WILEY | INTERNATIONAL JOURNAL OF DEVELOPMENTAL NEUROSCIENCE ISDN

RESEARCH ARTICLE

Virtual Reality Based Adaptive Game for Enhancing Joint Attention in Children With Autism

P. Valarmathi[1] | A. Packialatha[2]

[1]Department of Computer Science and Engineering, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, Tamilnadu, India | [2]Department of Computer Science and Engineering, Vels Institute of Science & Technology & Advanced Studies, Chennai, Tamilnadu, India

Correspondence: P. Valarmathi (nilapandy85@gmail.com)

Received: 15 March 2025 | **Revised:** 23 February 2026 | **Accepted:** 28 February 2026

Keywords: autism spectrum disorder | Bi-LSTM | improved DCNN | improved entropy | improved tanh normalization

ABSTRACT

Atypical brain development is a characteristic of autism spectrum disorder (ASD), a neurodevelopmental disorder that frequently shows up as deficits in joint attention, a crucial social-communication ability including gaze sharing and gaze following. Children with ASD frequently exhibit atypical gaze behaviours that hinder effective coordination of attention with others. Although recent studies have explored virtual reality (VR) to address these aspects, the existing systems rely primarily on performance-based learning and offer limited adaptability to individual behavioural patterns. To overcome these limitations, this study proposes a novel virtual reality and adaptive game-based system (VRAGS) designed to assess and enhance attention in children with ASD, thereby supporting improved learning capacity. The proposed framework begins with data acquisition through the 3D game environment. The collected data are preprocessed using an improved tanh normalization, which effectively suppresses noise while preserving informative variations for learning. Then, the feature extraction process takes place, in which the raw data features, statistical features and improved entropy features are derived. The improved entropy is computed using information gain, prioritizing features that maximally reduce uncertainty with respect to attention states, thereby enhancing discriminative power. Finally, the attention prediction is done via the hybrid classification (HC) model that integrates an Adam-optimized deep convolutional neural network (AODCNN) and a bidirectional long short-term memory (Bi-LSTM) network. This architecture collectively captures hierarchical spatial patterns and temporal dependencies within the data. The integration of the improved tanh activation and Adam optimization in the AODCNN model ensures stable convergence and robust feature learning. Moreover, the experimental evaluation demonstrates that the proposed model significantly outperforms existing methods, achieving an accuracy of 0.933, an F-measure of 0.925 and a negative predictive value (NPV) of 0.936. These findings demonstrate the potential of the suggested framework as a customized, adaptable tool for therapeutic intervention and support its efficacy in precisely predicting attention levels in children with ASD.

1 | Introduction

Asperger's condition, autism and infant disintegrative illness are among the complex neurodevelopmental brain disorders together referred to as autism spectrum disorder (ASD). As the name 'spectrum' implies, these illnesses include different symptoms and degrees of severity (Vakadkar et al. 2021; Amat et al. 2021; Xiao et al. 2020). This neurological disease is characterized by repeated behavioural patterns and social communication difficulties (Vakadkar et al. 2021; Goel et al. 2020; Ke et al. 2022; Johnston et al. 2020). ASD is becoming more common, while frequency varies by nation (Banire et al. 2021). Depending on how severe the present symptoms are, ASDs can range from very mild ASD to severe ASD (Jonsdottir et al. 2020; Zhao et al. 2018; Rusli et al. 2020). Early infancy, youth or adulthood may see the signs. As many diagnoses show a variety of typical neurodevelopment symptoms, it was particularly challenging to identify moderate forms of ASD early (Goel et al. 2020; Moreno et al. 2017; Zheng et al. 2016).

Automated facial landmark identification has led to applications that can be deployed in several fields. Examples of situations in which facial expressions act as intellectual systems are detecting pain, diagnosing infant syndromes, detecting driver fatigue and detecting engagement (Eni et al. 2020; Glennon et al. 2020; Puli and Kushki 2019). Through the application of face pattern recognition, advanced information technology that employs AI models has assisted in the early diagnosis of ASD (Herrero and Lorenzo 2020; Yaneva et al. 2020; Sheldrick et al. 2019). Artificial intelligence (AI) and machine learning (ML) algorithms were deployed in several realistic appliances to aid in resolving societal problems. AI (Herrero and Lorenzo 2020; Hadri and Bouramoul 2023) is used in all aspects of healthcare to aid medical professionals in dealing with disorders like autism. The extraction of ASD features has gained a lot of consideration since it could be utilized to differentiate between people without or with ASD. Deep learning (DL) methods (Hadri and Bouramoul 2023; Mazumdar et al. 2021), in particular, convolutional neural networks (CNNs) (Banire et al. 2021; Herrero and Lorenzo 2020) and recurrent neural networks (RNNs) as well as the bidirectional long short-term memory (Bi-LSTM) models, were used or suggested to diagnose autism in children.

Relevant to the current study, it is crucial to investigate the variables influencing the physical activity (PA) of children with ASD as well as whether they are given adequate opportunity to do so (Bejarano-Martín et al. 2020; Charpentier et al. 2018). To increase the effectiveness of therapies intended to promote the active life/well-being of children with ASD (Frolli et al. 2022), it will also be crucial to determine the contributing elements to PA. Also, researchers in the field of ASD have tried several methods to enhance attention evaluation for successful learning outcomes. The handling of specific data by the brain and behaviour during attention entails disregarding other information that is monitored (Garcia-Garcia et al. 2022; ElMoazen et al. 2020; Lai et al. 2020). It was sometimes referred to as the participants' behavioural (Banire et al. 2021) or cognitive (Herrero and Lorenzo 2020) involvement in the learning activity. Children with ASD struggled with focus since they were often diverted from their learning activities (Lu et al. 2022; de Mira Gobbo et al. 2021; Alcaniz Raya et al. 2020). For children with ASD, traditional intervention techniques including pivotal response training (PRT), physical prompting and fading treatments have demonstrated encouraging results; nevertheless, they are frequently quite expensive and resource-intensive. These approaches usually involve a significant cost burden for clinical settings, extensive hours of one-on-one therapy administered by qualified interventionists and little flexibility for therapists to implement changes in the task environment (Jyoti and Lahiri 2019). Furthermore, even though eye-tracking technology has shown promise in the diagnosis and evaluation of ASD, current applications encounter a number of difficulties, such as poor data accuracy, vulnerability to environmental influences and challenges in successfully integrating multimodal information to enhance diagnostic precision and reliability (Chen et al. 2025). These shortcomings spurred the development of a revolutionary virtual reality and adaptive game-based system (VRAGS) designed to assess and improve attention in children with ASD, fostering improved learning capacity. The main contributions of the proposed work are as follows:

- Proposes a new VRAGS model for ASD children to enhance their learning ability, in which an improved tanh normalization-based preprocessing is followed.
- Contributing to the extraction of improved entropy-based features, along with raw data and statistical features.
- Deploys a hybrid classification (HC) that integrates an Adam-optimized deep convolutional neural network (AODCNN) and a Bi-LSTM for the attention prediction of children with autism during gaming.

Section 2 reviews the extant VRAGS approaches. Section 3 describes the proposed VRAGS model for ASD children, and Section 4 describes data acquisition and preprocessing. Section 5 provides briefs about features, and Section 6 explains hybrid classifiers. Sections 7 and 8 explain outcomes and the conclusion.

2 | Literature Review

2.1 | Related Works

To develop and work on the emotional and social abilities of pupils with ASDs, Herrero and Lorenzo (2020) suggested the implementation and design of a head-mounted display (HMD) interactive virtual space system in 2019. They had chosen two groups for this. In the first group, they used immersive virtual reality (IVR) as a didactic tool to recreate the contexts of socializing (a classroom and a kindergarten) over the course of 10 sessions. The subsequent group served as the control and received no treatment throughout the intervention period. The degrees of improvement and adaptation showed that IVR in the proposed format was compatible with the tactile preference and visuospatial abilities of the ASD youngsters taking part in this analysis.

Rakić et al. (2020) presented a strategy for classifying ASD patients against control participants using structural and functional magnetic resonance images (MRI) data. Grey matter volume similarities between cortical parcels and functional connectivity patterns between brain areas were employed as features for operational and anatomical assessment pipelines, respectively. Stacks of unsupervised multilayer perceptrons (MLPs) were combined with supervised MLPs to build the classifying system.

Goel et al. (2020) suggested that a modified grasshopper optimization algorithm (MGOA) is capable of diagnosing ASD at all stages of life. The grasshopper optimization algorithm (GOA) was an algorithm that drew inspiration from nature and had the ability to successfully explore and use the search space. They aimed to address the limitations of the conventional GOA through their research, leading to an early identification of the illness.

Two strategies were presented in 2021 by Banire et al. (2021) for a face-based attention detection model. The first method translated geometric properties using a support vector machine (SVM) classifier, and the second method converted time-domain spatial information into two-dimensional (2D) spatial visuals using a CNN method. The outcomes demonstrated that the SVM classifier's geometric feature transformation beat the CNN method. Additionally, attention detection performed better in lower attention activities than in higher attention tasks and was more generalizable in normally developing children than in ASD groups.

In 2021, Mazumdar et al. (2021) presented a scheme depending upon the joint usage of ML and eye tracing data. More precisely, characteristics like the existence of objects and fixation in the core of an image were retrieved from the image content and examined behaviour. A classifier based on ML was trained using those characteristics. The results demonstrated that the factors taken into consideration enable the distinction between children with ASD and those who are growing normally.

Alsaade and Alzahrani (2022) created a method that used face recognition and social media to identify people with ASDs. DL approaches may be utilized to recognize these landmarks, but they need a reliable system for extracting and creating the right patterns of facial data. Using a simple web application built on a DL system, namely a CNN using transfer learning (TL) and the Flask system, aids groups and clinicians in empirically diagnosing autism using face traits. The trained schemes employed as classifiers were Visual Geometry Group Network (VGG19), Xception and NASNET Mobile.

Kuttala et al. (2022) presented a one-class paradigm for describing healthy people. In this one-class concept, any person with ASD/attention deficit hyperactivity disorder (ADHD) was seen as an oddity. The adopted model was a dense generative adversarial network (DGAN) design with a self-attention module. T1-weighted longitudinal structural magnetic resonance images (sMRI) were the input modalities utilized by the system. In addition, they exclusively trained the system with longitudinal data, which consisted of 2 scans of each person over time, as opposed to the cross-sectional data used in more conventional methods (a single scan per participant).

Hadri and Bouramoul (2023) concentrated on the development of the intellectual system for the support and companionship of ASD children. The suggested system was built using chatbots, which were conversational assistants that used AI. The contribution assisted autistic youngsters in understanding their circumstances by responding to their inquiries and offering guidance and suggestions. They felt that giving a non-human friend may assist the child and others around them in understanding the illness and avoiding depression.

In 2013, Cai et al. (2013) designed and developed an innovative virtual dolphinarium as a potential intervention for children with autism. Through the virtual dolphin interaction programme, children can take on the role of dolphin trainers and engage with virtual dolphins at the poolside using hand gestures to learn nonverbal communication. A dedicated immersive room with a 320° curved screen and an expensive five-panel projection system was used to increase engagement through the use of immersive visualization and gesture-based interaction. The study also reported a pilot investigation aimed at establishing a trial protocol for autism screening and evaluating participants' readiness to engage in the virtual dolphin interaction environment.

In 2025, Chen et al. (2025) have combined eye tracking with virtual reality settings to examine children with ASD's gaze patterns. To increase impartiality and accuracy, it suggests a new diagnostic framework called the Multiscale Search Enhanced Gaze Network (MSEG-Net). To forecast gaze direction, the gaze estimating model combines head as well as eye movement data. It uses multiscale convolutional kernels to derive hierarchical eye movement characteristics and improves precision through binocular fusion. In order to preserve crucial data, the model streamlines network connections. Long-range temporal dependencies within eye movements are modelled via a lightweight Transformer architecture. Nevertheless, these techniques frequently require particular hardware and environmental circumstances, which restricts their practicality.

In 2025, Nawghare and Prasad (2025) have developed a hybrid framework that integrates VGG16, a pretrained deep CNN, with a Random Forest (RF) classifier. This approach exploits VGG16 capability for extracting high-level image features. The results demonstrate that combining feature-rich DL representations with RF classification significantly improves performance, highlighting the suitability of this hybrid strategy for ASD classification tasks and enhancing prediction reliability. However, the developed model exhibits limited flexibility in incorporating additional DL models for more comprehensive classification. Table 1 shows the review of existing works.

2.2 | Research Gaps

Children with ASD struggle with focus since they quickly become diverted from their learning activities. Various schemes have been introduced for generalizing the attention tasks for children affected with ASD. Existing studies have demonstrated significant advancements in attention assessment and ASD-related learning systems; however, several critical limitations remain. The CNN model in (Kuttala et al. 2022) integrates with transfer learning and the Flash framework, achieving high accuracy but requiring a large volume of training data, while the DGAN/DAGAN frameworks and ensemble classifiers (Rakić et al. 2020) reported improved accuracy, yet they highlighted the need for further performance enhancement. Face-based attention recognition (Banire et al. 2021) achieved a high F1-score but lacked real-time validation, and the HMD-IVR learning system (Herrero and Lorenzo 2020) supported ASD learning but faced challenges in effectively tracking attention interruptions in group settings. Optimization-based and hybrid approaches such as MGOA (Goel et al. 2020) and RM3ASD (Mazumdar et al. 2021) maximized classification metrics but struggled with imbalanced datasets and failed to predict autism severity, while NN-based methods (Hadri and Bouramoul 2023) showed higher accuracy yet were unsuitable for severe autism cases. Additionally, the VGG16 + RF model (Nawghare and Prasad 2025) effectively combined deep feature extraction and RF for classification but lacked flexibility for incorporating temporal learning models. To address these challenges, the proposed

TABLE 1 | Reviews of extant works on ASD.

Authors	Methodologies	Feature	Limitation
Alsaade and Alzahrani (2022)	CNN with TL and the Flash framework	The highest accuracy was obtained	The main disadvantage of the model is it requires a huge amount of training data.
Banire et al. (2021)	Face-based attention recognition model	The F1-score was greater	The proposed model needs to be applied in a real-time environment
Herrero and Lorenzo (2020)	HMD IVR system	It is deployed as a learning tool for ASD children.	The interruptions among the large number of children could be tracked better.
Goel et al. (2020)	MGOA	Maximized accuracy, sensitivity and specificity	It is ineffective in handling imbalanced datasets
Kuttala et al. (2022)	DGAN and Dense Attentive Generative Adversarial Network (DAGAN) frameworks	Maximal accuracy was attained	Improve the performance of the proposed work
Rakić et al. (2020)	Ensemble classifiers	Higher accuracy, sensitivity, specificity and precision	Enhance the accuracy ratings
Mazumdar et al. (2021)	RM3ASD	Accuracy, specificity, sensitivity and precision were maximized	It does not predict the severity of autism
Hadri and Bouramoul (2023)	NN	Greater accuracy was obtained	This system was not conducive to severe autism cases
Chen et al. (2025)	MSEG-Net	The model effectively captured long-range temporal dependencies within eye movements.	The model requires particular environmental conditions and hardware configuration for implementing in the real-time environment.
Nawghare and Prasad (2025)	VGG16 + RF	The model effectively integrates feature-rich deep learning representations with RF classification significantly improved performance	The model demonstrates limited flexibility in incorporating additional deep learning models for effective classification.

VRAGS framework integrates an immersive VR environment for realistic data acquisition, employs improved tanh normalization to suppress noise, utilizes information gain–based entropy features for enhanced discriminative power and adopts a hybrid Adam-optimized DCNN with Bi-LSTM architecture to collectively capture spatial and temporal attention patterns, thereby enabling robust and scalable attention prediction for children with ASD.

3 | Proposed VRAGS Model for ASD Children

The incapacity of children with ASD to use Joint Attention (JA) to detect cues from social partners is a contributing factor in their social communication difficulties. Therapist-mediated JA therapies require a lot of work. Robot-assisted skill training requires specific expertise and is expensive. On the other hand, computer-based JA skill training platforms are less expensive, provide the creator more flexibility, but they do not allow for personalization. As a result, this study presents a VRAGS model to forecast children's gaming interest. This study examines two gaming systems: fishing and 3-point tracking. In this case, a 3-point tracking game is selected for gaze accessing and gaze following cooperative attention. The three-point gazing concept, such as avatar-toy-avatar, is advocated for children. The fishing game is intended to develop cognitive abilities and teach colours. Data acquisition, preprocessing, feature extraction and attention prediction constitute several crucial steps in the suggested method. This methodical approach enables the therapist to take the appropriate steps to enhance the children's capacity for learning. Real virtual reality data, including name, participant age, gender, age with a disability, minimum time, interaction, eye movement, speed, level in-game, number of attempts and number of levels, are first gathered in the virtual reality environment during the data gathering phase. The acquired data are next subjected to the improved Tanh normalization procedure to ensure that it is standardized and prepared for precise subsequent analysis. The pertinent elements such as statistical features, raw data and improved entropy-based features are then extracted from the preprocessed data in order to assess the data's uncertainty or unpredictability, which may provide insight into how the child interacted with the VR system. A hybrid approach for attention prediction, which combines the AODCNN and Bi-LSTM classifiers, receives the acquired feature set. The final prediction is determined by integrating the results from both models using the harmonic mean and produce the final prediction outcome as inactive, observer, avatar, 2-point gaze and 3-point gaze.

The adopted VRAGS model for ASD children is shown in Figure 1.

4 | Data Acquisition and Improved Preprocessing Phase

4.1 | Data Acquisition

The initial phase is data acquisition, which includes the process of collecting real virtual reality data such as name, participant age, gender, disability age, minimum time, interaction, eye movement, speed, level in-game, number of attempts and number of levels achieved. The acquired input data are represented as *ID*.

4.2 | Improved Preprocessing: Improved Tanh Normalization

Due to movement variations and uneven engagement, the raw data obtained during the data collecting phase using a 3D gaming system shows significant variability. Therefore, preprocessing is necessary to guarantee that clean, well-structured data are used for subsequent feature extraction and prediction. Specifically, the preprocessing intends to improve the data quality so that it can be analysed in a superior way. The tanh normalization is found to be reliable and very effective (Latha and Thangasamy 2011). This normalization method yields results that are very comparable to those of the Z-score normalization. The improved tanh normalization compresses outliers smoothly, preventing them from disproportionately affecting the learning model. By scaling values to a bounded range, the normalization reduces variability between features, as well as making the training more stable and efficient. The distribution in the converted domain has an average of 0.5 and an SD of around 0.01 due to the characteristics of the tanh distribution. The distribution of the normalized actual scores is controlled by a constant of 0.01 in tanh normalization (Latha and Thangasamy 2011). The tanh-estimators are modelled in Equation (1), in which, μ_{gh} signifies mean and σ_{gh} refers to SD, which are modelled as in Equations (2) and (3).

$$S'_k = \frac{1}{2}\left\{\tanh\left(0.01\left(S_k - \mu_{gh}\right)/\sigma_{gh}\right) + 1\right\} \tag{1}$$

$$\mu_{gh} = \frac{ID_1 + ID_2 +, \ldots ID_n}{n} \tag{2}$$

$$\sigma_{gh} = \frac{\sqrt{\sum\left(ID_i - \mu\right)^2}}{n} \tag{3}$$

However, this technique is not responsive to outliers. Therefore, we propose a new tanh estimator as shown in Equation (4).

$$S'_k = \frac{1}{2}\left\{\tanh\left(0.01\left(S_k - \mu_{mh}\right)/\sigma_{mh}\right) + 1\right\} \tag{4}$$

$$\mu_{mh} = \left[\frac{\frac{1}{n}\sum_{i=1}^{n}\left|ID_i - M(ID)\right| + \frac{n}{\frac{1}{ID_1} + \frac{1}{ID_2} \cdots \frac{1}{ID_n}}}{2}\right] \tag{5}$$

In Equation (5), $M(ID)$ refers to the mean of ID.

$$\sigma_{mh} = \frac{\sqrt{\sum\left(ID_i - \mu_{ma}\right)^2}}{n} \tag{6}$$

$$\mu_{ma} = median\left(\left|ID_i - median\left(ID_i\right)\right|\right) \tag{7}$$

The improved pre-processed data are represented as ID^{ITH}.

5 | Feature Extraction

Extraction of features intends to decrease the number of features by creating novel features from existing ones (by subsequently eliminating the original ones). The common data in the original feature is then summarized by this novel, smaller set of features. From this ID^{ITH}, statistical features and improved entropy features are drawn.

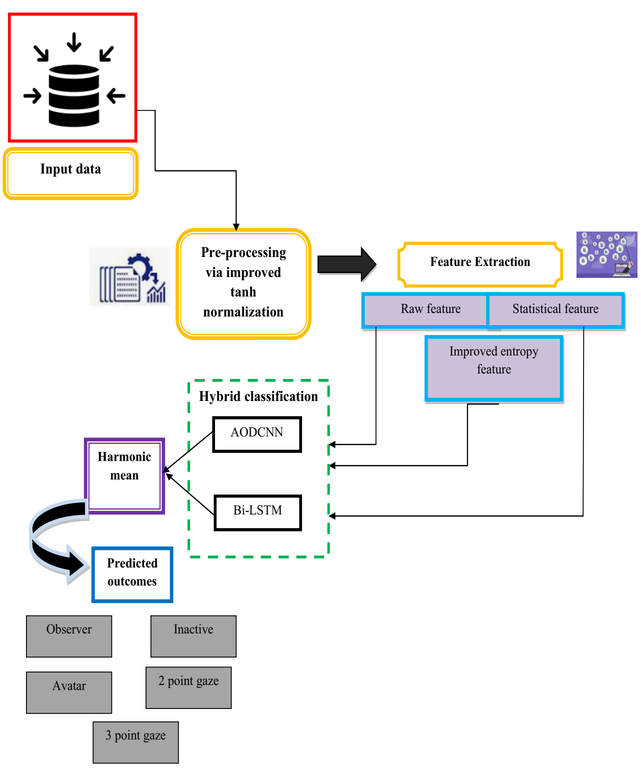

FIGURE 1 | Diagrammatic demonstration of the VRAGS model for ASD children.

5.1 | Raw Data Features

The raw data includes features such as name, participant age, gender, disability age, inactive, observer, avatar, 2-point gaze, 3-point gaze, minimum time, interaction, eye movement and speed, level in-game, number of attempts and number of levels achieved. The raw data features are represented as F^R.

5.2 | Improved Entropy-Based Features

The entropy measures the quantity of information required to portray the outcome of an arbitrary variable y for a known value of a new arbitrary variable x (Wikimedia Foundation, n.d.-a). The entropy of y conditioned on x is indicated by $h(y|x)$ is defined in Equation (8). Here, x refers to the features and y refers to the labels (Wikimedia Foundation, n.d.-a).

$$h(y|x) = -\sum_{X \in \chi, Y \in \Upsilon} P(X,Y) \log \frac{P(X,Y)}{P(X)} \tag{8}$$

Here, χ and Υ indicate the supporting sets of y and x. However, due to the limited problem resolving the capability of entropy features, we model a new entropy model based on information gain IG as shown in Equation (9). However, the conventional conditional entropy is prone to noise and less useful for identifying significant patterns from complicated VR-based interaction data since it primarily evaluates data uncertainty and is unable to assess a feature's relevance to the intended goal. By emphasizing features that lower uncertainty and contribute most to prediction, the improved entropy variant that incorporates information gain gets around this limitation. This leads to more robust, discriminative and informative extraction of features for attention analysis as well as higher prediction accuracy. The information gain of the data are computed, and then, the overall mean of the computed information gain is evaluated to get the final entropy features. Equation (9), μ refers to the mean of IG.

$$Ih(y|x) = -\sum_{X \in \chi, Y \in \Upsilon} P(X,Y) \log \frac{P(X,Y)}{P(X)} + \mu(IG(ID)) \tag{9}$$

5.3 | Statistical Features

The derived statistical features include SD, median and mean.

Mean (Wikimedia Foundation, n.d.-c): The procedure wherein the summation of every value is divided by the summation of values is known as the mean value.

$$\overline{Q} = \frac{1}{k} \sum_{m=1}^{k} Q_m \tag{10}$$

Equation (10), Q points out the observed value, k points out the count of values and $\overline{Q}$ point out the sample mean.

SD: It computes the quantity of variation or a group of distribution values. A small SD (Wikimedia Foundation, n.d.-b) shows that the resultants are close to the mean, whilst a great SD shows that the resultants are distributed over a wide range. Equation (11) provides the modelling of SD, where, σ refers to the standard deviation (SD).

$$\sigma = \sqrt{\frac{1}{k-1} \sum_{m=1}^{k} \left(Q_m - \overline{Q}\right)^2} \tag{11}$$

Median (Statistics Canada, n.d.): The median of a dataset is the number at which 50% of the data points are less than or equivalent to the median and 50% are greater or equivalent to the median.

The SD, median and mean features are together implied as SF. The raw data F^R, statistical feature SF and improved entropy features Ih are wholly represented by FS.

6 | Hybrid Model: AODCNN and Bi-LSTM for Attention Prediction of Children

A hybrid prediction model is the process of using two separate classifiers together to determine the prediction results. Compared to other classification methods, hybrid models improve prediction accuracy. Here, combine the AODCNN and Bi-LSTM models to create an HC model. Specifically, the AODCNN is highly effective at extracting spatial features from structured data, with its convolutional layers identifying patterns, edges and hierarchical structures that are critical for distinguishing attention levels. Incorporating the Adam optimizer enables adaptive learning rates, which accelerates convergence, reduces training time and enhances stability during backpropagation, allowing the network to efficiently learn complex patterns without becoming trapped in local minima. In contrast, the Bi-LSTM is designed to capture sequential dependencies and further improves this capability by processing data in both forward and backward directions, enabling the model to understand contextual information from both past and future time steps. By combining these two models, their complementary strengths provide a richer and more comprehensive representation of the data, resulting in improved classification performance. The process is as follows: Initially, the extracted features will be subjected to the models individually. The results from the models are combined by taking the harmonic mean, which determines the final prediction outcome.

6.1 | Adam Optimized-DCNN

The features FS are given to AODCNN (Gu et al. 2018) with 3 layers: 'convolutional layer, pooling layer and fully-connected layers'. A variety of convolution kernels are typically included in the convolution layer to help with the computation of diverse feature maps. Particularly in the feature map, each neuron is connected to the neurons in the previous layer. The feature values are designed via Equation (12) at the position (a,b) in l^{th} layer of c^{th} feature map. The weight vector and the bias of c^{th} filter related to the layer are implied by W_c^l and B_c^l. The patched input is signified as $FS_{a,b}^l$ at the centre position (a,b) at l^{th} layer.

$$Z_{a,b,c}^l = W_c^{l\,T} FS_{a,b}^l + B_c^l \tag{12}$$

In multilayer networks, the activation function creates non-linearity that helps find non-linear properties. The activation value $\left(act^{l}_{a,b,c}\right)$ related toward the convolution features $Z^{l}_{a,b,c}$ is designed by Equation (13).

$$act^{l}_{a,b,c} = act\left(Z^{l}_{a,b,c}\right) \tag{13}$$

In a conventional DCNN, the sigmoid activation function is used. Here, in the improved Adam Optimized DCNN, we deploy improved tanh as an activation function for enhanced training performances. Moreover, in a conventional DCNN, an optimizer such as an RMS probe was deployed to enhance the performance. With the improved modelling, this work deploys an 'Adam optimizer' instead of conventional optimizers so that better prediction accuracy can be achieved.

The conventional tanh activation is modelled as in Equation (14).

$$f(o) = \left(\frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}\right) \tag{14}$$

The proposed tanh activation is modelled as in Equation (15).

$$f_t(o) = \left(\frac{(e^{x} - e^{-x}) * t_1}{(e^{x} + e^{-x}) * t_2}\right) \tag{15}$$

The terms t_1 and t_2 in Equation (16) are modelled as in Equations (17) and (15).

$$t_1 = \alpha_1\left(2^{u/Ln(2)} - 1\right) \tag{16}$$

$$t_1 = \alpha_2\left(2^{u/Ln(2)} - 1\right) \tag{17}$$

The terms α_1 and α_2 in Equations (16) and (17) are modelled as in Equations (18) and (19), in which, υ_n refers to the input domain, a refers to a constant as given in Equation (20), n refers to the count of subdivisions of the sigmoid function.

$$\alpha_1 = \log \sum_{n=1}^{N} \exp\{\upsilon_n - a\} \tag{18}$$

$$\alpha_2 = \log \sum_{n=1}^{N} \exp\left\{\frac{\upsilon_n + a}{2}\right\} \tag{19}$$

$$a = \max\left(\upsilon_n\right) \tag{20}$$

6.2 | Bi-LSTM

On the other hand, the features FS are given to Bi-LSTM. Bi-LSTM includes 'input, output and forget gates in forward and backward directions' (Zhou et al. 2020). The neurons in Bi-LSTM are evaluated as exposed in Equations (21–25), and the input layer is represented as $\left(FS_1, FS_2, ..FS_3\right)$.

$$\ell_t = K\left(o_{\ell} + w_{\ell} \times \left[\mathrm{H}_{t-1}, FS_t\right]\right) \tag{21}$$

$$\kappa_t = K\left(o_{\kappa} + w_{\kappa} \times \left[\mathrm{H}_{t-1}, FS_t\right]\right) \tag{22}$$

$$\rho_t = K\left(o_{\rho} + w_{\rho} \times \left[c_t, \mathrm{H}_{t-1}, FS_t\right]\right) \tag{23}$$

$$c_t = \kappa_t \times c_{t-1} + \ell_t \times \tan\mathrm{H}\left(w_c \times \left[\mathrm{H}_{t-1}, FS_t\right]\right) \tag{24}$$

$$\mathrm{H}_t = \rho_t \times \tan\mathrm{H}\left(c_t\right) \tag{25}$$

Here, t denotes time; ℓ_t, κ_t and ρ_t indicates the input of the input and output gate, K = sigmoid function, FS_t represent the input neuron; o_l, o_m, and o_n indicates offset vector; H_t represent the output of the hidden layer; w_l, w_m and w_n indicates weight matrices of input and output cells.

Then, the harmonic mean is taken for the outputs obtained from Adam-otimized DCNN and Bi-LSTM to find if the output is inactive, observer, avatar, 2-point gazeor 3-point gaze. Figure 2 shows the representation of the proposed HC.

7 | Results and Discussion

7.1 | Simulation Set-Up

The HC (AODCNN + Bi-LSTM) method for the VRAGS model for ASD children was done in Python with a 64-bit system equipped with an 11th Gen Intel(R) Core (TM) i5-1135G7@2.40GHz (2.42 GHz) processor running at 2.40 GHz and 16 GB of RAM. For analysis, real data were collected that included the details of 60 participants with the details on their name, participant age, gender, disability age, minimum time, interaction, eye movement, speed, level in-game, number of attempts and number of levels achieved.

7.2 | Performance Analysis

For comparison, the conventional classifiers such as CNN, RNN, LSTM, RF, SVM, SVM + RF + Naïve Bayes (NB) + Linear Regression (LR) + K-nearest neighbour (KNN) (Vakadkar et al. 2021), CNN + TL (Alsaade and Alzahrani 2022), VGG16 + RF (Nawghare and Prasad 2025) and MSEG-Net (Chen et al. 2025) models were considered. These conventional classifiers were simulated over the presented HC (AODCNN + Bi-LSTM) classifier. The analysis was performed for varied rates of training data, and in addition, ablation as well as statistical studies were also conducted to prove the efficiency of the HC (AODCNN + Bi-LSTM) method over CNN, RNN, LSTM, RF, SVM, SVM + RF + NB + LR+ KNN (Vakadkar et al. 2021), CNN + TL (Alsaade and Alzahrani 2022), VGG16 + RF (Nawghare and Prasad 2025) and MSEG-Net (Chen et al. 2025) models.

7.2.1 | Accuracy

Accuracy is computed using a set of observations and their actual results. Equation (26), which includes the formula for calculated accuracy, where TP indicates true positive, TN represents true negative, FP referred to as false positive and FN defines false negative.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{26}$$

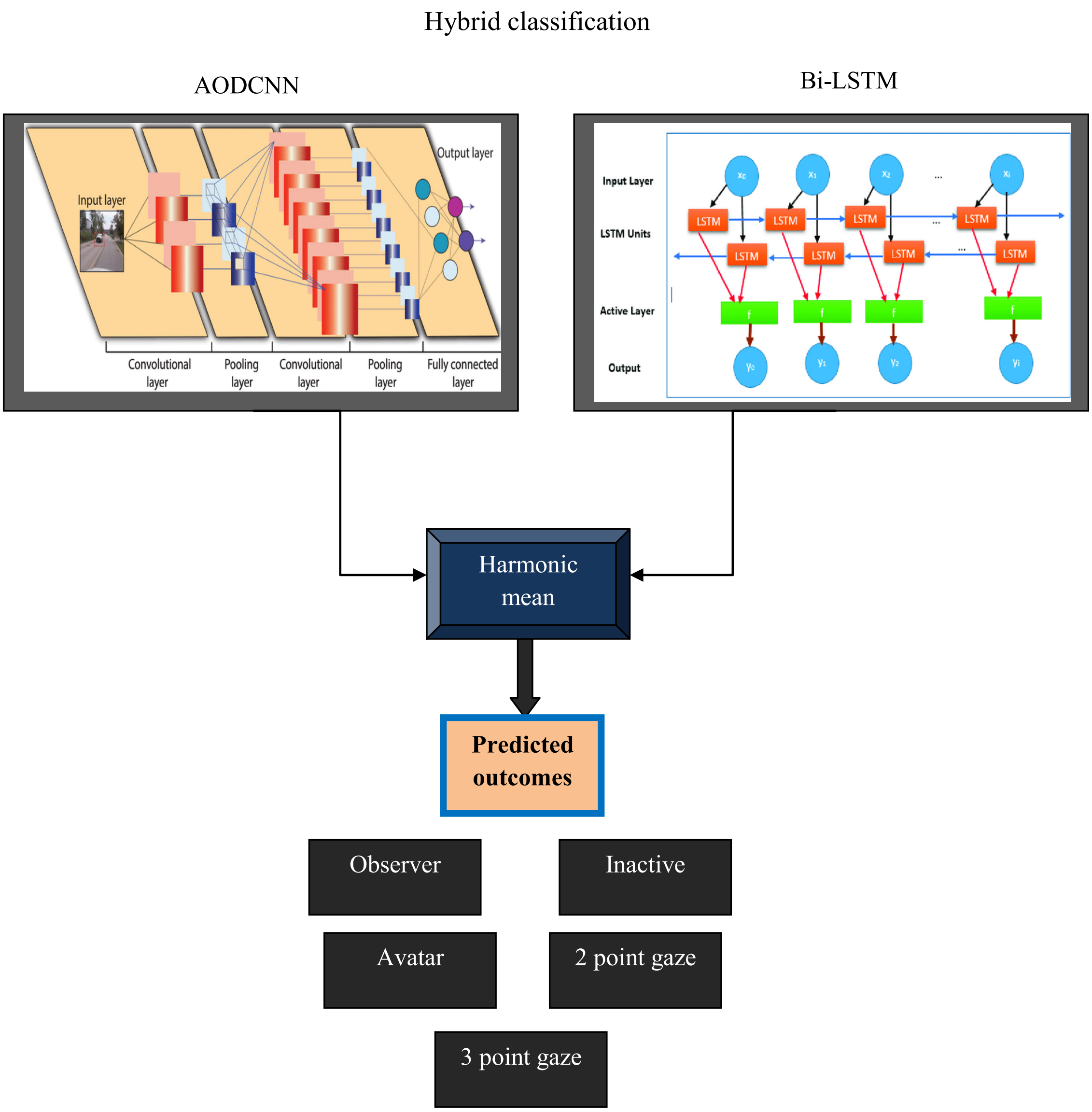

FIGURE 2 | Representation of the proposed hybrid classification.

7.2.2 | Precision

Precision is determined by the percentage of correctly classified instances. This is stated in Equation (27).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{27}$$

7.2.3 | Sensitivity

Sensitivity, which only considers positive true values, describes the probability of an optimal outcome. The formula of sensitivity is mentioned in Equation (28)

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100 \tag{28}$$

7.2.4 | Specificity

Equation (29) illustrates that specificity is defined as the probability of only having real negative values.

$$\text{Specificity} = \frac{TN}{FP + TN} \times 100 \tag{29}$$

7.2.5 | False Negative Rate (FNR)

It is defined as the potential for a test to miss a true positive, as given by Equation (30).

$$FNR = \quad FN(FN + TP) \tag{30}$$

7.2.6 | False Positive Rate (FPR)

The false positive rate, which can be found in Equation (31) and is defined as the ratio of the total number of true negative occurrences (independent of categorization) to the number of negative events that are incorrectly categorized as positive (false positive).

$$\text{FPR} = \frac{FP}{FP + TN} \tag{31}$$

7.2.7 | False Discover Rate (FDR)

FDR stands for false positive ratio, which is the predicted ratio of all positive classifiers (declines of the nil) to all false positive classifications. The FDR is described in Equation (32).

$$\mathrm{FDR} = \frac{FP}{TP + FP} \tag{32}$$

7.2.8 | F-Measure

To evaluate the accuracy results, the F-measure is employed. Equation (33) is used to represent it.

$$\mathrm{F-measure} = \frac{TP}{1/2(FN + FP) + TP} \tag{33}$$

7.2.9 | Matthews Correlation Coefficient (MCC)

The Mathews correlation coefficient is used to compute the Pearson's coefficient, which is utilized to ascertain the product–moment correlation between the actual and expected data using a contingency matrix method. The issue of imbalanced datasets is unrelated to this alternate metric. The calculation is indicated in Equation (34).

$$\mathrm{MCC} = \frac{TP.\,TN - FP.\,FN}{\sqrt{(TP + FP).\,(FN + TP).\,(TN + FP).\,(TN) + FN}} \tag{34}$$

7.2.10 | Negative Predictive Value (NPV)

The negative predictive value is expressed in Equation (35).

$$\mathrm{NPV} = \frac{TN}{FN + TN} \tag{35}$$

7.3 | Dataset Description

We chose a 3-point tracking game for co-operative attention of gaze accessing and gaze following. Children are encouraged to use the 3-point gazing idea, such as avatar-toy-avatar (Frazier et al. 2014). The Fishing game is used to teach colours and build cognitive skills. The dataset size is 60×9. The total number of features is 60×13. Table 2 shows the training and testing data analysis.

7.3.1 | 3-Point Tracking Game

In this case, individuals are instructed to follow 3-point gaze tracking (i.e., avatar-toy-avatar) (Garcia-Garcia et al. 2010) under the supervision of a therapist. By focusing on an item or subject that piques their interest, children can actively engage in or supervise activities with partners while employing three-point, 'coordinated' joint attention. Before they can participate in coordinated 3-point joint attention, the participant must first master some two-point gaze shifts in which they glance just once at an avatar and an object (e.g., avatar-toy or toy-avatar) (Gardiner and Iarocci 2018). As they start to effectively link people and objects, children with appropriate development are often anticipated to move their attention in a 3-point coordinated pattern. The evaluation of the outcome is based on the gaze points.

TABLE 2 | Analysis of training and testing data.

Percentage of training	Training data	Testing data
60	36×13	24×13
70	42×13	18×13
80	48×13	12×13
90	54×13	6×13

7.3.2 | Game Module

The 30-min session is broken up into 10 min of gameplay, 5 min of break time and 10 min of gameplay, followed by another five-minute break. It is possible to establish gaze sharing without adding to the stress that people with ASD may feel when forced to make direct, prolonged stares. According to the study, the minimal time for gaze focus is 200 ms (Geurts et al. 2009). The avatar will initiate the subsequent prompt by turning its eyes toward a toy object when a stare persists for more than 200 ms. The transition from one state to the next requires a 30-s wait for a stare to be detected on the avatar's eye area. The kids could become distracted and lose enthusiasm for the game if there is a longer wait period. The system played audio cues after 30 s if participants still did not stare at the avatar's eye region. A bell-ringing audio cue lasting 3 s was played. After 2 min, if no visual contact was made, the game was over. Participants are encouraged to follow the avatar's gaze with the aid of VR. The avatar's gaze prompt is determined by five gaze points: left, right, up, down and centre. The Avatar—Toy—Avatar rule was implemented (Gioia et al. 2000). The kids first look at the avatar. Then the toy emerges from the screen. Finally, the avatar gazes. Except for the central gaze point, the toy may be seen from any of the avatar's four gaze points. As a result, the avatar's sight changes depending on the toy that appears. The toy will vanish once the children achieve 2-point vision. They then return to avatar gazing. If the children complete 3-point gaze tracking, they will get visual rewards. The eye movement points acquired by this controller were transferred to the device, where they were logged alongside the timestamp. This will be repeated several times to enhance autistic children's joint attention. Table 3 depicts the five different states of the 3-point tracking game. Figure 3 shows the avatar's varied eye gaze configurations in (a) straight direction, (b) left direction with the toy and (c) right direction with the toy.

7.3.3 | Fishing Game

The next game is fishing, which may be played using the same strategy as the 3-point tracking game. A 4 weeks + 1 month is the overall duration of the evaluation period. A lesson, consisting of 10 min of gameplay, a 5-min break and then another 10 min of gameplay, is scheduled every 30 min. There are four lessons per week, with a 5-min of pause between them. The virtual reality learning environment (VLE) will accommodate 15 individuals with autism. The participants learn about colour theory with the use of the colourful 3D animated fish. It aids in the development of children with autism's cognitive abilities. Figure 4 shows an example of a fishing game output screen with various colours.

TABLE 3 | Five different states of the 3-point tracking game.

1. Inactive—The child is not engaged in anything. Example- The child's gaze is not concentrated on anything specific.
2. Observer—The child is watching but not participating in the game. Example—The child may be seeing what the avatar performs but is not getting involved.
3. Avatar—The child is occupied by the avatar. Example—Avatar may make eye contact with the child, and the child may react.
4. 2-point gaze—The child look from avatar to toy but do not return to an avatar or vice versa. A strong focus shift is necessary here. Example—The child looks at a toy, then at the avatar, or vice versa. The child failed to complete the third transition.
5. 3-point gaze—3-point attention shift from avatar to toy to avatar and vice versa. This must be a noticeable shift in focus. Example—The child looks at the avatar, then at the toy and then back at the avatar.

7.3.4 | Game Module

In the game module, the 3D sea environment is created by using a 3D graphics application called Autodesk Maya (Ionescu 2012; Kuschner et al. 2015). The participants are urged to watch the animation by the use of vibrant colours. Every class is divided into two halves. The video will take a while to fully understand the concepts. By employing VLE, it will immediately be stored in the main memory, and the principle is simple to grasp (Lane et al. 2014; Leigh and Morris 2020; Maechler et al. 2021). They have learnt a total of 10 colours: red, green, blue, yellow, orange, pink, black, white, brown and purple. Ten distinct coloured fish emerge from the dynamic three-dimensional sea cave scene. In the centre of the moving 3D marine scene is a rectangular frame. When the fish emerges from the cave and reaches the rectangle frame, the name of the colour appears on the screen. The audio is programmed to speak the colour name. The name will be repeated twice. For each fish, the timer is set for 20s. To forecast the gaze location, a rectangular frame is established. It will aid in making eye contact with the screen. If the participant maintains good eye contact, their gaze point will be in the rectangular frame. A buzzer will sound if their eye contact is weak. Participants are encouraged to make eye contact with the assistance of a therapist. They picked up the principles rapidly, and their cognitive ability has greatly increased (Manjur et al. 2022; McCulloch et al. 2019).

FIGURE 3 | Avatar's varied eye gaze configurations are shown in (a) Straight direction, (b) Left direction with toy and (c) Right direction with toy.

FIGURE 4 | Example of fishing game output screen with various colours.

TABLE 4 | Three different states of fishing game.

1. Inactive—The boy is not doing anything and just staring at the camera. Example—The gaze of a child is not fixated on a single object.
2. Observer—The child observes but does not participate in the action. Example—The child may be observing but is not actively interested in what the object is doing.
3. Engaged—The child is focused on the object. Example—The child is looking at the object and responds to it.

7.4 | Result Module

Table 4 depicts the three stages of the child during the activity. In the passive state, the child's eyes are not focused on the screen. The child is staring at the screen but not into the rectangular frame in the observer state. He is staring at something on the screen. In the engaged state, the child is completely absorbed in the activity depicted in the rectangular frame. Their eye contact is good, and he is listening and responding.

Due to ethical restrictions and privacy protection of child participants imposed by the institution, photography was not permitted. So, the experimental environment illustrated using a synthetically generated visualization is shown in Figure 5 to showcase the VR set-up, to aid understanding the environment.

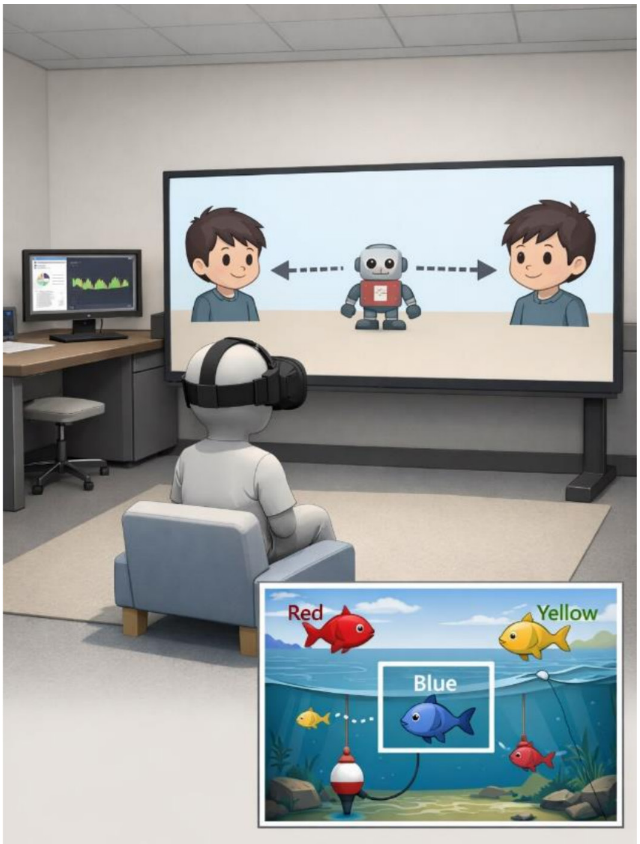

FIGURE 5 | Synthetically generated visualization of the experimental data acquisition environment.

7.5 | Performance Analysis

The evaluation of the HC (AODCNN + Bi-LSTM) over CNN, RNN, LSTM, RF, SVM, SVM + RF + NB + LR+ KNN (Vakadkar et al. 2021), CNN + TL (Alsaade and Alzahrani 2022), VGG16+RF (Nawghare and Prasad 2025) and MSEG-Net (Chen et al. 2025) is exposed in Figures 6–8 for the real-time dataset that included the details of individuals on name, participant age, gender, disability age, minimum time, interaction, eye movement, speed, level in-game, number of attempts and number of levels. The outputs are obtained on whether the participant is inactive, observer, avatar, 2-point gaze or 3-point gaze. The suggested HC (AODCNN + Bi-LSTM) has revealed improved values for all evaluated measures such as accuracy, MCC, NPV, precision and so on.

For the considered real-time dataset, the accuracy of HC (AODCNN + Bi-LSTM) is found to be higher at 60% of training data over CNN, RNN, LSTM, RF, SVM, SVM + RF + NB + LR+ KNN (Vakadkar et al. 2021), CNN + TL (Alsaade and Alzahrani 2022), VGG16 + RF (Nawghare and Prasad 2025) and MSEG-Net (Chen et al. 2025) methods. As the training data percentage increases from 60 to 70, a slight improvement in accuracy is obtained for HC (AODCNN + Bi-LSTM) over CNN, RNN, LSTM, RF, SVM, SVM + RF + NB + LR+ KNN (Vakadkar et al. 2021), CNN + TL (Alsaade and Alzahrani 2022), VGG16 + RF (Nawghare and Prasad 2025) and MSEG-Net (Chen et al. 2025) methods. Likewise, at 80% of training data, a slight improvement in accuracy is obtained compared to 70% of training data for HC (AODCNN + Bi-LSTM) over CNN, RNN, LSTM, RF, SVM, SVM + RF + NB + LR+ KNN (Vakadkar et al. 2021), CNN + TL (Alsaade and Alzahrani 2022), VGG16 + RF (Nawghare and Prasad 2025) and MSEG-Net (Chen et al. 2025) methods. However, for all

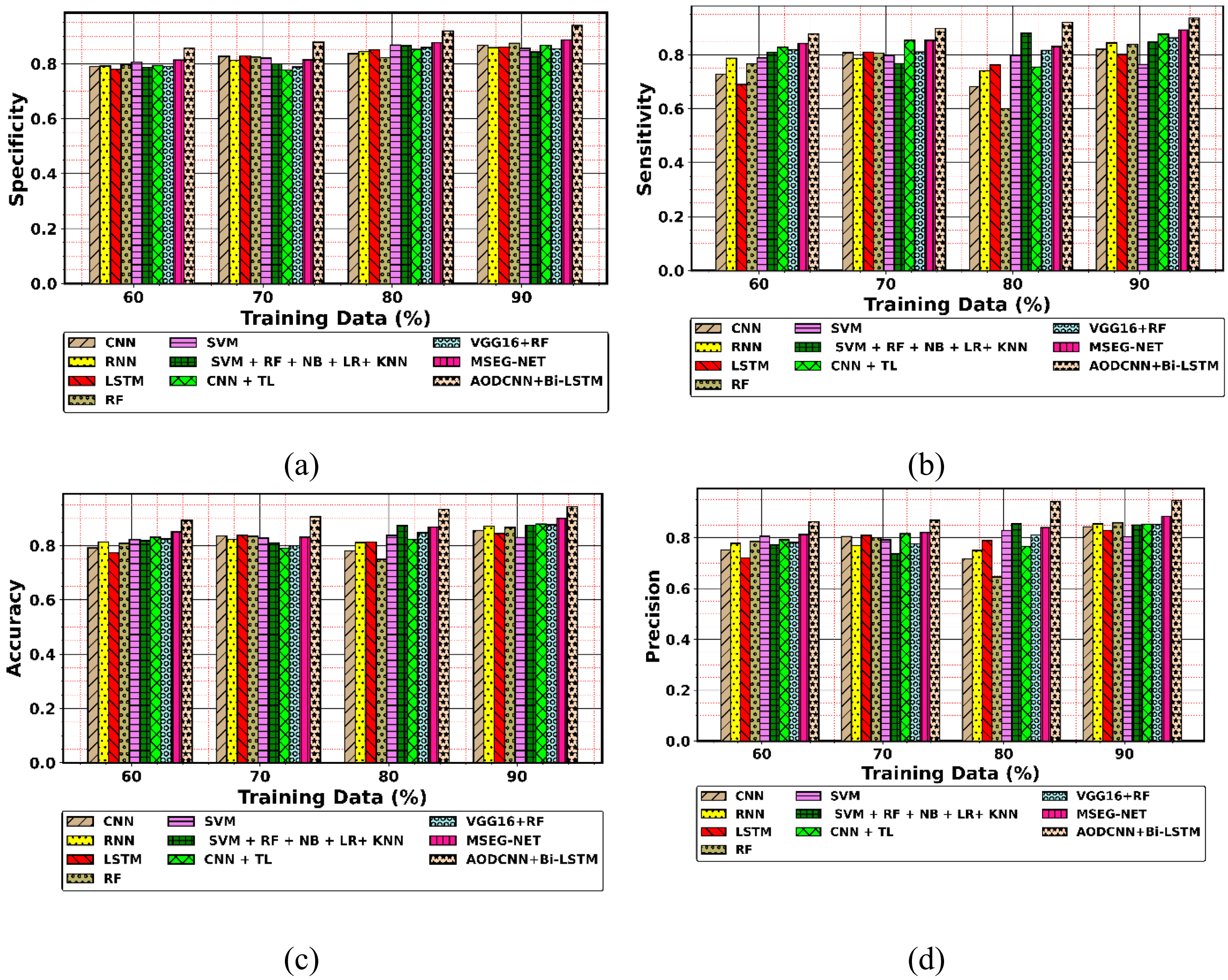

FIGURE 6 | Examination of the HC (AODCNN + Bi-LSTM) method for the VRAGS model for ASD children for (a) Specificity, (b) Sensitivity, (c) Accuracy and (d) Precision.

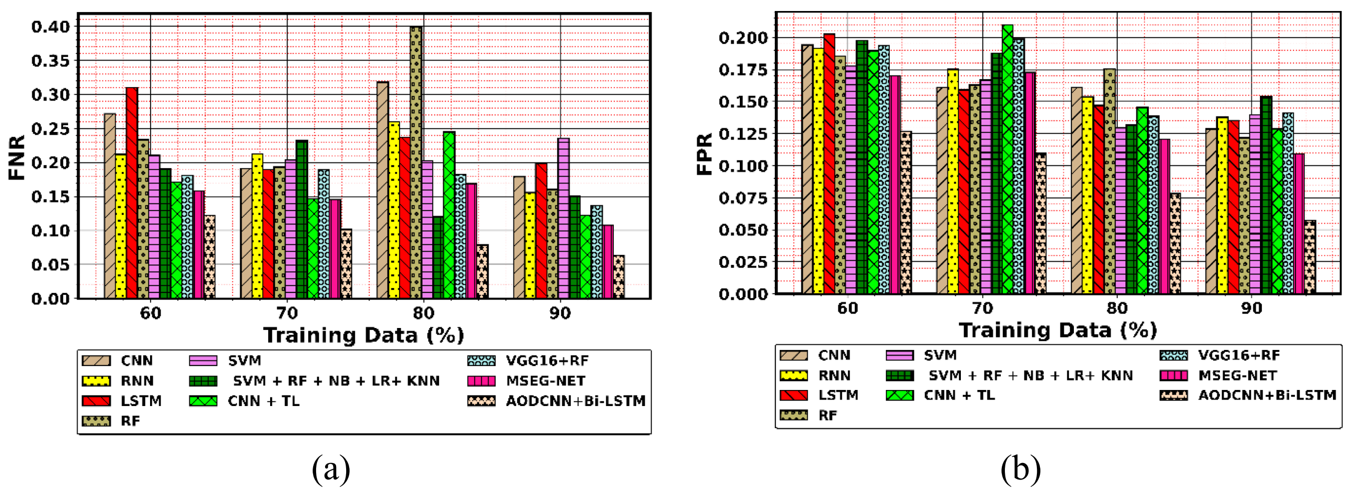

FIGURE 7 | Examination of the HC (AODCNN + Bi-LSTM) method for the VRAGS model for ASD children for (a) FNR and (b) FPR.

LPs, the proposed HC (AODCNN + Bi-LSTM) is found to be higher over CNN, RNN, LSTM, RF, SVM, SVM + RF + NB + LR+ KNN (Vakadkar et al. 2021), CNN + TL (Alsaade and Alzahrani 2022), VGG16 + RF (Nawghare and Prasad 2025), and MSEG-Net (Chen et al. 2025) methods. Likewise, in the case of FNR, a minimal FPR and FNR are obtained by the proposed HC (AODCNN + Bi-LSTM) model at 90% of training data. In other words, the FPR of HC (AODCNN + Bi-LSTM)

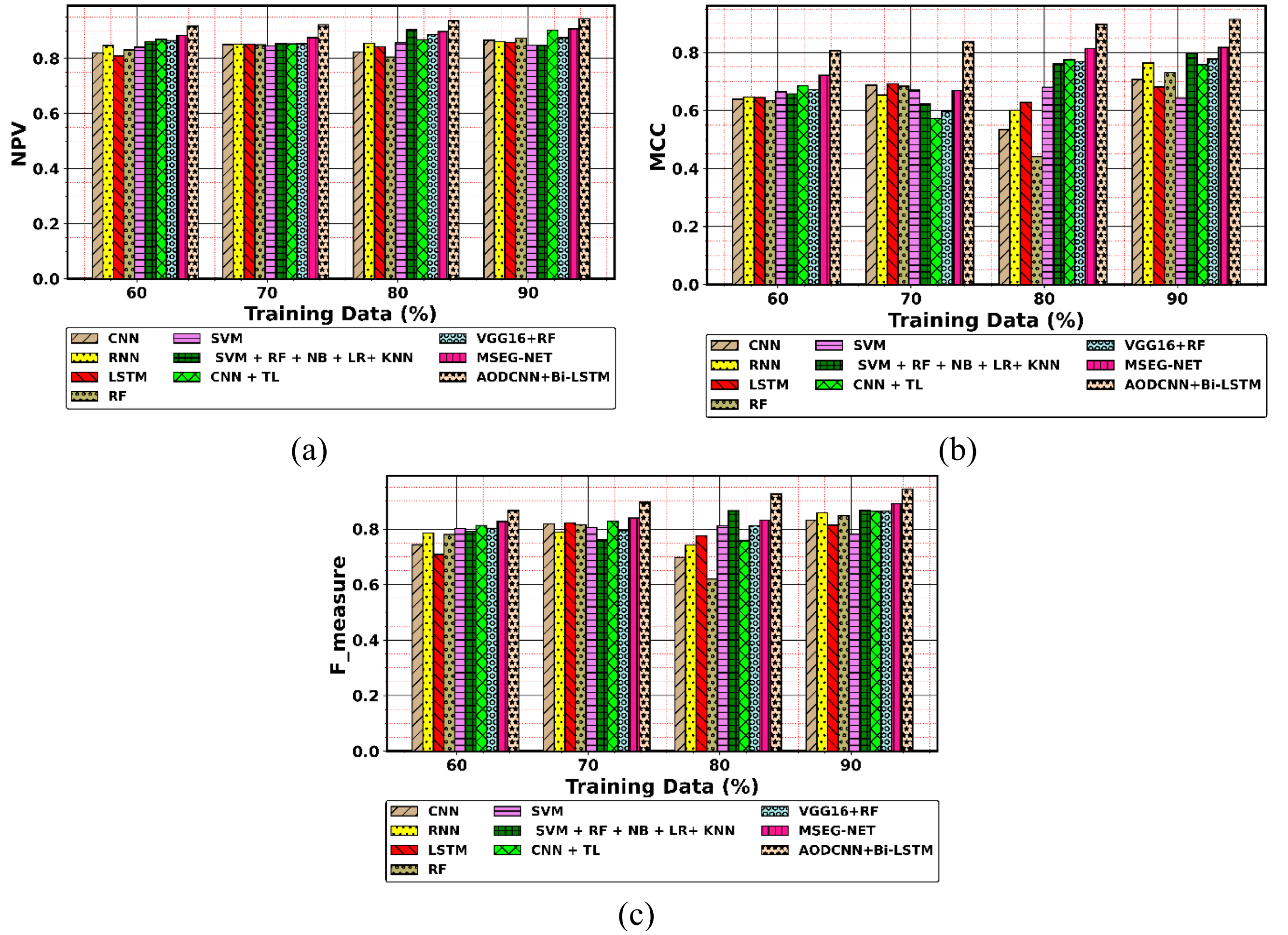

FIGURE 8 | Examination of the HC (AODCNN + Bi-LSTM) method for the VRAGS model for ASD children for (a) NPV, (b) MCC and (c) F measure.

TABLE 5 | Ablation study on the developed HC (AODCNN + Bi-LSTM) method over others for the VRAGS model for ASD children.

Metrics	HC (AODCNN + Bi-LSTM)	Proposed with conventional entropy	Proposed with traditional tanh normalization
Sensitivity	92.15%	78.06%	74.55%
FPR	7.81%	14.11%	16.11%
Accuracy	93.33%	83.34%	80.86%
NPV	93.63%	86.00%	83.68%
F-measure	92.60%	78.71%	75.81%
MCC	89.66%	65.37%	61.13%
Precision	94.35%	79.69%	76.16%
Specificity	91.94%	85.58%	83.14%
FNR	7.85%	21.94%	25.45%

is 0.125 at 60% of training data, 0.112 at 70% of training data, 0.075 at 80% of training data and 0.06 at 90% of training data. This shows that training data = 90% is the point at which better performance is achieved. When compared to CNN, RNN, LSTM, RF, SVM, SVM + RF + NB + LR+ KNN (Vakadkar et al. 2021), CNN + TL (Alsaade and Alzahrani 2022), VGG16 + RF (Nawghare and Prasad 2025) and MSEG-Net (Chen et al. 2025), the MCC of HC (AODCNN + Bi-LSTM) is high at about 0.92%. We achieve better results because we use upgraded tanh normalization with enhanced entropy features. Bi-LSTM with modifications and hybrid AODCNN also contributed to improved child ASD prediction.

7.6 | Ablation Study

The HC (AODCNN + Bi-LSTM) ablation study is presented in Table 5 and is suggested with extant tanh normalization instead of extant entropy. The sensitivity of the VRAGS model for ASD children using the HC (AODCNN + Bi-LSTM) classifier is 0.922, whereas sensitivity using HC (AODCNN + Bi-LSTM) with extant entropy and HC (AODCNN + Bi-LSTM) with extant tanh normalization is 0.780583 and 0.745466. Furthermore, NPV by means of HC (AODCNN + Bi-LSTM) with extant entropy, proposed with extant tanh normalization, is 0.836803 and 0.86001, whereas NPV employing the HC (AODCNN + Bi-LSTM) method is 0.936315. Better results are achieved by using an updated Tanh normalizing model with enhanced entropy features. In addition, the combination of hybrid AODCNN and Bi-LSTM with improvisations contributed to improved ASD prediction in children.

7.7 | Statistical Analysis

Table 6 shows the statistical study on the accuracy of the proposed HC (CNN + LSTM) over the traditional methods such as CNN, RNN, LSTM, RF, SVM, SVM + RF + NB + LR+ KNN (Vakadkar et al. 2021), CNN + TL (Alsaade and Alzahrani 2022), VGG16 + RF (Nawghare and Prasad 2025) and MSEG-Net (Chen et al. 2025). In case of maximum statistical metric, the HC (AODCNN + Bi-LSTM) obtained the greatest accuracy of 0.944, while CNN, RNN, LSTM, RF, SVM, SVM + RF + NB + LR+ KNN (Vakadkar et al. 2021), CNN + TL (Alsaade and Alzahrani 2022), VGG16 + RF (Nawghare and Prasad 2025) and MSEG-Net (Chen et al. 2025) have achieved accuracies of 0.856, 0.871, 0.844, 0.867, 0.837, 0.875, 0.879, 0.877 and 0.900 respectively. The mean case using HC (AODCNN + Bi-LSTM) is 0.919, while CNN, RNN, LSTM, RF, SVM, SVM + RF + NB + LR+ KNN (Vakadkar et al. 2021), CNN + TL (Alsaade and Alzahrani 2022), VGG16 + RF (Nawghare and Prasad 2025) and MSEG-Net (Chen et al. 2025) gained lower accuracies of 0.816, 0.830, 0.817, 0.815, 0.829, 0.844, 0.830, 0.837 and 0.862 respectively. An augmented Tanh Normalization model with enhanced entropy features yields better results. Moreover, hybrid AODCNN and Bi-LSTM produced improved children's ASD prediction.

7.8 | Overall Analysis of Classifier

Table 7 reveals the overall evaluation of classifiers for LP at 70%. The evaluation is performed on the proposed HC (CNN + LSTM) over CNN, RNN, LSTM, RF, SVM, SVM + RF + NB + LR+ KNN (Vakadkar et al. 2021), CNN + TL (Alsaade and Alzahrani 2022), VGG16 + RF (Nawghare and Prasad 2025) and MSEG-Net (Chen et al. 2025) schemes. The accuracy of HC (CNN + LSTM) is about 0.9333, while CNN, RNN, LSTM, RF, SVM, SVM + RF + NB + LR+ KNN (Zhao et al. 2018), CNN + TL (Vakadkar et al. 2021), VGG16 + RF (Nawghare and Prasad 2025) and MSEG-Net (Chen et al. 2025) models acquired lower accuracy values. The improvisations made in tanh normalization with enhanced entropy features resulted in superior results. Moreover, the prediction of ASD in children with a hybrid AODCNN and Bi-LSTM resulted in enhanced results.

TABLE 6 | Statistical analysis of accuracy for the HC (AODCNN + Bi-LSTM) method over others for the VRAGS model for ASD children.

	CNN	RNN	LSTM	RF	SVM	SVM + RF + NB + LR + KNN	CNN + TL	VGG16 + RF	MSEG-NET	AODCNN+Bi-LSTM
Mean	81.60%	83.00%	81.70%	81.50%	82.90%	84.40%	83.00%	83.70%	86.20%	91.90%
Median	81.40%	81.80%	82.60%	82.20%	82.90%	84.60%	82.60%	83.60%	85.90%	91.90%
Std	3.10%	2.40%	2.80%	4.30%	0.50%	3.00%	3.20%	2.80%	2.50%	2.00%
Min	78.10%	81.10%	77.40%	74.90%	82.20%	81.10%	78.90%	80.00%	83.20%	89.50%
Max	85.60%	87.10%	84.40%	86.70%	83.70%	87.50%	87.90%	87.70%	90.00%	94.40%

TABLE 7 | Overall analysis of developed HC (AODCNN + Bi-LSTM) method over others for VRAGS model for ASD children for training data = 70%.

Metrics	Sensitivity	F-measure	Accuracy	MCC	NPV	Precision	FPR	Specificity	FNR
CNN	0.6822	0.6975	0.7810	0.5346	0.8235	0.7177	0.1612	0.8363	0.3178
RNN	0.7404	0.7428	0.8110	0.6011	0.8542	0.7505	0.1536	0.8439	0.2596
LSTM	0.7633	0.7751	0.8132	0.6276	0.8423	0.7893	0.1469	0.8506	0.2367
RF	0.6011	0.6198	0.7489	0.4416	0.8046	0.6461	0.1754	0.8221	0.3989
SVM	0.7975	0.8119	0.8372	0.6806	0.8572	0.8286	0.1295	0.8680	0.2025
SVM + RF + NB + LR+ KNN (Zhao et al. 2018)	0.8797	0.8658	0.8731	0.7606	0.9037	0.8548	0.1319	0.8656	0.1203
CNN + TL (Vakadkar et al. 2021)	0.7550	0.7577	0.8222	0.7756	0.8681	0.7664	0.1454	0.8521	0.2450
VGG16 + RF (Nawghare and Prasad 2025)	0.8174	0.8117	0.8476	0.7681	0.8859	0.8106	0.1386	0.8589	0.1826
MSEG-Net (Chen et al. 2025)	0.8313	0.8318	0.8677	0.8134	0.8968	0.8401	0.1207	0.8768	0.1687
HC (AODCNN + Bi-LSTM)	0.9215	0.9260	0.9333	0.8966	0.9363	0.9434	0.0781	0.9194	0.0785

7.9 | Statistical Test Analysis

To determine whether the performance variations among models are statistically significant and not the result of chance, a statistical test is employed. Two non-parametric statistical tests, the Friedman test and the Wilcoxon signed-rank test, are used in this investigation. The Friedman test assesses the overall ranking differences across several models concurrently, whereas the Wilcoxon signed-rank test conducts pairwise comparisons between the suggested HC (AODCNN + Bi-LSTM) approach and each current approach. For the Wilcoxon test and Friedman test, the model having less than 0.05 indicates the model is statistically significant. The statistical test results are shown in Table 8. For the Wilcoxon test and Friedman test results, the model having a p-value less than 0.05 indicates the proposed model demonstrates improved performance over existing models. Specifically, for the Friedman test, the models such as CNN, RNN, CNN + TL and MSEG-NET have p-values below 0.05. These findings indicate that the proposed model achieves a significantly superior overall performance compared to the models. Overall, these results confirm the robustness and effectiveness of the proposed HC (AODCNN + Bi-LSTM) model in attention prediction, validating its advantage over existing approaches.

TABLE 8 | Statistical test comparison of the proposed HC (AODCNN + Bi-LSTM) model against the existing methods in terms of Wilcoxon test and Friedman test.

Proposed HC (AODCNN + Bi-LSTM) model Vs	Wilcoxon p	Friedman p
CNN	0.111	0.018
RNN	0.088	0.030
LSTM	0.083	0.066
RF	0.123	0.077
SVM	0.104	0.057
SVM + RF + NB + LR+ KNN	0.127	0.054
CNN + TL	0.079	0.040
VGG16 + RF	0.103	0.055
MSEG-NET	0.115	0.048

7.10 | K-Fold Analysis

K-fold analysis is a model evaluation technique in which the dataset is divided into K equal folds. The model is trained on K–1 folds and tested on the remaining fold, and this process is repeated K times so that each fold is used once for testing. The final performance is obtained by averaging the results across all folds, providing a reliable estimate of the model's generalization ability. Table 9 presents a K-fold comparison between the proposed HC (AODCNN + Bi-LSTM) model and existing techniques. The results demonstrate the superior and consistent performance of the proposed method across all five folds, highlighting its robustness and generalization capability. Existing models such as CNN, RNN, MSEG-NET, SVM + RF +

TABLE 9 | K-fold comparison of the proposed method against existing techniques.

Models	Fold = 1	Fold = 2	Fold = 3	Fold = 4	Fold = 5
CNN	0.927	0.939	0.910	0.922	0.916
RNN	0.914	0.917	0.924	0.926	0.932
LSTM	0.927	0.937	0.918	0.945	0.911
RF	0.937	0.927	0.932	0.916	0.918
SVM	0.942	0.949	0.923	0.938	0.945
SVM + RF + NB + LR+ KNN	0.946	0.913	0.912	0.917	0.945
CNN + TL	0.914	0.927	0.948	0.931	0.938
VGG16 + RF	0.944	0.931	0.917	0.927	0.945
MSEG-NET	0.920	0.927	0.921	0.935	0.921
Proposed HC (AODCNN + Bi-LSTM) model	0.960	0.962	0.963	0.958	0.963

NB + LR+ KNN and LSTM exhibit competitive performance, with accuracy values ranging between 0.91 and 0.95 across the folds. In contrast, the proposed HC (AODCNN + Bi-LSTM) model consistently achieves the highest accuracy across all five folds, ranging from 0.958 to 0.963. Overall, these results confirm that the proposed HC (AODCNN + Bi-LSTM) model not only outperforms existing state-of-the-art techniques but also maintains high stability and generalization across different data splits.

8 | Conclusion

A novel VRAGS model for children with ASD was created through this effort. In this case, the first step was the data collection phase. Following the stage of data collection, an improved tanh model was used to normalize the data. Next, the raw data, statistical characteristics, and enhanced entropy were retrieved. In the final phase, the prediction process was completed via the HC model that combined AODCNN and Bi-LSTM. Then, the harmonic mean was taken for AODCNN and Bi-LSTM outputs to find if the output was inactive, observer, avatar, 2-point gaze or 3-point gaze. Moreover, the proposed HC (AODCNN + Bi-LSTM) model was evaluated against traditional methods, including CNN, RNN, LSTM, RF, SVM, SVM+RF+NB+LR+KNN, CNN+TL, VGG16+RF and MSEG-Net using multiple performance measures such as accuracy, precision, F-measure and MCC. At a training data percentage of 90%, the proposed HC (AODCNN + Bi-LSTM) model achieved an accuracy of 0.934, outperforming CNN (0.813), RNN (0.820), LSTM (0.794), RF (0.821), SVM (0.784), SVM+RF+NB+LR+KNN (0.829) and CNN+TL (0.831). Additionally, as the training data percentage increased from 60% to 90%, the proposed model consistently demonstrated lower FNR and FPR, achieving values of 0.052 and 0.048, respectively, at LP = 90%. These results indicate that the efficiency of the suggested HC (AODCNN + Bi-LSTM) model was attributed to the improved tanh normalization-based preprocessing, improved entropy-based feature extraction, and the hybrid combination of the AODCNN and Bi-LSTM models. These innovations not only provide superior accuracy but also effectively reduce misclassification rates compared to conventional approaches. Although the results are promising, the system's effectiveness may differ among individuals due to the diverse characteristics of ASD. Its dependency on specialized VR hardware could also restrict accessibility for some users. Moreover, the advanced models employed demand significant computational resources and technical expertise. Future research should focus on expanding the participant pool to include a larger and more diverse group of children with ASD, in order to validate the VRAGS model's robustness and generalizability across different age ranges and severity levels. Longitudinal studies could provide insights into the sustained effects of VR-based interventions on cognitive development. Additionally, enhancing accessibility to the system through mobile VR platforms of the system could enable broader implementation.

Nomenclature

AI	artificial intelligence
AODCNN	Adam-optimized DCNN
ASD	autism spectrum disorder
Bi-LSTM	bidirectional long short-term memory
CNN	convolutional neural network
DAGAN	dense attentive GAN
DGAN	dense GAN
DL	deep learning
HMD	head-mounted display
IVR	immersive virtual reality
HC	hybrid classifier
KNN	K-nearest neighbour
ML	machine learning
MLP	multilayer perceptrons
MGOA	modified grasshopper optimization algorithm
NN	neural network
NB	Naïve Bayes
LSTM	long short-term memory

LR	linear regression
LP	learning percentage
SD	Standard Deviation
sMRI	structural magnetic resonance images
TL	transfer learning
SVM	support vector machine
RNNs	recurrent neural networks
RF	random forest
VGG19	Visual Geometry Group Network

Author Contributions

P. Valarmathi conceived the presented idea and designed the analysis. Also, he carried out the experiment and wrote the manuscript with support from A. Packialatha. All authors discussed the results and contributed to the final manuscript. All authors read and approved the final manuscript.

Acknowledgements

I would like to express my very great appreciation to the co-authors of this manuscript for their valuable and constructive suggestions during the planning and development of this research work.

Funding

The authors have nothing to report.

Ethics Statement

The authors have nothing to report.

Consent

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Alcaniz Raya, M., J. Marín-Morales, M. E. Minissi, G. Teruel Garcia, L. Abad, and I. A. Chicchi Giglioli. 2020. "Machine Learning and Virtual Reality on Body Movements' Behaviors to Classify Children With Autism Spectrum Disorder." *Journal of Clinical Medicine* 9, no. 5: 1260.

Alsaade, F. W., and M. S. Alzahrani. 2022. "Classification and Detection of Autism Spectrum Disorder Based on Deep Learning Algorithms." *Computational Intelligence and Neuroscience* 2022, no. 1: 8709145.

Amat, A. Z., H. Zhao, A. Swanson, A. S. Weitlauf, Z. Warren, and N. Sarkar. 2021. "Design of an Interactive Virtual Reality System, InViRS, for Joint Attention Practice in Autistic Children." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 29: 1866–1876.

Banire, B., D. Al Thani, M. Qaraqe, and B. Mansoor. 2021. "Face-Based Attention Recognition Model for Children With Autism Spectrum Disorder." *Journal of Healthcare Informatics Research* 5, no. 4: 420–445.

Bejarano-Martín, Á., R. Canal-Bedia, M. Magán-Maganto, et al. 2020. "Early Detection, Diagnosis and Intervention Services for Young Children With Autism Spectrum Disorder in the European Union (ASDEU): Family and Professional Perspectives." *Journal of Autism and Developmental Disorders* 50, no. 9: 3380–3394.

Cai, Y., N. K. Chia, D. Thalmann, N. K. Kee, J. Zheng, and N. M. Thalmann. 2013. "Design and Development of a Virtual Dolphinarium for Children With Autism." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 21, no. 2: 208–217.

Charpentier, J., K. Kovarski, E. Houy-Durand, et al. 2018. "Emotional Prosodic Change Detection in Autism Spectrum Disorder: An Electrophysiological Investigation in Children and Adults." *Journal of Neurodevelopmental Disorders* 10, no. 1: 28.

Chen, H., X. Zhang, Z. Chen, Y. Ren, and R. Liu. 2025. "Auxiliary Diagnostic Method for Children With Autism Spectrum Disorder Based on Virtual Reality and Eye-Tracking Technology." *Scientific Reports* 15, no. 1: 40552.

de Mira Gobbo, M. R., C. R. S. C. de Barbosa, M. Morandini, F. Mafort, and J. L. V. M. Mioni. 2021. "ACA Game for Individuals With Autism Spectrum Disorder." *Entertainment Computing* 38: 100409.

ElMoazen, D., O. Sobhy, R. Abdou, and H. AbdelMotaleb. 2020. "Binaural Interaction Component of the Auditory Brainstem Response in Children With Autism Spectrum Disorder." *International Journal of Pediatric Otorhinolaryngology* 131: 109850.

Eni, M., I. Dinstein, M. Ilan, I. Menashe, G. Meiri, and Y. Zigel. 2020. "Estimating Autism Severity in Young Children From Speech Signals Using a Deep Neural Network." *IEEE Access* 8: 139489–139500.

Frazier, T. W., K. R. Ratliff, C. Gruber, Y. Zhang, P. A. Law, and J. N. Constantino. 2014. "Confirmatory Factor Analytic Structure and Measurement Invariance of Quantitative Autistic Traits Measured by the Social Responsiveness Scale-2." *Autism* 18, no. 1: 31–44.

Frolli, A., G. Savarese, F. Di Carmine, et al. 2022. "Children on the Autism Spectrum and the Use of Virtual Reality for Supporting Social Skills." *Children* 9, no. 2: 181.

Garcia-Garcia, J. M., V. M. Penichet, M. D. Lozano, and A. Fernando. 2022. "Using Emotion Recognition Technologies to Teach Children With Autism Spectrum Disorder How to Identify and Express Emotions." *Universal Access in the Information Society* 21, no. 4: 809–825.

Garcia-Garcia, M., F. Barceló, I. C. Clemente, and C. Escera. 2010. "The Role of the Dopamine Transporter DAT1 Genotype on the Neural Correlates of Cognitive Flexibility." *European Journal of Neuroscience* 31, no. 4: 754–760.

Gardiner, E., and G. Iarocci. 2018. "Everyday Executive Function Predicts Adaptive and Internalizing Behavior Among Children With and Without Autism Spectrum Disorder." *Autism Research* 11, no. 2: 284–295.

Geurts, H. M., B. Corbett, and M. Solomon. 2009. "The Paradox of Cognitive Flexibility in Autism." *Trends in Cognitive Sciences* 13, no. 2: 74–82.

Gioia, G. A., S. C. Guy, and L. Kenworthy. 2000. "Behavior Rating Inventory of Executive Function." *Child Neuropsychology*.

Glennon, J. M., H. D'Souza, L. Mason, A. Karmiloff-Smith, and M. S. Thomas. 2020. "Visuo-Attentional Correlates of Autism Spectrum Disorder (ASD) in Children With Down Syndrome: A Comparative Study With Children With Idiopathic ASD." *Research in Developmental Disabilities* 104: 103678.

Goel, N., B. Grover, D. Gupta, A. Khanna, and M. Sharma. 2020. "Modified Grasshopper Optimization Algorithm for Detection of Autism Spectrum Disorder." *Physical Communication* 41: 101115.

Gu, J., Z. Wang, J. Kuen, et al. 2018. "Recent Advances in Convolutional Neural Networks." *Pattern Recognition* 77: 354–377.

Hadri, S. A., and A. Bouramoul. 2023. "Towards a Deep Learning Based Contextual Chat Bot for Preventing Depression in Young Children With Autistic Spectrum Disorder." *Smart Health* 27: 100371.

Herrero, J. F., and G. Lorenzo. 2020. "An Immersive Virtual Reality Educational Intervention on People With Autism Spectrum Disorders (ASD) for the Development of Communication Skills and Problem Solving." *Education and Information Technologies* 25, no. 3: 1689–1722.

Ionescu, T. 2012. "Exploring the Nature of Cognitive Flexibility." *New Ideas in Psychology* 30, no. 2: 190–200.

Johnston, D., H. Egermann, and G. Kearney. 2020. "Soundfields: A Virtual Reality Game Designed to Address Auditory Hypersensitivity in Individuals With Autism Spectrum Disorder." *Applied Sciences* 10, no. 9: 2996.

Jonsdottir, S. L., E. Saemundsen, S. Gudmundsdottir, G. S. Haraldsdottir, A. H. Palsdottir, and V. Rafnsson. 2020. "Implementing an Early Detection Program for Autism in Primary Healthcare: Screening, Education of Healthcare Professionals, Referrals for Diagnostic Evaluation, and Early Intervention." *Research in Autism Spectrum Disorders* 77: 101616.

Jyoti, V., and U. Lahiri. 2019. "Virtual Reality Based Joint Attention Task Platform for Children With Autism." *IEEE Transactions on Learning Technologies* 13, no. 1: 198–210.

Ke, F., J. Moon, and Z. Sokolikj. 2022. "Virtual Reality–Based Social Skills Training for Children With Autism Spectrum Disorder." *Journal of Special Education Technology* 37, no. 1: 49–62.

Kuschner, E. S., I. W. Eisenberg, B. Orionzi, et al. 2015. "A Preliminary Study of Self-Reported Food Selectivity in Adolescents and Young Adults With Autism Spectrum Disorder." *Research in Autism Spectrum Disorders* 15: 53–59.

Kuttala, D., D. Mahapatra, R. Subramanian, and V. R. M. Oruganti. 2022. "Dense Attentive GAN-Based One-Class Model for Detection of Autism and ADHD." *Journal of King Saud University, Computer and Information Sciences* 34, no. 10: 10444–10458.

Lai, M., J. Lee, S. Chiu, et al. 2020. "A Machine Learning Approach for Retinal Images Analysis as an Objective Screening Method for Children With Autism Spectrum Disorder." *EClinicalMedicine* 28: 100588.

Lane, A. E., M. E. Geraghty, G. S. Young, and J. L. Rostorfer. 2014. "Problem Eating Behaviors in Autism Spectrum Disorder Are Associated With Suboptimal Daily Nutrient Intake and Taste/Smell Sensitivity." *Infant Child Adolescent Nutrition* 6, no. 3: 172–180.

Latha, L., and S. Thangasamy. 2011. "Efficient Approach to Normalization of Multimodal Biometric Scores." *International Journal of Computer Applications* 32, no. 10: 57–64.

Leigh, S. J., and M. J. Morris. 2020. "Diet, Inflammation and the Gut Microbiome: Mechanisms for Obesity-Associated Cognitive Impairment." *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1866, no. 6: 165767.

Lu, S. C., P. Rowe, C. Tachtatzis, et al. 2022. "Swipe Kinematic Differences in Young Children With Autism Spectrum Disorders Are Task- and Age-Dependent: A Smart Tablet Game Approach." *Brain Disorders* 5: 100032.

Maechler, M., P. Rousseeuw, C. Croux, et al. 2021. "Robustbase: Basic Robust Statistics." http://robustbase.r-forge.r-project.org/.

Manjur, S. M., M. B. Hossain, P. A. Constable, et al. 2022. "Detecting Autism Spectrum Disorder Using Spectral Analysis of Electroretinogram and Machine Learning: Preliminary Results." In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 3435–3438. IEEE.

Mazumdar, P., G. Arru, and F. Battisti. 2021. "Early Detection of Children With Autism Spectrum Disorder Based on Visual Exploration of Images." *Signal Processing: Image Communication* 94: 116184.

McCulloch, D. L., M. Kondo, R. Hamilton, et al. 2019. "ISCEV Extended Protocol for the Stimulus–Response Series for Light-Adapted Full-Field ERG." *Documenta Ophthalmologica* 138, no. 3: 205–215.

Moreno, J. C., E. M. S. Morales, and E. P. Seller. 2017. "Case Study of the Vulnerabilities That Children With Autism Spectrum Disorder Have in Education: The Importance of Early Detection." *Procedia-Social and Behavioral Sciences* 237: 661–666.

Nawghare, P., and J. Prasad. 2025. "Hybrid CNN and Random Forest Model With Late Fusion for Detection of Autism Spectrum Disorder in Toddlers." *MethodsX* 14: 103278.

Puli, A., and A. Kushki. 2019. "Toward Automatic Anxiety Detection in Autism: A Real-Time Algorithm for Detecting Physiological Arousal in the Presence of Motion." *IEEE Transactions on Biomedical Engineering* 67, no. 3: 646–657.

Rakić, M., M. Cabezas, K. Kushibar, A. Oliver, and X. Llado. 2020. "Improving the Detection of Autism Spectrum Disorder by Combining Structural and Functional MRI Information." *NeuroImage: Clinical* 25: 102181.

Rusli, N., S. N. Sidek, H. M. Yusof, N. I. Ishak, M. Khalid, and A. A. A. Dzulkarnain. 2020. "Implementation of Wavelet Analysis on Thermal Images for Affective States Recognition of Children With Autism Spectrum Disorder." *IEEE Access* 8: 120818–120834.

Sheldrick, R. C., E. Frenette, J. D. Vera, et al. 2019. "What Drives Detection and Diagnosis of Autism Spectrum Disorder? Looking Under the Hood of a Multi-Stage Screening Process in Early Intervention." *Journal of Autism and Developmental Disorders* 49, no. 6: 2304–2319.

Statistics Canada. n.d. "Median." https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch11/median-mediane/5214872-eng.htm.

Vakadkar, K., D. Purkayastha, and D. Krishnan. 2021. "Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques." *SN Computer Science* 2, no. 5: 386.

Wikimedia Foundation. n.d.-a. "Conditional Entropy." https://en.wikipedia.org/wiki/Conditional_entropy#Definition.

Wikimedia Foundation. n.d.-b. "Standard Deviation." https://en.wikipedia.org/wiki/Standard_deviation.

Wikimedia Foundation. n.d.-c. "Statistic." https://en.wikipedia.org/wiki/Statistic.

Xiao, W., M. Li, M. Chen, and A. Barnawi. 2020. "Deep Interaction: Wearable Robot-Assisted Emotion Communication for Enhancing Perception and Expression Ability of Children With Autism Spectrum Disorders." *Future Generation Computer Systems* 108: 709–716.

Yaneva, V., S. Eraslan, Y. Yesilada, and R. Mitkov. 2020. "Detecting High-Functioning Autism in Adults Using Eye Tracking and Machine Learning." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28, no. 6: 1254–1261.

Zhao, H., A. R. Swanson, A. S. Weitlauf, Z. E. Warren, and N. Sarkar. 2018. "Hand-in-Hand: A Communication-Enhancement Collaborative Virtual Reality System for Promoting Social Interaction in Children With Autism Spectrum Disorders." *IEEE Transactions on Human-Machine Systems* 48, no. 2: 136–148.

Zheng, Z., Q. Fu, H. Zhao, et al. 2016. "Design of an Autonomous Social Orienting Training System (ASOTS) for Young Children With Autism." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25, no. 6: 668–678.

Zhou, X., J. Lin, Z. Zhang, Z. Shao, S. Chen, and H. Liu. 2020. "Improved Itracker Combined With Bidirectional Long Short-Term Memory for 3D Gaze Estimation Using Appearance Cues." *Neurocomputing* 390: 217–225.