applications) enables continuous, real-time monitoring outside of clinical settings. The combination of such technologies with robotic systems to improve patient participation during assessments and telehealth systems offers significant potential for remote monitoring and screening, especially for regions with restricted access to healthcare[66]. Further questions should focus on developing efficient integrated algorithmic implementations (like EEG processing with energy efficiency prioritization[65]) compatible with edge devices or mobile devices to allow real-time processing in interactive environments such as gamification

activities or robot-aided therapy sessions and ensure smooth deployment for broader reach under telehealth[65-66].

8.4 Personalized ASD prediction and intervention systems

ASD manifests itself uniquely in each individual. Future AI systems must move beyond binary diagnosis toward personalization. This involves:

- Subtyping: Using unsupervised learning and clustering on multimodal data to identify distinct ASD subtypes with shared biological or behavioral markers[55][67].
- Prediction of Severity Trajectories: Using longitudinal data sets and ML methods to predict personal developmental trajectories and the course of severity[55][64].
- Tailored Therapeutic Interventions: Adapting therapeutic protocols -such as changing levels of difficulty in robot-mediated therapy through the introduction of fuzzy logic[64] and making personalized skill development recommendations- according to each patient's expected subtype, severity profiles, and behavior response patterns. The first steps toward this goal are being taken by adopting adaptive models, such as fuzzy cognitive maps (FCMs), within games[55][64][66].

8.5 Federated learning for the use of privacy-preserving data

Data protection is of paramount importance, and compiling sensitive patient data, especially multimodal data such as EEG, genetic data, and video recordings of behavior, into centralized data pools raises significant privacy and legal conundrums. Federated learning (FL), which allows training models on a distributed set of disconnected devices or establishments without requiring data exchange, is an essential future direction. Although not specifically elaborated on in the mentioned publications, the explicit emphasis on medical data privacy [55][67] and the need for large datasets forcefully point to the need for FL. This framework allows for collaborative model training on heterogeneous datasets held by various hospitals and clinics worldwide while maintaining patient privacy and compliance with rules such as HIPAA and GDPR, thereby making it possible to develop robust models without compromise of confidentiality[55][67].

8.6. Alternative Explanations and Limitations

Depending on the promising results, DL models for ASD may be influenced by factors like dataset bias, overfitting, or artifacts in data collection and preprocessing. These issues can lead to inflated performance that does not generalize to diverse populations. To ensure robustness, future research should emphasize cross-dataset validation, benchmarking on standardized

datasets, and transparent reporting. Addressing these limitations is crucial to improving the reproducibility, reliability, and clinical applicability of DL-based ASD diagnostic systems.

9. Conclusions

DL has indeed shown revolutionary potential in powering ASD diagnosis with unprecedented capacity to detect intricate, nonlinear biomarkers in various data modalities. Recent studies have witnessed architectures such as CNNs, Transformers, GCNs, and their ensembles repeatedly report high diagnostic accuracies, highly frequently over 90% for imaging-based methods, and demonstrate tremendous potential in neuroimaging, kinematics, and multimodal methods. Transfer learning increases the input feature space, while attention-based mechanisms further improve feature extraction heterogeneity and system interpretability. This fosters an environment in which DL can not only better deliver diagnostic accuracy but can also be implemented for earlier intervention, personalized treatment, and scalable screening. Realizing this potential requires interdisciplinary high-level collaboration. Closing the loop between algorithmic development and clinical translation requires synergy between computer science, neuroscience, clinical psychology, and ethics. Multidimensional data streams -neuroimaging to health records- require co-designed preprocessing procedures and validation systems with a focus on biological relevance. Interdisciplinary collaborations are required to assemble representative data, avoid algorithmic biases, solve privacy issues, and advocate representatively for fair deployment to diverse populations. To bridge research advances to clinical significance, subsequent studies must solve the translational robustness, which includes

- Preprocessing and standardization of data: to increase reproducibility and generalizability of the model.
- Easy explainability tools (eg, attention heatmaps, feature importance maps) to facilitate clinician trust.
- Strict prospective validation: in real-world environments and demographic populations.
- Development of regulatory frameworks: for clinical uptake, dispelling ethical and practical deployment issues.

As striking as outcomes are in idealized environments, practice requires practice translation to recognize heterogeneity in real-world data and for models to generalize across practice sites. Ultimately, DL can transform ASD diagnosis, but success depends on the application of technological advances within a collaborative, patient-focused system prioritizing clinical validity, equity, and ethical integrity.

References

[1] J. Zeidan, E. Fombonne, J. Scorah, A. Ibrahim, M. S. Durkin, S. Saxena, A. Yusuf, A. Shih, and M. Elsabbagh, "Global prevalence of autism: A systematic review update," Autism Res., vol. 15, no. 5, pp. 778–790, 2022. https://doi.org/10.1002/aur.2696

[2] M. Elsabbagh, G. Divan, Y.-J. Koh, Y. S. Kim, S. Kauchali, C. Marcín, C. Montiel-Nava, V. Patel, C. S. Paula, C. Wang, M. T. Yasamy, and E. Fombonne, "Global prevalence of autism and other pervasive developmental disorders," Autism Res., vol. 5, no. 3, pp. 160–179, 2012. https://doi.org/10.1002/aur.239

[3] S. Timimi, D. Milton, V. Bovell, S. Kapp, and G. Russell, "Deconstructing diagnosis: Four commentaries on a diagnostic tool to assess individuals for autism spectrum disorders," Autonomy (Birmingham, England), vol. 1, p. AR26, 2019. https://pubmed.ncbi.nlm.nih.gov/31396391

[4] V. Dahiya, E. DeLucia, C. G. McDonnell, and A. Scarpa, "A systematic review of technological approaches for autism spectrum disorder assessment in children: Implications for the COVID-19 pandemic," Res. Dev. Disabil., vol. 109, p. 103852, 2021. https://doi.org/10.1016/j.ridd.2021.103852

[5] S. Bhat, U. R. Acharya, H. Adeli, G. M. Bairy, and A. Adeli, "Autism: cause factors, early diagnosis and therapies," Rev. Neurosci., vol. 25, pp. 841–850, 2014. https://doi.org/10.1515/revneuro-2014-0056

[6] Nafea, A.-M. Manar, K. M. A. Alheeti, M. S. I. Alsumaidaie, and M. M. AL-Ani, "A Hybrid Method of 1D-CNN and Machine Learning Algorithms for Breast Cancer Detection," Baghdad Sci. J., vol. 21, pp. 3333–3343, 2024. https://doi.org/10.21123/bsj.2024.9443

[7] S. H. Majeed, A. K. Kareem, A. A. Nafea, and K. M. Ali Alheeti, "Efficient cardiovascular disease classification using supervised machine learning on photoplethysmography data," in IET Conf. Proc. CP906, IET, pp. 431–435, 2024. https://doi.org/10.1049/icp.2025.0121

[8] S. Y. Mohammed, "Enhancing COVID-19 Patients Detection using Deep Transfer Learning Technique Through X-Ray Chest Images," J. Adv. Res. Appl. Sci. Eng. Technol., vol. 32, pp. 290–302, 2023. https://doi.org/10.37934/araset.32.1.290302

[9] M. Khodatars, A. Shoeibi, D. Sadeghi, N. Ghaasemi, M. Jafari, P. Moridian, A. Khadem, R. Alizadehsani, A. Zare, Y. Kong, A. Khosravi, S. Nahavandi, S. Hussain, U. R. Acharya, and M. Berk, "Deep learning for neuroimaging-based diagnosis and rehabilitation of Autism Spectrum Disorder: A review," Comput. Biol. Med., vol. 139, p. 104949, 2021. https://doi.org/10.1016/j.compbiomed.2021.104949

[10] F. Jassam, A. A. Mukhlif, A. A. Nafea, M. A. Tharthar, and A. I. Khudhair, "A Review of Breast Cancer Histological Image Classification: Challenges and Limitations," Iraqi J. Comput. Sci. Math., vol. 6, pp. 1–23, 2025. https://doi.org/10.52866/2788-7421.1351

[11] H. Ibadi and A. Lakizadeh, "ASDvit: Enhancing autism spectrum disorder classification using vision transformer models based on static features of facial images," Intell.-Based Med., vol. 11, p. 100226, 2025. https://doi.org/10.1016/j.ibmed.2025.100226

[12] X. Liu, M. R. Hasan, T. Gedeon, and M. Z. Hossain, "MADE-for-ASD: A multi-atlas deep ensemble network for diagnosing Autism Spectrum Disorder," Comput. Biol. Med., vol. 182, p. 109083, 2024. https://doi.org/10.1016/j.compbiomed.2024.109083

[13] T. Thanarajan, Y. Alotaibi, S. Rajendran, and K. Nagappan, "Eye-Tracking Based Autism Spectrum Disorder Diagnosis Using Chaotic Butterfly Optimization with Deep Learning Model," Comput. Mater. Contin., vol. 76, pp. 1995–2013, 2023. https://doi.org/10.32604/cmc.2023.039644

[14] W.-C. Su, J. Mutersbaugh, W.-L. Huang, A. Bhat, and A. Gandjbakhche, "Using deep learning to classify developmental differences in reaching and placing movements in children with and without autism spectrum disorder," Sci. Rep., vol. 14, p. 30283, 2024. https://doi.org/10.1038/s41598-024-81652-z

[15] K. Dick, P. Moore, L. J. Filliter, T. Boni, S. J. Moore, S. MacEachern, A. J. S. Moore, S. Ahmed, F. Samiee-Zafarghandi, and J. R. Green, "Transformer-based deep learning ensemble framework predicts autism spectrum disorder using health administrative and birth registry data," Sci. Rep., vol. 15, p. 11816, 2025. https://doi.org/10.1038/s41598-025-90216-8

[16] S. Abdullah, K. V, S. Geetha, and U. Mishra, "Leveraging deep learning for enhanced diagnosis of autism spectrum disorder using resting-state functional magnetic resonance imaging and clinical data," Results in Engineering, vol. 25, p. 104444, 2025. https://doi.org/10.1016/j.rineng.2025.104444

[17] M. Laguna, S. Pusil, A. L. Paltrinieri, and S. Orlandi, "Automatic Cry Analysis: Deep Learning for Screening of Autism Spectrum Disorder in Early Childhood," J. Autism Dev. Disord., pp. 1–7, 2025. https://doi.org/10.1007/s10803-025-06811-1

[18] J. Balasubramani, S. Rajendran, M. Zakariah, and A. Alnuaim, "Ensemble of Deep Learning with Crested Porcupine Optimizer Based Autism Spectrum Disorder Detection Using Facial Images," Comput. Mater. Contin., vol. 83, pp. 2793–2807, 2025. https://doi.org/10.32604/cmc.2025.062266

[19] H. Göker, "1D-convolutional neural network and fast Walsh–Hadamard transform approach for diagnosing autism spectrum disorder," Neural Comput. Appl., vol. 37, pp. 13039–13057, 2025. https://doi.org/10.1007/s00521-025-11208-3

[20] S. S. Rajagopalan, Y. Zhang, A. Yahia, and K. Tammimies, "Machine Learning Prediction of Autism Spectrum Disorder From a Minimal Set of Medical and Background Information," JAMA Netw. Open, vol. 7, p. e2429229, 2024. https://doi.org/10.1001/jamanetworkopen.2024.29229

[21] T. Wang, H. Ma, H. Ge, Y. Sun, T. T.-O. Kwok, X. Liu, Y. Wang, W. K. W. Lau, and W. Zhang, "the use of gamified interventions to enhance social interaction and

communication among people with autism spectrum disorder: A systematic review and meta-analysis," Int. J. Nurs. Stud., vol. 165, p. 105037, 2025. https://doi.org/10.1016/j.ijnurstu.2025.105037

[22] L. dos Santos, I. I. Barreto, A. C. F. da Silva, J. F. B. Soriano, J. de L. S. Castro, L. S. Tristão, and W. M. Bernardo, "Behavioral therapies for the treatment of autism spectrum disorder: A systematic review," Clinics, vol. 80, p. 100566, 2025. https://doi.org/10.1016/j.clinsp.2024.100566

[23] M. D. Gangayah, D. Zhao, E. J. Y. Liew, N. A. M. Nor, T. Paramasivam, Y. Y. Lee, N. I. A. Hasan, and S. Shaharuddin, "Accelerating autism spectrum disorder care: A rapid review of data science applications in diagnosis and intervention," Asian J. Psychiatr., vol. 108, p. 104498, 2025. https://doi.org/10.1016/j.ajp.2025.104498

[24] S. Q. Abd Al-Rahman and S. A. Jassim, "Using Deep Learning Neural Networks to Recognize and Authenticate the Identity of the Speaker," in 2024 21st Int. Multi-Conf. on Systems, Signals & Devices (SSD), IEEE, pp. 185-192, 2024. https://doi.org/10.1109/SSD61670.2024.10548534

[25] M. M. Taye, "Understanding of machine learning with deep learning: architectures, workflow, applications and future directions," Computers, vol. 12, p. 91, 2023. https://doi.org/10.3390/computers12050091

[26] S. Cong and Y. Zhou, "A review of convolutional neural network architectures and their optimizations," Artif. Intell. Rev., vol. 56, pp. 1905–1969, 2023. https://doi.org/10.1007/s10462-022-10213-5

[27] R. Qamar and B. A. Zardari, "Artificial neural networks: An overview," Mesopotamian J. Comput. Sci., vol. 2023, pp. 124–133, 2023. https://doi.org/10.58496/MJCSC/2023/015

[28] J. Yuan, Q. Wu, J. Zhou, S. Yu, X. Xin, J. Liu, and X. Cui, "A review of deep learning models for food flavor data analysis," J. Future Foods, vol. 6, no. 4, pp. 533–544, 2026. https://doi.org/10.1016/j.jfutfo.2025.01.001

[29] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," IEEE Trans. Neural Netw. Learn. Syst., vol. 33, pp. 6999–7019, 2021. https://doi.org/10.1109/TNNLS.2021.3084827

[30] R. Asghar, M. Quercio, L. Sabino, A. Mahrouch, and F. R. Fulginei, "A novel dual-stream attention-based hybrid network for solar power forecasting," IEEE Access, 2025. https://doi.org/10.1109/ACCESS.2025.3555810

[31] M. S. Meghana, D. Abhijith, S. Aysha, and P. K. Kollu, "Sentiment analysis on Amazon product reviews using LSTM and naïve Bayes," in 2023 7th Int. Conf. Comput. Methodol. Commun. (ICCMC), IEEE, pp. 626–631, 2023. https://doi.org/10.1109/ICCMC56507.2023.10084052

[32] M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning," Decision Analytics J., vol. 3, p. 100071, 2022. https://doi.org/10.1016/j.dajour.2022.100071

[33] S. Wettewa, L. Hou, and G. Zhang, "Graph Neural Networks for building and civil infrastructure operation and maintenance enhancement," Adv. Eng. Inform., vol. 62, p. 102868, 2024. https://doi.org/10.1016/j.aei.2024.102868

[34] Nawaz, S. S. Khan, and A. Ahmad, "Ensemble of autoencoders for anomaly detection in biomedical data: A narrative review," IEEE Access, pp. 1–1, 2024. https://doi.org/10.1109/ACCESS.2024.3360691

[35] D. Mienye and T. G. Swart, "Deep autoencoder neural networks: A comprehensive review and new perspectives," Arch. Comput. Methods Eng., p. 110254, 2025. https://doi.org/10.1007/s11831-025-10260-5

[36] M. Liu, J. Zhang, Y. Wang, Y. Zhou, F. Xie, Q. Guo, F. Shi, H. Zhang, Q. Wang, and D. Shen, "A common spectrum underlying brain disorders across lifespan revealed by deep learning on brain networks," iScience, vol. 26, p. 108244, 2023. https://doi.org/10.1016/j.isci.2023.108244

[37] S. S. Alotaibi, T. A. Alghamdi, and R. Alharthi, "Two-tier nature inspired optimization-driven ensemble of deep learning models for effective autism spectrum disorder diagnosis in disabled persons," Sci. Rep., vol. 15, p. 10059, 2025. https://doi.org/10.1038/s41598-025-93802-y

[38] S. H. Sanjai, M. Arjun, S. Harshavardhan, P. Yugander, and M. Jagannath, "Autism Spectrum Disorder Detection using attention-based CNN and ML classifiers," Procedia Comput. Sci., vol. 258, pp. 4216–4227, 2025. https://doi.org/10.1016/j.procs.2025.01.776

[39] R. Thakur, D. Malhotra, and M. Mengi, "Ensembler_fMRI: An Intelligent Approach for the Early Prediction of Autism Disorder," Procedia Comput. Sci., vol. 259, pp. 1863–1873, 2025. https://doi.org/10.1016/j.procs.2025.04.142

[40] Y. Xu, Z. Yu, Y. Li, Y. Liu, Y. Li, and Y. Wang, "Autism spectrum disorder diagnosis with EEG signals using time series maps of brain functional connectivity and a combined CNN–LSTM model," Comput. Methods Programs Biomed., vol. 250, p. 108196, 2024. https://doi.org/10.1016/j.cmpb.2024.108196

[41] T. Farhat, S. Akram, Z. Ali, A. Ahmad, and A. Jaffar, "Facial image-based autism detection: A comparative study of deep neural network classifiers," Comput. Mater. Contin., vol. 78, pp. 105–126, 2024. https://doi.org/10.32604/cmc.2023.045022

[42] Z. Sherkatghanad, M. Akhondzadeh, S. Salari, M. Zomorodi-Moghadam, M. Abdar, U. R. Acharya, R. Khosrowabadi, and V. Salari, "Automated detection of autism spectrum disorder using a convolutional neural network," Front. Neurosci., vol. 13, p. 1325, 2020. https://doi.org/10.3389/fnins.2019.01325

[43] M. Almars, M. Badawy, and M. A. Elhosseini, "ASD²-TL∗ GTO: Autism spectrum disorders detection via transfer learning with gorilla troops optimizer framework," Heliyon, vol. 9, p. e21530, 2023. https://doi.org/10.1016/j.heliyon.2023.e21530

[44] R. A. Rasul, P. Saha, D. Bala, S. M. R. U. Karim, M. I. Abdullah, and B. Saha, "An evaluation of machine learning approaches for early diagnosis of autism spectrum disorder," Healthcare Analytics, vol. 5, p. 100293, 2024. https://doi.org/10.1016/j.health.2023.100293

[45] S. C. Rajkumar, S. Cirillo, D. Yuvasinsi, and L. Solimando, "A hybrid approach combining images and questionnaires for early detection and severity assessment of autism spectrum disorder," Image Vis. Comput., vol. 160, p. 105547, 2025. https://doi.org/10.1016/j.imavis.2025.105547

[46] P. Wang, X. Wen, L. Yi, Y. Guo, J. Li, Y. Hao, R. Cao, C. Gao, and R. Cao, "MCDGLN: Masked Connection-based Dynamic Graph Learning Network for Autism Spectrum Disorder," Brain Res. Bull., vol. 224, p. 111290, 2025. https://doi.org/10.1016/j.brainresbull.2025.111290

[47] K. Kareem, M. M. Al-ani, and A. A. Nafea, "Detection of Autism Spectrum Disorder Using a 1-Dimensional Convolutional Neural Network," Baghdad Sci. J., vol. 20, pp. 1182–1193, 2023. https://doi.org/10.21123/bsj.2023.8564

[48] H. S. Nogay and H. Adeli, "Multiple classification of brain MRI autism spectrum disorder by age and gender using deep learning," J. Med. Syst., vol. 48, p. 15, 2024. https://doi.org/10.1007/s10916-023-02032-0

[49] R. C. Contreras, M. S. Viana, V. J. S. Bernardino, F. L. dos Santos, Ö. Toygar, and R. C. Guido, "A multi-filter deep transfer learning framework for image-based autism spectrum disorder detection," Sci. Rep., vol. 15, p. 14253, 2025. https://doi.org/10.1038/s41598-025-97708-7

[50] P. Nawghare and J. Prasad, "Hybrid CNN and random forest model with late fusion for detection of autism spectrum disorder in toddlers," MethodsX, vol. 14, p. 103278, 2025. https://doi.org/10.1016/j.mex.2025.103278

[51] W.-J. Ocampo-Pazos and C. S. González-González, "Agile framework for systematic reviews of literature in formative research," in 2023 XIII Int. Conf. Virtual Campus (JICV), IEEE, pp. 1–4, 2023. https://doi.org/10.1109/JICV59748.2023.10565691

[52] S. D. Kevelson, R. Elmaghraby, F. Patel, H. Brown, M. Gorenstein, J. Bain, Z. M. Grinspan, E. Pedapati, J. Veenstra-VanderWeele, and P. Vandana, "Novel therapeutics in autism spectrum disorder," Neurotherapeutics, vol. 23, no. 1, p. e00857, 2026. https://doi.org/10.1016/j.neurot.2026.e00857

[53] Lord, S. Risi, L. Lambrecht, E. H. Cook Jr., B. L. Leventhal, P. C. DiLavore, A. Pickles, and M. Rutter, "The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism," J.

Autism Dev. Disord., vol. 30, pp. 205–223, 2000. https://doi.org/10.1023/A:1005592401947

[54] S. Levy, M. Duda, N. Haber, and D. P. Wall, "Sparsifying machine learning models identify stable subsets of predictive features for behavioral detection of autism," Mol. Autism, vol. 8, pp. 1–17, 2017. https://doi.org/10.1186/s13229-017-0122-1

[55] S. Pandey and S. Sharma, "ML Algorithms that have been Utilised to Classify Neuro-Developmental Disorders: A Review," in 2023 6th Int. Conf. Inf. Syst. Comput. Netw. (ISCON), IEEE, pp. 1–5, 2023. https://doi.org/10.1109/ISCON57409.2023.10134730

[56] M.-S. Cahart, A. Amad, S. B. Draper, R. G. Lowry, L. Marino, C. Carey, C. E. Ginestet, M. S. Smith, and S. C. R. Williams, "The effect of learning to drum on behavior and brain function in autistic adolescents," Proc. Natl. Acad. Sci. U.S.A., vol. 119, no. 23, p. e2106244119, 2022. https://doi.org/10.1073/pnas.2106244119

[57] J. A. Barnes, C. H. Park, A. Howard, and M. Jeon, "Child-robot interaction in a musical dance game: An exploratory comparison study between typically developing children and children with autism," Int. J. Hum.-Comput. Interact., vol. 37, no. 3, pp. 249–266, 2021. https://doi.org/10.1080/10447318.2020.1801224

[58] H. Feng, M. H. Mahoor, and F. Dino, "A music-therapy robotic platform for children with autism: A pilot study," Front. Robot. AI, vol. 9, p. 855819, 2022. https://doi.org/10.3389/frobt.2022.855819

[59] S. M. Nguyen, N. Collot-Lavenne, C. Lohr, S. Guillon, P. Tula, A. Paez, M. Bouaida, A. Anin, and S. El Qacemi, "An implementation of an imitation game with ASD children to learn nursery rhymes," arXiv preprint arXiv:2004.05886, 2020. https://doi.org/10.48550/arXiv.2004.05886

[60] F. L. Cibrian, M. Madrigal, M. Avelais, and M. Tentori, "Supporting coordination of children with ASD using neurological music therapy: A pilot randomized control trial comparing an elastic touch-display with tambourines," Research in Developmental Disabilities, vol. 106, p. 103741, 2020. https://doi.org/10.1016/j.ridd.2020.103741

[61] M. I. Ahmed, B. Spooner, J. Isherwood, M. Lane, E. Orrock, and A. Dennison, "A systematic review of the barriers to the implementation of artificial intelligence in healthcare," Cureus, vol. 15, no. 10, 2023. https://doi.org/10.7759/cureus.38263

[62] K. Mössler, W. Schmid, J. Aßmus, L. Fusar-Poli, and C. Gold, "Attunement in music therapy for young children with autism: revisiting qualities of relationship as mechanisms of change," J. Autism Dev. Disord., vol. 50, no. 11, pp. 3921–3934, 2020. https://doi.org/10.1007/s10803-020-04445-z

[63] E. Hernandez Ruiz and B. B. Braden, "Improving a parent coaching model of music interventions for young autistic children," J. Music Ther., vol. 58, no. 3, pp. 278–309, 2021. https://doi.org/10.1093/jmt/thab002

[64] S. Rani and A. Alphy, "Fuzzy based approaches for autism spectrum disorder detection: A review," in Int. Conf. Commun., Secur. Artif. Intell. (ICCSAI), IEEE, pp. 353-357, 2023. https://doi.org/10.1109/ICCSAI58591.2023.10384403

[65] K. Lalli and M. Senbagavalli, "Identification of biomarker for autism spectrum disorder using EEG: A review," in Proc. Int. Conf. Adv. Comput. (ICAC), pp. 1-6, 2023. https://doi.org/10.1109/ICAC57847.2023.10123786

[66] K.-F. Kollias, P. Sarigiannidis, C. K. Syriopoulou-Delli, and G. F. Fragulis, "Implementation of robots in autism spectrum disorder research: Diagnosis and emotion recognition and expression," in Proc. 12th Int. Conf. Mod. Circuits Syst. Technol. (MOCAST), IEEE, pp. 1-4, 2023. https://doi.org/10.1109/MOCAST58375.2023.10188140

[67] W. J. Ocampo-Pazos and C. S. González-González, "Machine learning as a support tool for early diagnosis in children with autism spectrum disorder: A systematic review of the literature," in Proc. XIII Int. Conf. Virtual Campus (JICV), IEEE, pp. 1-4, 2023. https://doi.org/10.1109/JICV60470.2023.10386262

مناهج التعلم العميق لاضطراب طيف التوحد: مراجعة شاملة

سميح عبد الغفور جاسم[1،*]، ايثم خيري كريم[2]، احمد عادل نافع[3]

1. قسم علوم الحاسوب، كلية العلوم، جامعة المعارف، الأنبار، العراق.
2. قسم تربية هيت، المديرية العامة للتربية في الأنبار، وزارة التربية، هيت، الأنبار، العراق.
3. قسم الذكاء الاصطناعي، كلية علوم الحاسوب وتكنولوجيا المعلومات، جامعة الأنبار، الرمادي، العراق.

المستخلص

اضطراب طيف التوحد (ASD) هو اضطراب نمو عصبي معقد ذو مسببات معقدة تُمثل عقبات أمام التشخيص المبكر والموثوق. أحدثت التطورات الحديثة في مجال التعلم العميق (DL) ثورة في أبحاث اضطراب طيف التوحد، إذ أتاحت التعرف التلقائي على البصمات الدقيقة وغير الخطية من مصادر متعددة الوسائط، غنية بالبيانات، وغير متجانسة. في هذا العمل، نُقدم أول دراسة مقارنة منهجية لثلاثة وعشرين نموذجًا حديثًا للتعلم العميق للكشف عن اضطراب طيف التوحد، تغطي مجموعات بيانات متعددة الوسائط من تصوير الوجه، والتصوير العصبي (الرنين المغناطيسي الوظيفي/تخطيط كهربية الدماغ)، وتتبع حركة العين، والأنماط الحركية، والسجلات الصحية الإلكترونية، وذلك من خلال هياكل دمج مبتكرة. تُحقق التصاميم المعمارية الحديثة، مثل محولات الرؤية (ViTs)، وأطر التعلم العميق الهجينة، والنماذج المُعززة بالانتباه، واستراتيجيات المجموعات، دقة تشخيصية عالية باستمرار، تتجاوز بشكل ملحوظ قدرات أساليب التعلم الآلي التقليدية. ثانيًا، يُسهم تحسين تعلم النقل والدمج متعدد الوسائط، إلى جانب أساليب التفسير (مثل GradCAM وخرائط الانتباه الحرارية)، في جعل السمات قابلة للتمثيل وتعزيز تعميم النماذج. على الرغم من التقدم الهائل المُحرز في هذا المجال، إلا أن تبني التطبيقات العملية لا يزال يُعيقه تباين مجموعات البيانات، ونقص التمثيل الديموغرافي والتنوع بين مجموعات التحقق، والاختلافات في البيئات السريرية. سيتطلب هذا اعتماد خطوط أنابيب موحدة، وتحققًا صارمًا بين السكان، وأطر ذكاء اصطناعي تفسيرية مُصممة لضمان الشفافية والفائدة السريرية. باختصار، يعتمد تحقيق القدرة التحويلية للتعلم العميق في تشخيص اضطراب طيف التوحد على شراكات متعددة التخصصات مستمرة بين علماء الحاسوب والأطباء وعلماء الأعصاب لتقديم حلول جديدة دقيقة وشاملة (أي يجب أن تكون ذات صلة سريرية)، وملتزمة بالمبادئ الأخلاقية.

الكلمات المفتاحية : اضطراب طيف التوحد، التعلم العميق، محولات الرؤية (Vision Transformers)، التصوير العصبي ، الذكاء الاصطناعي القابل للتفسير، التشخيص السريري.