Deep learning with image-based autism spectrum disorder analysis: a systematic review

Md. Zasim Uddin[a,*], Arif Shahriar[a], Md. Nadim Mahamood[a], Fady Alnajjar[b], Md. Ileas Pramanik[a], Md Atiqur Rahman Ahad[c]

[a]*Department of Computer Science and Engineering, Begum Rokeya University, Rangpur, Bangladesh*
[b]*Department of Computer Science and Software Engineering, United Arab Emirates University, UAE*
[c]*Department of Computer Science and Digital Technologies, University of East London, London, UK*

Abstract

Autism spectrum disorder (ASD) is a collection of neuro-developmental disorders associated with social, communication, and behavioral difficulties. It is necessary for early detection to mitigate the adverse effects of this disorder by starting special education in a school and rehabilitation center to enhance children's daily lives. There are two types of methods available to diagnose and rehabilitate ASD. One of them is the manual method (i.e., observation or interview-based approach), which is diagnosed through observation or interview of parent or caregiver. It is time-consuming, subjective, and mostly depended on examining behavioral symptoms. Another method is the automatic diagnosis using traditional machine learning (ML) and modern deep learning (DL)-based approaches using images. This systematic review aims to examine the application of the DL-based approach using images or videos in autism research. It includes the publications indexed on PubMed, IEEE Xplore, ACM Digital Library, and Google Scholar, conducted from 2017 to 2022. The result is reported on the PRISMA statement. A total of 130 studies are included in this analysis. Eligible papers are categorized based on the different features extracted to feed the DL-based approach. Existing well-known public and private datasets, including images or videos for autism research, are extensively reviewed and discussed in this systematic review. Moreover, different rehabilitation strategies that are highly helpful for ASD individuals are included in this review. Finally, various current challenges for the automated detection, classification, and rehabilitation of ASD are presented. The review concludes that the application of deep learning for precise and affordable diagnosis of autism is rising substantially.

Keywords: Autism Spectrum Disorder, Deep Learning, Image or Video, Detection and Classification, MRI

1. Introduction

Autism spectrum disorders (ASD) are a diverse group of neuropsychiatric conditions. They are characterized by some degree of difficulty with impairments in social communication, personal interaction, academic functioning, and restricted and repetitive behaviors. Notably, people with ASD may behave, communicate, and learn in ways different from most others. The Autism and Developmental Disabilities Monitoring Network (ADDM) of the Centers for Disease Control and Prevention (CDC), USA estimated that about one in 44 children had been identified with ASD (Maenner et al., 2021). Diagnosing ASD can be difficult as there is no direct pathological or radiological examination to diagnose the disorder. However, individuals with ASD can exhibit different signs and symptoms, which are conspicuous in an early stage of life (Lord et al., 2006), including but not limited to joint attention (Wilkinson, 1998), trying to avoid eye contact, the obsessiveness of activities (Tanguay et al., 1998), stereotypical motor movements (Großekathöfer et al., 2017), atypical sensory responsivity (Kanner et al., 1943). In particular, ASD children have less visual attention in contrast than the typically developed (TD) children (Tanaka and Sung, 2016). The level of these symptoms varies within individuals. Therefore, it is sometimes considered a spectrum condition (Lord et al., 2018).

Early detection and diagnosis are essential to ensure that reasonable treatment and/or therapy for children with ASD symptoms can be managed. ASD subjects

*Corresponding author. E-mail: zasim@am.sanken.osaka-u.ac.jp

require to receive the services to achieve their full potential for bringing a more significant outcome for society (Pickles et al., 2016). Therefore, it is important to employ proper diagnostic and rehabilitation techniques. There are two ways to diagnose and monitor children with ASD: the manual system, and the automatic diagnosis system. The automatic systems explore various computer vision- or image-based strategies with traditional machine learning (ML) as well as deep learning (DL).

Observation- and interview-based methods are two widespread manual ASD detection and diagnosis systems. The Childhood Autism Rating Scale (CARS) (Schopler et al., 1980) consists of 15 items to assess ASD. CARS has a range of scores to indicate the ASD levels, e.g., a score between 30-37 is considered as mild ASD, while 38-60 is used as severe ASD. On the other hand, interview-based detection and diagnosis systems depend on the interview with parents or caregivers. The Developmental, Dimensional, and Diagnostic Interview (3DI) consists of 183 items regarding the children's developmental delay history and family background (Skuse et al., 2004). It can be used to identify children and adults with ASD. Similarly, the Autism Diagnostic Interview-Revised (ADI-R) is an investigator-based interview where parents and possible ASD cases need to be present in-person (Lord et al., 1994). Moreover, the Asperger Syndrome Diagnostic Interview (ASDI) is also an investigator-based interview where physicians present and investigate for 15–20 min whether or not persons meet the criteria of ASD (Gillberg et al., 2001). Furthermore, Gilliam Autism Rating Scale (GARS) (Lecavalier, 2005) contains 56 items with four categories: stereotyped behaviors, communication, social interaction, and developmental disturbance. They rated the severity of ASD by scoring a range of items. However, the manual systems depend on behavioral symptoms and parents' or caregivers' observations and require an expert physician to make judgments. Therefore, it cannot capture data on real situations of typical daily-life activities. Moreover, these processes are costly and time-consuming (Galliver et al., 2017), e.g., the ADI-R experiment takes around 2-3 hours to diagnose (Rutter et al., 2003).

To mitigate the problem of manual detection and diagnosis, researchers tend to develop automatic tools to analyze ASD, which provide more accuracy and reduce diagnosis time (Noorbakhsh-Sabet et al., 2019). Additionally, it offers an early ASD diagnosis at the age of two years (Chen et al., 2022; Chang et al., 2021). Initially, the computer vision with traditional machine learning (ML) approaches (Thabtah, 2019; Hossain et al., 2021; Sapiro et al., 2019) were employed to develop automated ASD screening tools that can be more time-efficient and inexpensive than regular manual diagnosis. Meanwhile, the deep learning (DL)-based approaches have been exploited effectively in disease detection and diagnosis (e.g., brain tumors, breast cancer, and cardiac diseases, etc.). The significant advantages of DL-based methods are that they can extract features automatically, reduce error, and mostly outperforms traditional machine learning (ML)-based approaches (Niu et al., 2020). Recently, researchers employed DL-based methods with the image(s) and videos in autism research to detect, classify, diagnose and/or monitor ASD children.

A large number of studies on autism research have been published in the last five years (i.e., 2017 to 2022) using the deep learning-based method with the image(s) and videos. However, to the best of our knowledge, no systematic review studies have yet been published using those papers. Therefore, this paper aims to provide a systematic and comprehensive review of the published deep learning studies with images or videos to analyze ASD.

2. Related Work and Our Contribution

In the following section, relevant reviews related to ASD are discussed, and finally, the contributions of this systematic review are summarized.

Some reviewed works regarding ASD are available in the literature considering supervised and unsupervised traditional ML and modern DL-based approaches. For example, Hyde et al. (2019) provided a review of supervised machine learning on ASD, including 45 papers. They also examined text mining to uncover probable ASD genes and look into unclear connections between ASD and other domains. They, however, considered only five studies about deep learning. Regarding unsupervised approaches, Parlett-Pelleriti et al. (2022) provided a review of ASD with only three studies about deep learning.

In contrast, computer vision techniques were employed to analyze ASD, and these approaches were reviewed in (Rahman et al., 2021; De Belen et al., 2020; Minissi et al., 2021). For example, Rahman et al. (2021) provided a review to detect ASD using different human activity analyses. They summarized the work related to capturing and analyzing sensor data from a person's movement, gesture, or motion, while De Belen et al. (2020) provided a review based on different computer vision-based features and datasets for ASD detection and classification of the published work from 2009 to

2019. However, These reviews considered only 14 and 20 DL-based approaches. Furthermore, Minissi et al. (2021) reviewed only 11 papers and focused on classifying ASD based on ML and social visual attention towards social stimuli. They also discussed various ML-based models and eye movement as biomarkers. However, considering only eye movement biomarkers, they reviewed four DL-based studies. Moreover, Song et al. (2021) reviewed the study on traditional ML to distinguish ASD from TD, while DL-based approaches were employed for rehabilitation in (Khodatars et al., 2021). However, they limit themselves to considering only structural and functional neuroimaging data. In contrast to the above, our comprehensive systematic review considered only DL-based approaches, with 130 articles published from 2017 to 2022 regarding the broad aspects of image and video modalities, including RGB images, neuroimages, images generated from other domains (e.g., spectrogram from EEG signal). To the best of our knowledge, this study is the first work with an extensive and systematic review of DL-based approaches with image modalities to analyze ASD and the corresponding public and private datasets along with the different rehabilitation procedures to enhance the daily-live of the individual with ASD. The significant contribution of our studies are below:

- Presenting an extensive and systematic review conducted on deep learning with image-based research studies covering 130 articles from 2017 to 2022 for ASD detection, classification and diagnosis, rehabilitation therapy, and ASD research.

- Presenting a depth analysis of the publicly available, along with private datasets that were employed with deep learning-based approaches to analyze the ASD research. Furthermore, the performance evaluation metrics are presented.

- Finally, opinions on current challenges and future directions in ASD research are provided. In addition, by providing a thorough summary of existing deep learning with image-based approaches, including the focus, network structure, loss function, activation function, result, etc., as well as dataset and performance metrics in ASD, this systematic review can serve as a convincing resource that helps develop a DL-based approach to analyze ASD.

3. Materials and Methods

We follow a systematic way to explore and analyze the papers to select. All procedures were conducted as the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Moher et al., 2009).

3.1. Eligibility Criteria

All titles and abstracts were filtered to include studies that fulfill the following criteria: (i) Deep learning-based approaches were employed to extract features, detection, classification, or rehabilitation. (ii) The study must use images or videos to extract features or off-the-self features extracted from images or videos to study related to autism. (iii) The articles were written in English. (iv) The article should focus on autism, or part of the study was related to autism research. (v) The studies related to human beings only. (vi) Paper published between 2017 to 2022.

3.2. Search Process

In this review, PubMed, Scopus, Springer Link, ACM, IEEE Xplore, Google Scholars, as well as other conferences or journals, were used to acquire the articles on ASD detection, classification, and rehabilitation using a DL-based approach with image or video. Furthermore, the search query employed includes the combination of the following terms “Autism Spectrum Disorder,” “Autism,” “ASD,” “Detection,” ”Classification,” ”with Video,” “with Image,” “Rehabilitation,” “Therapy,” “Deep learning,” “Convolutional Neural Network,” and “CNN.”

3.3. Quality Assessment

The three authors screened and examined individual article abstracts and titles to determine whether the corresponding article fulfilled the selection criteria. If an article fulfills the inclusion criteria, the corresponding value of the paper is ”2” (value ”2” means selected for the next step) in the excel sheet, while ”0” is for the excluded article. On the other hand, if there is a confusion that an article might fulfill the criteria, then the article’s value is ”1” (one waiting for recheck).

3.4. Study Selection and Result

A total of 525 articles were collected after removing duplication, examined through abstract and title, and put the values either 0, 1, or 2. If an article has value one, we re-examined it by the other authors and discussed with all authors whether it would be included. In this screening phase, 220 articles were excluded due to failed criteria. Next, we tried to find the full text of 305 articles and determine whether they were focusing on other topics and excluded 60 articles. Finally, 130 articles were included to review from 2017 to 2022; Fig. 1 depicts the number of considered articles using deep learning with the image(s) or video-based features. Moreover, the above discussion is represented as a flow diagram according to PRISMA (Moher et al., 2009) in Fig. 2.

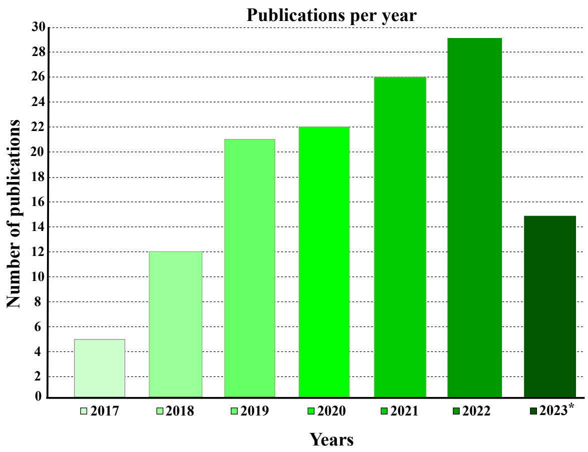

Fig. 1: The number of publications per year for ASD detection, classification, and diagnosis, along with rehabilitation using deep learning with images or videos.

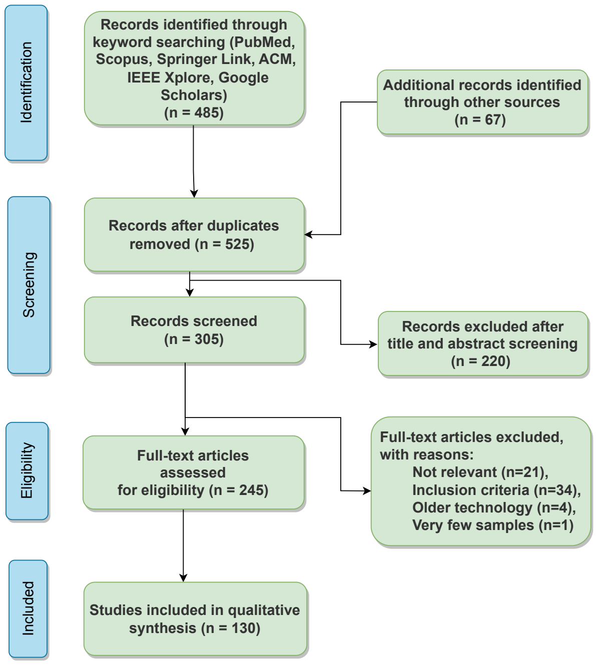

Fig. 2: PRISMA flow diagram for the article selection process.

4. Dataset and Performance Metrics

Many different datasets were developed in the literature to demonstrate autism research considering different modalities, including face, eye gaze, and MRI. An example including different modalities datasets is shown in Fig. 4.

4.1. Public Datasets

The publicly available datasets to analyze autism research considering different image modalities are summarized in Table 1.

4.1.1. Eye Gaze Datasets

Analysis of eye-tracking data is one of the basic fundamental approaches for detecting ASD. The Saliency4ASD (Duan et al., 2019b) is a public dataset of children with ASD's eye movements. Tobii T120 eye tracker was employed to collect the data with equally distributed ASD and TD (i.e., 14 subjects for each group). At the same time, they viewed 300 visual stimuli images of objects, natural scenes, animals, and persons. However, fixation maps and scan paths were collected from the participants. In addition, eye-tracking technology was used in (Carette et al., 2018) to translate dynamic eye movement into gradient color images and make a public dataset. The dataset includes 547 visual images of the eye scan path from 29 ASD and 30 TD participants, where 328 and 219 images are respective for ASD and TD. In another study, Chong et al. (2017) collected a video dataset for detecting gaze, which included 100 children in 156 distinct play sessions where children with ASD and TD are equally distributed. They collected data based on the interaction between an adult and a child. Finally, they annotated the image sequence (i.e., around 2 million frames) using video annotation software ELAN [1] and INTERACT Mangold [2]. Ground truth annotation considered flagging the image-level onset and offset of each instance of the participant making eye contact with the examiner. Its mentioned that a portion of their training data was already publicly available as part of the MMDB (Rehg et al., 2013) dataset.

4.1.2. Magnetic Resonance Imaging Datasets

Magnetic Resonance Imaging (MRI) is a non-invasive imaging tool that generates three-dimensional comprehensive anatomical images that differ significantly between ASD and TD. To expedite knowledge

[1] http://tla.mpi.nl/tools/tla-tools/elan/

[2] https://www.mangold-international.com/en/products/software/behavior-research-with-mangold-interact

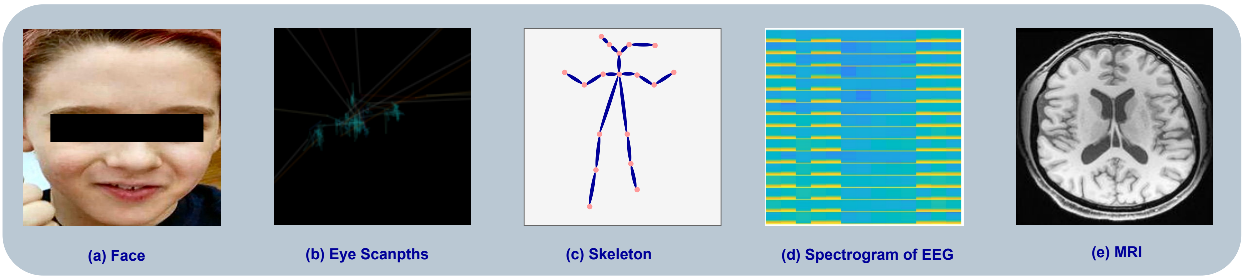

Fig. 3: Different modalities for the image(s)-based dataset for autism research: (a) A face image for an ASD child (Piosenka, 2021); (b) An image of scan path for an ASD subject (Carette et al., 2018); (c) An example of a skeleton image; (d) The spectrogram image of EEG signal (Tawhid et al., 2021); and (e) A single 2D MRI image of an ASD subject (Di Martino et al., 2017).

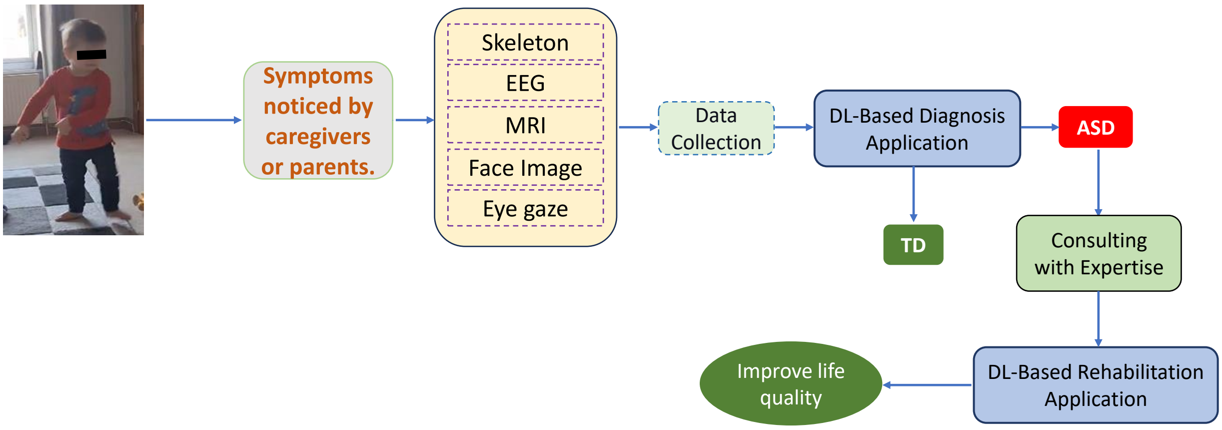

Fig. 4

of the neurological roots of autism, Autism Brain Imaging Data Exchange (ABIDE) (Di Martino et al., 2014, 2017) gathered functional and structural brain imaging data from laboratories throughout the world. ABIDE I and ABIDE II are two large-scale collections in the ABIDE effort. Each collection was built by aggregating datasets gathered individually across more than 24 international brain imaging laboratories and is now available to researchers.

ABIDE I (Di Martino et al., 2014) represents the primary version and involved 17 international sites, sharing earlier collected resting-state functional magnetic resonance imaging (rs-fMRI) data. Altogether, it included 1,112 records, where ASD and TD participants were 539 and 573, respectively, along with ages ranging from 7 to 64 years. Later, a more diverse large-scale dataset ABIDE II (Di Martino et al., 2017), was released, including 1,044 records, where ASD and TD participants were 487 and 593, respectively.

4.1.3. *Face Datasets*

The Autism Facial Image Dataset (AFID) is the only publicly available dataset for face images for autism research presented in (Piosenka, 2021). Images were collected from various websites and Facebook pages with equal distribution of ASD and TD. They proposed a benchmark protocol: 2,540 images for training and 300 and 100 for testing and validation, respectively.

4.1.4. *Multi-modal Datasets*

When systems use a single trait or modality for the detection, classification, or analysis of ASD is called uni-modal systems (Uddin et al., 2017) and are regarded as a conventional system because of their simplicity. These systems are, however, commonly affected by some problems, such as noisy sensor data and low accuracy. One solution to these problems is using data from multiple modalities, called multi-modal data. The De-Enigma (Shen et al., 2018) is a publicly available multi-modal (e.g., audio, depth, and video) dataset of autistic children that can be used for behavioral analy-

Table 1: Publicly available datasets for ASD detection, classification, diagnosis, and rehabilitation research.

Authors	Focus	Modality	Dataset	Age (Years)	Instance
Cai et al. (2022)	Extracted head-related features from video data	Multi-modal	N/A	1.6 - 13.0	Sub: 57 - ASD, 25 - TD
Piosenka (2021)	Collect ASD and TD images from various websites	Face	AFID	2.0 - 14.0	Image: 1470 - ASD, 1470 - TD
Billing et al. (2020)	Captured joint 3D skeleton, Head orientation and Eye gaze data during robot-enhanced therapy	Multi-modal	DREAM	3.0 - 6.0	Sub: 61 - ASD
Duan et al. (2019b)	Record of eye movement while watching image	Eye Gaze	Saliency4 ASD	8.0 (Avg.)	Sub: 14 - ASD, 14 - TD
Carette et al. (2018)	Captured eye movement to visualizations of eye-tracking scan paths	Eye Gaze	N/A	3.0 - 13.0	Sub: 30-ASD, 29-TD
Zunino et al. (2018)	Collect gesture data when grasping bottle from video	Multi-modal	N/A	9.6 (Avg.)	Sub: 20 - ASD, 20 - TD
Shen et al. (2018)	Multi-modal (audio,video and depth) data collected during robot-assisted activities	Multi-modal	De-Enigma	5.0 - 12.0	Sub: 128 Children
Chong et al. (2017)	Captured eye-gaze data during child-adult interaction	Eye Gaze	N/A	3.0 - 13.7	Sub: 50 - ASD, 50 - TD
Di Martino et al. (2017)	Increase sample size, greater phenotypic charaterization from of ABIDE I	MRI	ABIDE II	5.0 - 64.0	Sub: 521 - ASD, 539 - TD
Di Martino et al. (2014)	Collect and combine functional and structural brain MRI from various laboratories	MRI	ABIDE I	7.0 - 64.0	Sub: 539 - ASD, 573 - TD
Rehg et al. (2013)	Multi-modal (Audio, video, and physiological) data recorded from toddlers	Multi-modal	MMDB	1.0 – 2.0	Video: 160 Sessions; 3-5 mins
Rajagopalan et al. (2013)	Collecting videos of childrens naturalistic behaviours from public domain websites.	Multi-modal	SSBD	N/A	Video: 75; Avg. 90 sec.

sis. It included 62 British and 66 Serbian children aged 5 to 12 years who participated in De-Enigma studies on emotion recognition. Each child was randomly assigned to robot-assisted or adult-assisted activities with 4-5 sessions. In addition, it captured 152 hours of interaction resulting in 13 TB of multi-modal data.

Another publicly available multi-modal ASD dataset is the Development of Robot-Enhanced Therapy for Children with Autism Spectrum Disorders (DREAM) (Billing et al., 2020), which is also a behavioral dataset gathered from 61 children with ASD. Participating children undergoing robot-enhanced therapy, which consists of 3,000 therapy sessions. As a result, it captured around 300 hours of therapy. Three RGB cameras and two RGBD (i.e., the Kinect sensor) cameras were employed to capture the children's behavior. The main features extracted are ten joint 3D skeletons covering the upper body (head, shoulders, elbows, wrists, and hands), head orientation, and eye gaze. In addition, metadata: age, gender, and autism diagnosis are included in the dataset. Again, Cai et al. (2022) collected videos during social interactions. This study included 57 and 25 children with ASD and TD, with ages ranging from 1.7 to 13.0 years. Later, they extracted head-related features such as head position, head rotation, direction, eye position, eye gaze direction, facial position, facial action units, rigid face shape features, and non-rigid face shape features from videos using Openface (Baltrusaitis et al., 2018).

Similarly, Zunino et al. (2018) presented a video dataset containing gesture data from ASD and TD children. They collected 1,837 video sequences from 20 children for each group of ASD with an average age of 9.8 years TD with an average age of nine and a half a year. The participants were instructed to hold a bottle in hand and perform various tasks such as placing, pouring, passing to pour, and passing to place. In addition, Rajagopalan et al. (2013) collected videos from publicly domain platforms (e.g., Youtube, Vimeo, Dailymotion) of children employed in an uncontrolled natural setting. The dataset, i.e., the Stimulatory Behaviour Dataset (SSBD), consists of 75 videos, each 90 seconds long on average. All videos are divided into three categories, i.e., arms flapping, head banging, and spinning, and this public dataset is used for ASD diagnosis.

4.2. Private Datasets

Private datasets are not publicly available; in our systematic review, a good number of studies employed private datasets, which are discussed below.

4.2.1. Eye Gaze Datasets

Studies in (De Belen et al., 2021) considered 34 participants to collect data for the dynamic eye track, where half of them are ASD according to the criteria of the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) of the American Psychiatric Association. Using a Tobii X2-60 eye tracker, they collected fixations and saccades of eyes when participants stimulus by a different scene. Another scan path visualization data were generated using SMI RED250 eye-tracking technology in (Cilia et al., 2021) while participants watched videos or photos of different stimuli. Fifty-nine children participated in the study, where 29 children are ASD according to ADI-R, while 30 to TD.

Chrysouli et al. (2018) generated a dataset of eye movement from videos of children interacting with adults. A total of 43 subjects participated in constructing the dataset. After being possessed, 3,37,815 images were extracted from videos with only the subject's eyes. Individual gaze patterns were also extracted from the captured video. In (Li et al., 2020) employed the tracking-learning-detection approach to monitor eye movement in the video, where a total of 83 videos of ASD children were added, with 189 recordings of 53 and 136 children, respectively, for ASD and TD.

4.2.2. Face Datasets

Some of the studies developed their own face dataset. For example, Leo et al. (2018b) collected a face dataset for facial expression recognition with 17 children with ASD. Each child was asked to produce four facial expressions (i.e., happiness, sadness, fear, and anger) sequentially and capture the events using a video camera. Altogether, 17 videos were manually annotated based on whether the children produced correct facial expressions. Again, Rani (2019) collected 25 images of ASD from different internet sources with four emotions (i.e., angry, neutral, sad, and happy) for study. Another facial expression dataset was collected in (Han et al., 2018), where 25 participants produced seven different emotions. Finally, they managed 150 expressions images. Moreover, from online sources, Patnam et al. (2017) also collected an image dataset, including 4,000 images of gestures covering ears and faces. Furthermore, Shukla et al. (2017) constructed a developmental disorders dataset including ASD with 1,126 face images from various sources to recognize different developmental disorders. The dataset is annotated by age, gender, and type of developmental disorder. Similarly, Lu and Perkowski (2021) collected 1,122 face images to analyze the ethnic-racial factors. The images were collected from the same race with equally distributed ASD and TD.

Furthermore, Banire et al. (2021) collected a face dataset including 20 and 26 children, respectively, for ASD and TD. They captured 95 videos and labeled them using iMotions software as attention and inattention. Similarly, Tang et al. (2018) constructed a video dataset while the mother and infant interacted. A total of 34 participants were included, among them 11 ASD and 23 TD. They labeled 77,000 frames manually into a smile, non-smile, and occluded faces. In addition, Ganesh et al. (2021) constructed a dataset of thermal face images to detect ASD, including 50 children with ASD and TD.

4.2.3. Skeleton Datasets

Skeleton data consists of the 2D/3D coordinates of the human joints. Kojovic et al. (2021) made a video dataset of social interaction, including 136 subjects with equally distributed ASD and TD. Later, they extract skeletal key points from videos using OpenPose (Cao et al., 2017).

4.3. Performance Metrics

Evaluating the performance of a deep learning method is one of the crucial steps while designing a model. Many different metrics are used to assess the model's performance, and these metrics are known as performance metrics. For the classification models to classify ASD and TD, the accuracy, sensitivity or recall, specificity, Precision, and F1-score are employed and its acquired from the confusion matrix. An example of the confusion matrix is shown in Fig. 5. The accuracy gives the ratio of the correctly predicted observations out of the total number of tested observations. Moreover, sensitivity represents the ratio of correctly classified positive cases, while specificity defines the proportion of true negative data correctly classified. The precision gives the ratio of correctly predicted positive cases and true positives and false positives. The F1-score is a critical assessment statistic that is defined as the harmonic mean of the model's precision and recall.

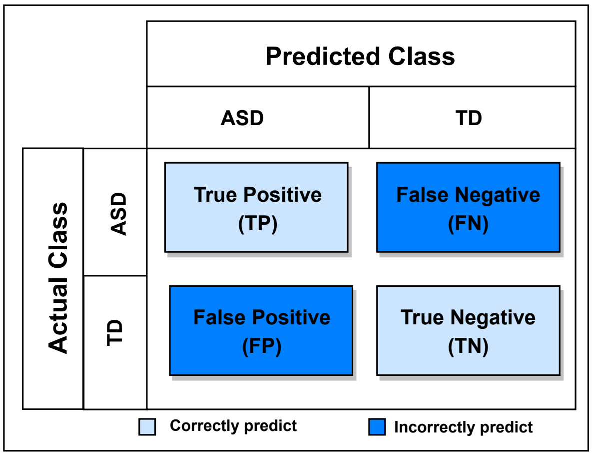

Fig. 5: An example of a confusion matrix for the classification of ASD and TD.TP: true positive; TN: true negative; FP: false positive; FN: false negative.

Table 2: Metrics are used to evaluate the performance of a method for detection, classification, and regression problems in autism research.

Metric	Equation	Puporse
Accuracy	$\frac{\|TP+TN\|}{\|TP+TN+FP+FN\|}$	Classification
Sensitivity/Recall	$\frac{\|TP\|}{\|TP+FN\|}$	Classification
Specificity	$\frac{\|TN\|}{\|TN+FP\|}$	Classification
Precision	$\frac{\|TP\|}{\|TP+FP\|}$	Classification
F1-score	$\frac{2(Precision)(Sensitivity)}{Precision+Sensitivity}$	Classification
MSE	$\frac{\sum_{i=1}^{N}(\bar{x}_i - x_i)^2}{N}$	Regression

$\bar{x}_i$: Observed sample; x_i: actual sample; N: number of samples.

Mainly, its value ranges from 0 to 1, where 0 indicates a bad prediction performance while 1 for excellent.

Furthermore, the area under the receiver operating characteristic (ROC) curve (AUC) is another metric used to evaluate the performance, which is a graph showing the evaluation performance of a method at all classification thresholds, where the curve plots the true positive rate versus the false positive rate while AUC measures the entire two-dimensional area underneath the entire ROC curve from 0 to 1. Moreover, the values of AUC range from 0.0 to 1.0. For example, if a method's predictions are 100% wrong has an AUC of 0.0, whereas 1.0 for 100% correct. Furthermore, some variants of AUC, such as AUC_Judd (Judd et al., 2009), and AUC_borji (Borji et al., 2012), are employed to evaluate the performance. The Mean Square Error (MSE) is also used to measure performance. These metrics calculations are summarized in Table 2.

5. Results of Features Found in Study

Feature extraction is finding key points or characteristics that can be used for further analysis, such as detection, classification, rehabilitation, etc.; it can be done manually or automatically. In manual feature extraction, a specialist recognizes the features and devises a strategy to extract them, while automatic feature extraction is carried out automatically (e.g., using a DL-based approach). A general framework for the feature extraction and classification using a DL-based approach and the handcraft-based feature is shown in Fig. 6. It can be noted that a large number of samples with known labels are first feed during training to learn a model. In the first step, various preprocessing techniques are employed, such as data augmentations and noise removal, and the features are extracted. Finally, the weight values of the DL-based method are updated to make a robust method to test an unknown sample during the test stage.

Convolutional Neural Networks (CNN) can be employed to extract features, which is a special feed-forward neural network (FFNN) including convolution, Rectified Linear Units (ReLU), and pooling layers along with a fully connected (FC) layer. The convolutional layers are good for feature extraction from the image and video as they deal with spatial redundancy by weight sharing. It includes at least one kernel to slide across the input and perform a convolutional operation between each input region and the kernel. The results are stored in the activation maps containing features extracted by different kernels, which can be considered the convolutional layer's output. Pooling, also known as downsampling, is a dimensionality reduction procedure. Usually, a pooling layer is inserted between a convolutional layer and the following layer. The FC layers are a basic hidden layer of FFNN where all the neurons from the previous layer are connected to every neuron in the final activation unit of the next layer. The softmax and sigmoid are the two most often utilized activation functions for ASD and TD classification. A simple CNN architecture is illustrated in Fig. 7.

Long Short-Term Memory (LSTM) is a variant of a recurrent neural network (RNN) (see section 6.5 for more details) that can recognize order dependency in sequence prediction problems and address the shortcomings of RNNs (e.g., handling long-term dependencies). Furthermore, LSTM addresses the vanishing gradient in sequence prediction problems. LSTM, combined with the extracted feature from CNN, can be used as a feature extractor and a classifier. Examples of a simple LSTM architecture and CNN-LSTM architecture are illustrated in Figs. 8 and 9 respectively.

5.1. Facial Expression or Emotion Features

Emotion recognition refers to recognizing a person's emotions, which include joy, sadness, anger, fear, disgust, and surprise. Facial landmarks can help to identify an individual's expressions or emotions; therefore, they were employed to classify ASD and TD. To find the facial landmarks, Banire et al. (2021) present a face-based attention recognition network because of its ubiquitousness compared to other methods. They employed the iMotions software to extract 34 facial landmarks from a video and then transformed them into geometric-based features using the Euclidian Distance. Similarly, in (Leo et al., 2018a,b; Del Coco et al., 2017) proposed methods to automatically analyze facial expressions produced by ASD children using conditional local neural field (CLNF) (Baltrusaitis et al., 2013) to recognize and track facial landmarks. The CLNF consists of point distribution models for capturing landmarks shape variations, while patch experts for appearance variations. Moreover, Wu et al. (2021) used facial key points such as face and eye landmarks, facial action units (AUs), head pose, eye gaze direction, etc., to predict the behavior of ASD children. They extracted those features using OpenFace (Baltrusaitis et al., 2018); finally, they detected a smile, a look face, and vocalization, while Cai et al. (2022) extracted head-related features. Moreover, they introduced a head-related characteristic attention mechanism to select the most discriminative features.

In addition, some studies used existing off-the-self CNN-based approaches for feature extraction. For example, Shukla et al. (2017) used pre-trained AlexNet (Krizhevsky et al., 2017) for feature extraction. During training, four parts of the segmented face and the original face image are fed into the networks. Similarly, Cao et al. (2023) separates the image into a number of patches. Each image patch was given positional encoding before using Vision Transformer (VIT) (Dosovitskiy et al., 2020) to automatically categorize ASD and TD. Again, (Mujeeb Rahman and Subashini, 2022; Alam et al., 2022) extract facial landmarks from images and identify children with ASD and TD using MobileNet (Howard et al., 2017), Xception (Chollet, 2017), and different versions of Efficient-Net (Tan and Le, 2019). Again, Hosseini et al. (2021), Alkahtani et al. (2023) and Akter et al. (2021) employed the MobileNet, while Alsaade and Alzahrani (2022) used Xception and VGG-19.

Moreover, Rabbi et al. (2023) also used VGG-19 along with Inception-V3 and DenseNet-201 for facial feature extraction and classification.

Besides RGB images, thermal imaging can also be employed to extract facial landmarks to analyze ASD and TD. For example, Ganesh et al. (2021) used thermal images focused on the forehead, eyes, cheek, and nose thermal variations, varying between ASD and TD. Similarly, in (Tamilarasi and Shanmugam, 2020) also used thermal images in the ResNet-50 networks for facial feature extraction and classification.

Some soft attributable features can be used together with the extracted feature from the image to influence the classification performance for individuals with ASD and TD. For example, Lu and Perkowski (2021) observe racial factors play a vital role in classifying ASD and TD from facial images. They demonstrated their experiment on their own dataset called East Asian ASD facial image datasets and publicly available datasets in the Kaggle repository, the AFID (Piosenka, 2021), and a mixture of these two datasets. The East Asian datasets included people of the same race, while the AFID differed. They found that classification accuracy is better in East Asian datasets due to symmetry in ethnic characteristics.

Furthermore, gesture characteristics can also be used for emotion recognition. For example, Patnam et al. (2017) develop a system that can recognize the meltdown action of kids with ASD. They collected all the meltdown gesture data (i.e., covering ears, covering the face, biting hands, and flapping hands) from various sources. Two instances of the recognized behavior covering the face and the ears were used to identify meltdowns.

5.2. Eye Gaze Feature

Children with ASD may exhibit atypical patterns of gaze perception due to disruptions in their early visual processing. Therefore, its probable to evaluate an observer who has ASD based on their eye-gaze feature. Tao and Shyu (2019) established a framework for identifying ASD and TD based on the observer's scan paths at a given image. First, the saliency prediction model for a certain image from ordinary people generates a reference saliency map based on SalGAN (Pan et al., 2017), a pre-train saliency prediction CNN network. The image patches of the predicted saliency map are then constructed based on the given scan path. Finally, patch features are employed in the proposed approach to classify ASD and TD children. Review 3, Question:1,

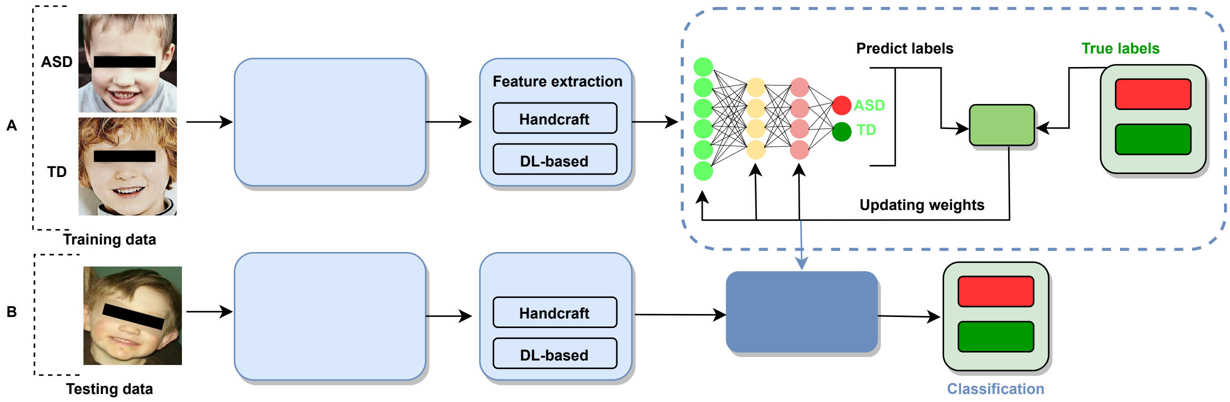

Fig. 6: A general framework for image(s) or video-based deep learning approach for the classification of ASD and TD: (A) Training phase to train the network using known data with true labels; (B) The trained model is employed to classify the unknown samples. Example input images were taken from AFID dataset (Piosenka, 2021).

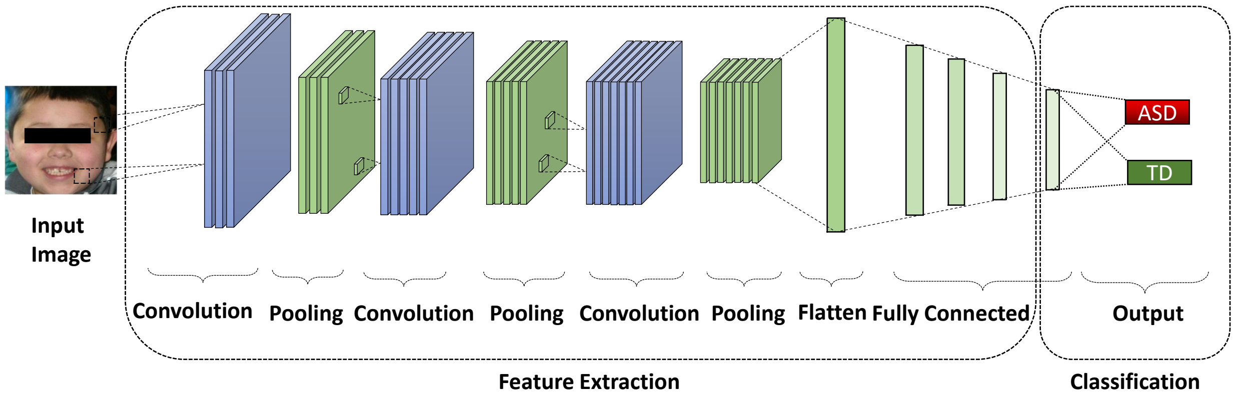

Fig. 7: The basic structure of a CNN-based approach. A stack of convolutional and pooling layers is used for learning features from images. FC layers classify these features gathered from the convolutional layer.

part:1 (atayabi2023stratification) Atyabi et al. (2023) combines spatial information (eye-gaze scan-paths) and temporal information (velocity of eye movement) to classify ASD and TD. Similarly, Wei et al. (2021) extracts the spatiotemporal feature combined with the scan path for classification, which outperformed the above approach mentioned in (Tao and Shyu, 2019).

Eye movement data can also be used to identify ASD. Liaqat et al. (2021) employed a synthetic saccade pattern model (Wloka et al., 2017) to represent the baseline combined with the original scan pattern along with many auxiliary data for classification. They forwarded the image and processed fixation sequences as data points for classification. On the other hand, Cilia et al. (2021) transformed the eye-track data into a visual representation that binds into a set of images. Later, this set of images and their corresponding feature is further used to analyze ASD. The scan path features may use as a biomarker for classifying individuals with ASD and TD (Kanhirakadavath and Chandran, 2022; Xia et al., 2020) and can also be fused with several other attributes like temporal information and pupil velocity data (Atyabi et al., 2022).

Some studies analyze an individual's dynamic gaze patterns while viewing a natural image (i.e., visual attention) and may demonstrate the salient region of a particular image to analyze ASD and TD. For example, Fang et al. (2019) make a saliency map and demonstrate that there is a difference between ASD and TD eye fixation maps. They diagnose ASD by comparing the fixations map and utilizing an objective loss function with the PN-MSE (i.e., the positive and negative equilibrium mean square error), which helps identify the salient regions. Similarly, Jiang and Zhao (2017) present an approach for analyzing eye movement patterns in ASD and TD individuals while free viewing natural images. They generated a Fisher score (PEH, 2001) of the images, indicating that the most crucial feature is at the top because Fisher's score placed the data of the same type closely, while data of different types were set far apart to generate discriminative features. Finally, these discriminative features are further employed to diagnose ASD. Similarly, Wei et al. (2019) also proposed a model for saliency prediction using multi-level features extracted using a CNN-based approach and fused them for further prediction of ASD and TD. Moreover, De Belen et al. (2021) employed ACLNet (Wang et al., 2018) networks, a combination of CNN and LSTM used for feature extraction from eye movement.

Gaze patterns can also be found in daily-live social activities like interacting with others, hearing sounds, making eye contact while talking, and playing games. By employing these daily-live social activities, Chong et al. (2017) analyzed eye contact and implicitly estimated head pose from a child's naturalistic dynamic social interaction video to detect the individual with ASD. Meanwhile, the VGG-face model (Parkhi et al., 2015) was employed to extract facial features. Finally, the extracted features are forwarded to further classification.

In addition, Eye gaze data can also be used to analyze an individual's affective state. For example, Chrysouli et al. (2018) explore individuals' affective states (e.g., bored, frustrated, engaged, etc.) while interacting with a computer by using the flow of eye movement. They extract individual faces using IntraFace (De la Torre Frade et al., 2015) from the videos. Furthermore, it provides 49 facial landmarks points, where six key points are neighbors of the eye. They crop the image, which contains only the subject's eye, with the help of these six key points. Finally, these eye images are further used to find an affective state, which helps to analyze ASD.

5.3. Skeleton Feature

The behavior feature extraction from eye gaze and face has some limitations. For example, data must be collected in a controlled environment. Therefore, its troublesome to collect data, particularly to capture data from children. To overcome this problem, Kojovic et al. (2021) extract features from video while adults and children socially interact at a distance. They extract the skeleton using DL-based multi-person pose estimator OpenPose (Cao et al., 2017) network from social interactions video. Similarly, in (Marinoiu et al., 2018), the video was captured during robot-assisted treatment sessions with autistic children in an uncooperative environment; further, it was employed to develop action categorization and emotion prediction. They used high-level 3D pose and shape features to comprehend the children better. In their framework, they employed the modified Deep Multi-task DMHS (Popa et al., 2017) network for fully automatic 2D and 3D human sensing with feedforward and feedback components to get a 3D skeleton.

Action can be recognized from the skeleton key point to identify individuals with ASD. For example, Pandian et al. (2022) employed the skeleton points for action recognition, combining raw videos and key points of the skeleton for detecting action. They used the high-resolution network (Wang et al., 2020a) for the pose estimation to generate the key point and limb in the form of a heatmap. Finally, the heatmap and raw videos are passed to their network to recognize the actions.

5.4. Electroencephalography Feature

Electroencephalography (EEG) is an examination system to identify abnormalities in a person's brain waves that can be used to analyze ASD. For example, Tawhid et al. (2021) developed a model that can classify an individual with ASD and TD based on time-frequency spectrogram images of EEG signals. They preprocessed raw EEG signals using common average referencing, infinite impulse response filter, and normalization. After preprocessing, they segmented each signal and employed a Short-time Fourier Transform (STFT) on each segment to get the images. Finally, its employed for the classification of ASD and TD. Again, Baygin et al. (2021) used a deep lightweight feature extractor for ASD detection from EEG signals. They employed a one-dimensional Local Binary Pattern (LBP) to generate features from a one-dimensional signal. Then, these features fed an input to the STFT to generate an image of an EEG signal. Later, MobilNetV2 (Sandler et al., 2018), SqueezeNet (Iandola et al., 2016), and ShuffleNet (Hluchyj and Karol, 1991) were employed to extract discriminative feature from these images. Furthermore, in (Mayor-Torres et al., 2021; Torres et al., 2022) used EEG images to classify facial expressions of ASD and TD children.

5.5. Magnetic Resonance Imaging (MRI) Features

A medical imaging technology, Magnetic Resonance Imaging (MRI), is a non-invasive imaging technology that generates three-dimensional anatomical images and can differentiate between normal and abnormal tissue. Eventually, its employed to detect, classify, diagnose and analyze ASD. A three-dimensional MRI image is mixed with many layers and is a complete package of structure and function. It, therefore, is challenging to

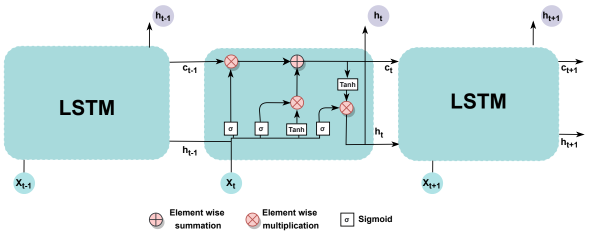

Fig. 8: A simple architecture of LSTM. Here, each LSTM unit has three inputs (h_{t-1}, c_{t-1}, and x_t) and two outputs (h_t, and c_t). For a given time t, c_t is the hidden state, h_t is the cell state, and x_t is the current input. The first sigmoid layer σ has two inputs: the hidden state from the previous cell h_{t-1}, and current input x_t.

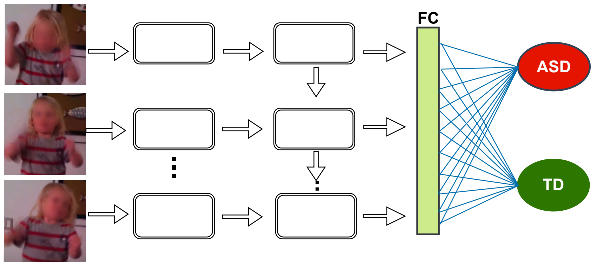

Fig. 9: An example of CNN-LSTM architecture. Here, CNN is generally used for feature extraction, and LSTM is used for classification.

consider a whole brain at a time. Hence, its employed as an atlas that can define the shape and location of brain regions in a common coordinate space. Further, an atlas can parcellate the brain image into several Regions of Interest (ROIs). Finally, the feature extracted from the time series of each ROI to analyze ASD. Anatomical, functional, and data-driven atlases are commonly used to generate ROIs. Some popular atlas used in the literature: Bootstrap Analysis of Stable Clusters (BASC) (Bellec et al., 2010), Craddock 200 (CC200) (Craddock et al., 2012), Craddock 400 (CC400) (Kunda et al., 2020), Dosenbach (DOH) (Dosenbach et al., 2010), Power (Power et al., 2011), Automated Anatomical Labeling (AAL) (Tzourio-Mazoyer et al., 2002), Harvard-Oxford (HO) (Desikan et al., 2006), Talaraich and Tournoux (TT) (Talairach, 1988), Eickhoff-Zilles (EZ) (Eickhoff et al., 2005), Multi-Subject Dictionary Learning (MSDL) (Varoquaux et al., 2011).

The AAL atlas has divided the brain's cerebrum into parcels by anatomical landmarks in which various labeling nodes are generated manually in different versions. Li et al. (2018a) employed an AAL atlas to resting state-functional MRI (rs-fMRI) data to classify ASD and TD. They calculate the brain Functional Connectivity Matrix (FCM) as a 90 × 90 adjacency matrix representing the connection between each pair of ROI. The Pearson Correlation Coefficients (PCC) of the ROI pair determine the cell weight of the FCM. Finally, they extracted 4,005 dimensional features from the FCM. They forwarded it to classify while Wang et al. (2019a) employed AAL to generate 6,670-dimensional feature vectors by PCC and Fisher's z transformation to classify the individual with ASD and TD. Similarly, Lu et al. (2022) used the AAL atlas to analyze rs-MRI's instability, leading to FCM ambiguity; eventually, its impairs ASD diagnosis. Therefore, they employed the Takagi-Sugeno-Kang Fuzzy inference systems to decrease the uncertainty and instability of rs-fMRI. In addition to FCM, in (Al-Hiyali et al., 2021b,a), they used the scalogram image from the AAL atlas. A continuous Wavelet transform generates the scalogram images and extracts dynamic temporal features to detect ASD. Furthermore, Tang et al. (2020) took the AAL atlas and the full-brain connection matrix into consideration. The full-brain connectivity of functional magnetic resonance imaging (fMRI) voxels and the ROIs correlation matrix was employed to extract the feature to analyze ASD. The community structures are more efficient in diagnosing ASD than PCC. For example, Liao and Lu (2018) implements a normalized mutual information statistic matrix considering the AAL atlas and achieves better accuracy than PCC.

The Craddock atlas is a data-driven parcellate approach that takes whole-brain rs-fMRI. In (Heinsfeld et al., 2018; Almuqhim and Saeed, 2021) generated the FCM using the CC200 atlas with 200 ROIs while PCC determined the value of each cell in the matrix to indicate brain regions strongly linked to anti-correlated ones. Afterward, they generated 19,900-dimensional functional connectivity features. Similarly, Zhang et al. (2022b) also used the CC200 atlas to generate FCM and employed the Fisher score selection method to select the top features to detect ASD. Some studies (Sherkatghanad et al., 2020; Zhang et al., 2022a; Yang et al., 2020; Othmani et al., 2023; Wadhera et al., 2023) extracted feature from CC400 atlas, where Sherkatghanad et al. (2020); Wadhera et al. (2023) generated 400 ROIs and make a 392 × 392 FCM where PCC or ROI average time series are employed to describes the weight of FCM. Finally, this FCM was forwarded to the classifier to classify ASD and TD. In addition, Zhang et al. (2022a) extracted 76,636-dimensional features from 392 ROIs by PCC. Later, they employed the step distribution curves to select 3,170-dimensional features for classification, while Yang et al. (2020) extracted 77,028-dimensional features using PCC for classification. Furthermore, Kiruthigha and Jaganathan (2021) extracted features from the CC400 atlas and explored 3D CNN to reduce the dimensions of 3D volume data.

The Power atlas comprises the cerebral cortex, subcortical tissues, and cerebellum for generating ROI. Yin et al. (2021) followed the Power atlas to parcellate the brain region, which consists of 264 ROIs for time series extraction. The weight of the brain network is defined by PCC, which shows the relation of two ROIs' time series data. Finally, these features were employed to further analysis.

The BASC altas is another data-driven fMRI atlas that employs unsupervised clustering to parcellate the whole brain. For example, Bayram et al. (2021) employed a BASC atlas of 122 ROIs to generate a connection matrix. These connectivity matrices were used to further the classification of ASD and TD.

The HO altas encompasses structural regions in the cortical and subcortical brains obtained from structural data and segmentations. Cao et al. (2021) utilized the HO atlas and generated features by PCC and Fisher z transformation. They reduced the dimensions of the feature vector by the recursive feature elimination process; finally, low dimensional features are further used for the analysis of ASD.

Multiple atlases are also considered in the literature to analyze ASD. For example, Subah et al. (2021) built an FCM feature using tangent-embedded atlases and compared it with different structural and functional atlases (e.g., BASC, CC200, AAL, and Power atlases) where AAL is a structural atlas, and the remaining are functional. Similarly, Wang et al. (2022) explored six (e.g., AAL, EZ, HO, TT, CC200, and DOH) atlases and generated features using PCC. Later they employed the support vector machine (SVM)-recursive feature elimination method to reduce the dimensions of the features. On the other hand, Yang et al. (2022) consider one structural atlas, such as AAL, and five other functional atlases, such as CC200, CC400, Power 264, BASC 197, BASC 444. They employed canonical ICA and dictionary learning to generate an FCM, which was used later for classification. Additionally, to generate an effective FCM for each individual, Pavithra et al. (Pavithra et al., 2023) brings time series from 48 regions of interest identified by the HO atlas and 122 regions of interest established by BASC. These features are supplied to the identification model.

Personal attributes can also be explored along with atlases to analyze ASD. For example, Niu et al. (2020) employed the AAL atlas, HO atlas, and CC200 to generate 90×90, 110×110, and 200×200 connectivity matrices, respectively, using PCC. Along with the extracted feature, they combine personal attributes such as sex, handedness, full-scale, verbal, and performance IQs to classify the individual with ASD and TD.

Similarly, in (Rathore et al., 2019) combined CC200 and CC400 atlas features with topological features, including persistence pictures, landscapes, and diagrams. Again, Mellema et al. (2019) also used seven atlases (HO, Power, MSDL, CC, and variation of BASC) to represent an FCM by projecting into tangent space along with structural data for classification.

Deep learning can also be explored to extract the highly discriminative features from MRI to analyze ASD. For example, Elakkiya and Dejey (2022) employed Bernoulli Restricted Boltzmann Machine (RBM) to extract features from fMRI, while Kashef (2022) used CNN to diagnose ASD. On the other hand, Li et al. (2022) employed the 3D ResNet (Tran et al., 2018) to extract features to diagnose ASD. Furthermore, the LSTM can also be used to extract features to analyze ASD. For example, Liu et al. (2021) extracted abstract features, which are then fed to an autoencoder (see Sec. 6.4 for further details) to extract final features to diagnose ASD, while Kang et al. (2022) extracted dynamic spatiotemporal features using LSTM along with CNN. Similarly, in (Jiang et al., 2022b) attempt to keep both the spatial and temporal features. They extracted spatial information from fMRI using CNN and a series of spatial characteristics input into a Gated Recurrent Unit (GRU) to extract temporal data. In addition, some works employed two-stage networks to extract features. For example, Li et al. (2018c) explored a 2-stage network to distinguish between ASD and TD and clarify the brain biomarkers. They employed a frequency sampling method that corrupts the original image. The corrupted image is forwarded to CNN, which helps to find brain biomarkers with the discriminative feature, where they used the AAL atlas with 116 ROIs. Nogay and Adeli (2023) devised a two-stage strategy for categorizing ASD and TD based on sMRI images that includes preprocessing and a grid search optimization algorithm which was applied to deep CNN.

Morphological technique can also be employed to extract features to classify the ASD, and TD (Gao et al., 2021; Sharif and Khan, 2022). Sharif and Khan (2022), for example, utilized morphological data from the corpus callosum and intracranial brain volume to differentiate ASD from TD, while in (Pugazhenthi et al., 2019) segmented brain images into the white matter, gray matter, and cerebrospinal fluid by their threshold values; finally these segmented images along with original images were fed into the classifier. In addition, the shape features from rs-MRI also contribute to the diagnosis of ASD. For example, Ismail et al. (2017) merged eight lobes from the cerebral cortex and cerebral white matter to obtain 64 attributes of shape variants per sample where each element is represented by its cumulative distribution function, which generates 64×4000 points that are subsequently classified as ASD and TD.

5.6. Multi-modal Feature

Multi-modal features extracted from the multiple modalities as already discussed in Sec. 4.1.4 for the detection and classification as well analyze autism. For example, Chen and Zhao (2019) proposed a privileged modality framework for classifying individuals with ASD and TD that combines two distinct modalities of visual attention. The first is a photo-taking task where participants are instructed to take photos in various scenarios. The second is an image-viewing task where participants' eye fixations were extracted. These two modalities are then fed into CNN-LSTM architecture to extract features and classify ASD and TD. Moreover, Javed and Park (2020) used human movement and facial key points features to identify the risk of ASD. They extracted facial key points, and body tracking data using OpenPose (Cao et al., 2017) and then used laban movement analysis (Groff, 1995) to derive movement features; eventually, they extracted three movement features and 68 facial key points. In addition, Duan et al. (2019a) explored atypical and typical features to compare the visual attention of ASD and TD children from the facial images. Similarly, Saranya and Anandan (2021) developed a framework that combined facial expression and gait, which can be defined as the manner of walking for a person (Uddin et al., 2019) to predict ASD. The gait features are extracted from video data: heel strike, foot flat, mid-stance, heel-of, pre-swing, terminal swing, and mid-swing.

In addition, Wang et al. (2019b) explored the gesture and eye gaze as the primary criterion for judging the performance of the expressing needs with the index finger pointing task, which helps in early diagnosis of ASD. Where gaze is estimated by a combination of eye center localization (Daugman, 1993) and head pose estimation (Baltrušaitis et al., 2016). On the other hand, the gesture is recognized by the single-shot detector algorithm. Similarly, Ali et al. (2022) try to understand behaviors (e.g., clapping, arm-flapping, to-taste, jump-up, headbanging, and spinning) of children to help the diagnosis of ASD. They extract several features from raw human-human or human-object interaction video. They employed YOLO-V5 for person detection, followed by DeepSORT (Wojke et al., 2017) for tracking and recurrent all pairs of field transformers for optical flow. Finally, the optical flow features, along with RGB image further employed to classify ASD and TD.

Some studies combine EEG along with other modalities such as eye tracking, facial images, etc. For example, Han et al. (2022) explored the EEG signal and eye-tracking features to identify ASD. They extracted the relative power energy, multiscale entropy, and brain network for the EEG signal; on the other hand, a TX300 eye tracker was used to record eye gaze data and extract 96-dimensional eye-tracking features. Similarly, in (Haputhanthri et al., 2020) explored the EEG feature together with the feature of facial thermography to classify ASD and TD, where the standard deviation and Shannons entropy are calculated from EEG features, while the mean temperature of nine ROIs in facial thermographic images was selected for facial features. Finally, these extracted features were fused in a feature-level fusion for the classification of ASD and TD.

6. Deep Learning-based Methods

Deep learning can be employed to extract features and classifiers along with feature extraction and classification in an end-to-end manner. The feature extraction procedure was explored in Sec. 5. Here we will explore the deep learning-based method for classification and summarized in Table 3.

6.1. Artificial Neural Networks

Artificial Neural Networks (ANN), also referred to as Feed-Forward Neural Networks (FFNN), are modeled after biological neurons to mimic how they communicate with one another in the human brain. It has three layers: an input layer, one or more hidden layers, and an output layer. There are no FC layers and only one direction of travel for input data. On the other hand, the Multilayer Perceptron (MLP) is a type of FFNN in which every layer is FC and can backpropagate. MLP serves as the fundamental building block for more advanced deep-learning architectures.

Some of the studies (Rani, 2019; Ahmed et al., 2022a) explored ANN for the classification of ASD and TD. Among them, Ahmed et al. (2022a) explored ANN of 126 input layers followed by ten interconnected hidden layers, and finally, two classes were produced. They achieve the classification accuracy of 99.8% over figshare repository[3]. On the other hand, MLP network explored in (Niu et al., 2020; Haputhanthri et al., 2020) where Niu et al. (2020) construct MLP with five layers.

[3]https://figshare.com/articles/dataset/Visualization_of_Eye-Tracking_Scanpaths_in_Autism_Spectrum_Disorder_Image_Dataset/7073087/1

Among these, one dropout layer with an input size of 4,005 and four dense layers of size 1,024, 512, 128, and 32, respectively, were employed.

6.2. Deep Neural Network

Deep Neural Network (DNN) is an ANN with more than one hidden layer between the input and output layers. In DNN, each node is connected with every node of the previous and forward layers. It takes an input and has some FC layers to process inputs to get the final output. In each layer, artificial neurons learn to extract increasingly abstract features from input data which increases their strength. Subah et al. (2021) employed a DNN classifier using the preprocessed rs-fMRI data, including two hidden layers, each with 32 neurons with a dropout value of 0.8 between each layer. Finally, a sigmoid activation function was used in the output layer to predict ASD. They achieved an accuracy of 88.0% on ABIDE I dataset. In another study, Yang et al. (2022) implemented DNN with eight hidden layers sizes 2,600, 2,048, 1,024, 512, 256, 128, 64, and 32, respectively, to reduced dimension. Finally, an output layer with a softmax activation was employed for classification. They achieved an accuracy of 68.4% using the same ABIDE I dataset.

Deep Belief Network (DBN) (Hinton, 2009) is a probabilistic generative model composite of the N number of Restricted Boltzmann Machine (RBM). The DBN was trained in two phases; firstly, it reconstructed input in an unsupervised manner and, finally, fine-tuned using a supervised way. Lu et al. (2022) construct a DBN using three hidden layers with dimension sizes of 512, 256, and 128. Finally, the output sizes of two with softmax activation. They achieve an accuracy of 68.6%, a sensitivity of 67.1%, and a specificity of 70.0% on ABIDE I dataset. Similarly, in (Huang et al., 2020) stacked three hidden RBM to implement DBN and achieved an accuracy of 76.4% on ABIDE I dataset. In addition, Bhandage et al. (Bhandage et al., 2023) employed an Adam war strategy optimization (AWSO) based DBN. The AWSO is designed by the Adam optimizer integrated with War Strategy Optimization. They used ABIDE I and ABIDE II datasets by varying training sets and got accuracy, sensitivity, and specificity of 92.4%, 93.0%, and 93.5%, respectively using ABIDE I dataset.

6.3. Convolutional Neural Network

A Convolutional Neural Network (CNN) is a form of ANN primarily intended to analyze pixel input and used mainly in image and computer vision (Yoo, 2015; Roy and Bhaduri, 2023), biometrics (Uddin et al., 2018), Natural Language Processing (NLP) (Wang and Gang, 2018), and medical imaging (Anwar et al., 2018). We already analyzed how CNN can extract features for autism research in Sec. 5 along with a basic feature extraction and classification structure by using CNN in Fig. 7. Here we will explore how CNN can be explored in autism research as a classifier and an end-to-end network together with feature extraction and classification.

A simple CNN network consisting of convolution, pooling, and FC layers can be used to feature extractors and classifiers. Sherkatghanad et al. (2020) constructed Functional Connectivity Matrix (FCM) between pairs of ROIs into a simple CNN with one convolutional layer, max-pooling, and densely connected layers along with sigmoid activation. They demonstrated their model on ABIDE I dataset and achieved a classification accuracy of 70.2%. Similarly, Bayram et al. (2021) developed a CNN model with nine layers, including convolutional, dropout, and max-pooling, and the FC layers along with the sigmoid activation. They demonstrated on ABIDE I dataset and achieved a classification accuracy of 70.2%. Furthermore, Marinoiu et al. (2018) construct a CNN model which takes temporal series of 3D skeletons as the input obtained from a Kinect sensor. Their CNN model consists of convolutional, pooling layers that are repeated twice, and lastly, an FC layer is added for action recognition; and explored their proposed model on the DE-ENIGMA dataset and achieved an accuracy of 53.1%. On the other hand, Mishra et al. (Mishra and Pati, 2023) proposed an ensemble model of CNN with different optimizers. The ensemble model of CNN with Adam and Nadam optimizer has achieved an accuracy of 81.3%, 77.6%, and 77.5% on the train-test ratio of 90:10, 80:20, and 70:30, respectively a total of 975 samples from ABIDE I dataset.

Existing pre-train CNN model can be explored to extract features and perform classification with learned weight using a large-scale dataset. Usually, It takes less training time and effort to develop the model's architecture. Mujeeb Rahman and Subashini (2022) studies five pre-trained CNN models, i.e., the XceptionNet, MobileNet, and different versions of EfficientNet, to extract the facial landmark feature. Then the feature is forwarded to the DNN, consisting of a hidden layer and a sigmoid activation used as a classifier. They demonstrated that the XceptionNet achieved the best AUC at 96.6% on the AFID (Piosenka, 2021).

In addition, some studies explored different versions of the VGG network (Simonyan and Zisserman, 2014b), one of the simplest and most popular pre-trained models. Lu and Perkowski (2021) employed a modified architecture of the pre-trained VGG-16 model followed by two hidden dense layers and a dropout layer to avoid possible over-fitting along with ReLU as an activation during training. They achieved a classification accuracy of 95.0% on their privately collected dataset, the East Asian Dataset (Lu and Perkowski, 2021). In (Jiang and Zhao, 2017) introduced a network architecture for analyzing eye-tracking data and learning discriminative features from images that follow the network of SALICON (Huang et al., 2015). It uses two parallel VGG-16 networks to process the input image. The first network uses the original image, while the second uses images that reduce the size by half of the original one. Finally, the concatenated features of 1,024 dimensions followed by the SVM classifier (Cortes and Vapnik, 1995) to classify the individuals with ASD and TD. They demonstrated their framework on the eye tracking dataset (Wang et al., 2015) where images were collected from OSIE dataset (Xu et al., 2014) and achieved an accuracy of 92.0%.

Multi-stream CNN architecture can be employed to extract more than one feature to analyze ASD. Chrysouli et al. (2018) used two-stream CNN architecture consisting of spatial and temporal blocks based on the model described in (Simonyan and Zisserman, 2014a) for recognizing affective state. The spatial one handles eye images, and the temporal one takes the eye's motion and merges them later. It achieved an accuracy of 95.3% for the privately collected dataset for the two classes (engagement vs. non-engagement) while 92.7% for the three classes (engagement, boredom, and frustration) classification.

In addition, Chong et al. (2017) proposed the Pose-implicit CNN (Pi-CNN) model, which jointly extracts the head pose and eye contact feature for analyzing ASD. They employed the modified AlexNet (Krizhevsky et al., 2017) architecture with a smaller kernel (i.e., 7x7 instead of 11x11) with stride 2 to find out more details about the face. The model generates two branches from the fully connected (FC) layer, one for the head pose and another for eye contact classification. They achieved the best F1-score, precision, and recall at 78.0%, 75.0%, and 80.0%, respectively, compared with the state-of-the-art in the literature (Krizhevsky et al., 2017; Smith et al., 2013; Ye et al., 2015; Rehg et al., 2013) using a publicly available MMDB (Rehg et al., 2013) dataset (see Table 4.1 for more details about the datasets). Similarly, Li et al. (2018c) explored a two-stage CNN to classify ASD and analyze brain biomarkers in ASD. The first stage is the framework of DNN (Li et al., 2018b) consisting of six convolutional, four max-pooling, and two FC layers, and lastly, a sigmoid output layer for the classification of ASD and TD. The second stage uses the anatomical structure of the brain fMRI to analyze the brain's bio marks. Here, they corrupt the ROI of the image and put it into a well-trained DNN to find a prediction to help to develop the importance of ROI. It achieved an accuracy of 87.1% over the ABIDE I dataset for the classification of ASD and TD.

The Region-based Convolutional Neural Network (R-CNN) (Girshick et al., 2015) was also employed in the research of autism. The key concept behind the R-CNN is a series of regional proposals. Region proposals are used to localize objects within an image. Patnam et al. (2017) reconstruct R-CNN for recognizing meltdown action. They added a classifier layer and achieved an accuracy of 92.0% in their custom dataset, which is 30.0% better than the benchmark R-CNN. Again, Prakash et al. (2023) used hand/finger-pointing annotated images from various hand gestures for training the R-CNN model for joint attention tasks. They achieved 93.4% accuracy for the detection of whether a child points to someone or something.

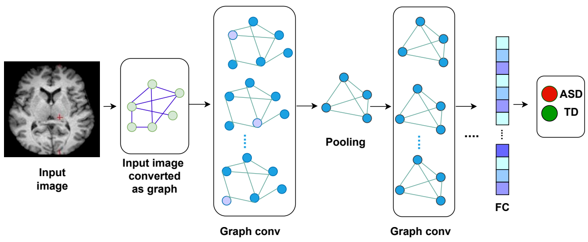

Fig. 10: An architecture of Graph Convolutional Network (GCN). First, input images are converted into a graph structure, i.e., nodes and edges. After some convolutional operations, the FC layer is added for the classification of ASD and TD.

Graph Convolutional Network (GCN) is a class of CNN for semi-supervised learning on graph-structured data, and it may operate directly on graphs and utilize structural data (Jiang et al., 2022a). A simple GCN architecture is illustrated in Fig. 10. Some variants of GCN were also explored for the analysis of autism research. For example, Wang et al. (2022) construct six graphs from six different atlases of the brain and then perform graph convolution operation on each graph. Finally, they achieved a classification accuracy of 75.8% on ABIDE dataset, while Wen et al. (2022) employed multi-view GCN that combines graph structure and multi-task graph embedding learning to improve classification performance. They achieved an average accuracy of 69.3% over ABIDE dataset.

Park and Cho (2023) developed a model that uses functional brain connectivity between STS and the visual cortex to diagnose ASD. First, it extracts both the spatial and temporal features from 4D fMRI brain images using residual CNN and Bi-LSTM with self-attention. These features are then converted into FCM to use by the GCN. They achieved an accuracy of 97.6% over the ABIDE-I dataset.

6.4. *Autoencoder*

Autoencoder (AE) (Rumelhart et al., 1985) is a simple FFNN consisting of input, output, and hidden layers divided into two phases, i.e., encoder and decoder. It firstly downsamples the whole input into a lower dimension by using input relation into the encoding phase, called latent space, while in the decoding phase, this down-sampled latent feature is upsampled to reconstruct the input as output. During the upsampling, it can produce novel samples with similar characteristics to the original data. However, the latent space features are used in the analysis of ASD research. A simple AE architecture is illustrated in Fig. 11.

Yin et al. (2021) developed an AE-based diagnostic approach. The first three layers of the AE are input and hidden layers (latent space representation), followed by a DNN classifier with a softmax activation. They demonstrated that the accuracy of the pre-trained AE along with a pre-trained DNN classifier is 3% more than without a pre-trained DNN classifier using ABIDE I dataset. Similarly, Sewani and Kashef (2020) employed AE for feature extraction and CNN as a classifier. They achieved an accuracy of 84.0% over ABIDE I dataset, while Mostafa et al. (2019b) achieved accuracy at 79.2% by employing AE followed by a pre-trained DNN classifier.

Sparse Autoencoder (SAE) (Ng et al., 2011) is a variant of the AE which uses sparsity to create an information bottleneck. Almuqhim and Saeed (2021) implemented an SAE which takes 9,500-dimensional features as input and reduces them to 4,975-dimensional features in bottleneck layers, followed by a DNN of three layers with the size of 2,487, 500, and 2, respectively, along with a softmax layer for classification.

Stacked Autoencoder is stacked by N number of AE where the output of i^{th} AE acts as input of $(i+1)^{th}$ AE. Studied in (Kong et al., 2019) explored a DNN-based model consisting of input, two SAE, and output, where SAEs reduce the feature dimension and extract hidden features, followed by a softmax in the output layer. They demonstrated their model on ABIDE I dataset and achieved an accuracy of 90.3%. Similarly, Li et al. (2018a) employed stacked three SAEs in the encoding part to generate a stacked SAE prototype to be learned in an unsupervised manner and then

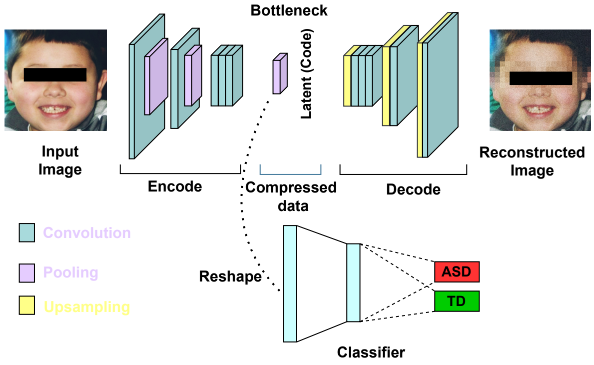

Fig. 11: A simple autoencoder architecture. First, in the encoding phase, the input image is compressed into a lower dimension (i.e., latent code generation) while upsampling the latent code into the output image. The generated latent code help to classify ASD and TD.

combined with softmax, subsequently fed into a deep transfer learning NN. They achieved an average accuracy of 67.1% on ABIDE I dataset. Again, Wang et al. (2019a) employed a stacked SAE with two hidden layers followed by a softmax, achieving 93.5% accuracy on ABIDE I dataset.

Variational Autoencoder (VAE) (Kingma and Welling, 2013; Milano et al., 2023) are probabilistic generative models in the latent space. The encoder can produce multiple samples from the same distribution while the decoder maps from the latent space to the input. The authors in (Zhang et al., 2022a) employed a VAE that first trained the model in an unsupervised way. Then the pre-trained encoder portion of the VAE is concatenated with additional layers for fine-tuning in a supervised manner. The model takes 3,170-dimensional features and reduces dimensions to 250 and then 150. Finally, these 150-dimensional features are fed into the softmax layer for the classification of ASD and TD. They achieve an accuracy of 78.1% over the ABIDE I dataset.

Table 3: Summary of articles published using DL-based methods for detecting, classifying, and rehabilitating ASD with the image(s) or video. CE: Cross Entropy, LOO: Leave One Out, N/A: Not Available or Applicable; Acc.: Accuracy; Sen.: Sensitivity; Spe.: Specificity; AUC: Area Under Curve; F1: F1 Score; Pre.: Precision.

Author (Year)	**Focus**	**Modality**	**Method**	**Datasets**	**Activation Function**	**Loss Function**	**Result [%]**	**K-Fold**
Park and Cho (2023)	Classification of ASD using functional brain connectivity between STS and visual cortex	MRI	GCN	ABIDE I	Softmax	N/A	Acc: 97.6; Sen: 98.0; F1: 98.0	10
Pavithra et al. (2023)	Identification of ASD and TD by RCNN based model and MRI data	MRI	CNN	ABIDE	N/A	N/A	Acc: 85.0; Sen: N/A; Spe: N/A	5
Bhandage et al. (2023)	Classify ASD and TD by using DBN and MRI data	MRI	DBN	ABIDE I	N/A	N/A	Acc: 92.4; Sen: 93.0; Spe: 93.5	N/A
Prakash et al. (2023)	Emotional and skill assessment test of ASD children from play-based intervention sessions	Multi-modal	R-CNN, Resnet	Primary	Softmax	N/A	Acc: 72.3; Sen: N/A; Spe: N/A (activity comprehension) Acc: 97.0; Sen: 95.5; Spe: 98.0 (joint attention of eye gaze) Acc: 95.1; Sen: N/A; Spe: N/A (facial expression recognition)	N/A
Mishra and Pati (2023)	Detect ASD and TD by CNN and MRI data	MRI	CNN	ABIDE I	N/A	N/A	Acc: 81.3; Sen: N/A; Spe: N/A	N/A
Milano et al. (2023)	Diagnosed ASD and TD analyzing their motor abnormalities	Multi-modal	VAE	Primary	Softmax	Proposed	Acc: 91.2; Sen: N/A; Spe: N/A	10
Wadhera et al. (2023)	Diagnosed ASD and TD using MRI image and hybrid DL model	MRI	VGG + ResNet	ABIDE I	Softmax	N/A	Acc: 88.1; Sen: 91.3; Spe: 86.3	N/A
Cao et al. (2023)	Classify ASD and TD by using patch based VIT and facial images	Face	VIT	AFID	N/A	MSE	Acc: 94.5; Sen: N/A; Spe: N/A; AUC: 97.9	N/A
Sabegh et al. (2023)	Classify ASD and TD by using resting-state fMRI data	MRI	CNN	ABIDE I	N/A	N/A	Acc: 73.5; Sen: N/A; Spe: N/A	N/A
Rabbi et al. (2023)	Detection of ASD and TD by using facial images	Face	VGG-19, Inception-V3, DenseNet-201	AFID	N/A	N/A	Acc: 85.0; Sen: N/A; Spe: N/A; AUC: 92.3 Acc: 78.0; Sen: N/A; Spe: N/A; AUC: 85.9 Acc: 83.0; Sen: N/A; Spe: N/A; AUC: 91.0	N/A
Atyabi et al. (2023)	Analyzing ASD and TD using spatio-temporal features of their scan-paths	Eye Gaze	CNN	Primary	N/A	N/A	Acc: 80.2; Sen: N/A; Spe: N/A; AUC: 83.8	N/A
Othmani et al. (2023)	Diagnose ASD and TD from MRI images	MRI	LeNet-5	ABIDE I	Sigmoid	CE	Acc: 95.0; Sen: 95.0; Spe: N/A; F1: 95.0	5
Alkahtani et al. (2023)	Identify ASD and TD based on facial landmark	Face	MobileNet, VGG-16	AFID	Softmax	CE	Acc: 92.0; Sen: 92.0; Spe: N/A; F1: 92.0 Acc: 82.1; Sen: 82.0; Spe: N/A; F1: 82.0	N/A
Nogay and Adeli (2023)	Diagnosed ASD and TD using structural brain MRI images and grid search optimization	MRI	CNN	ABIDE	Softmax	CE	Acc: 100; Sen: 100; Spe: N/A	5
Han et al. (2022)	Identify ASD and TD using Multi-modal (EEG, Eye track) framework	Multi-modal	DAE	Primary	N/A	N/A	Acc: 95.5; Sen: 92.5; Spe: 98.0	10
Atyabi et al. (2022)	Ordering ASD and TD using spatial and spatio-temporal scanpaths generated from eye gaze pattern	Eye Gaze	CNN	Primary	N/A	N/A	Acc: 80.2; Sen: N/A; Spe: N/A	N/A
Alam et al. (2022)	Identify ASD by transfer-learning-based method using facial images	Face	Xception, ResNet-50	AFID	N/A	CE	Acc: 95.0; AUC: 98.0; Pre: 95.0 Acc: 94.0; AUC: 96.0; Pre: 94.0	N/A
Kanhirakadavath and Chandran (2022)	Diagnose ASD and TD based on scan path	Eye Gaze	CNN	Figshare	Sigmoid	BCE	Acc: N/A; Sen: 93.2; Spe: 91.3; AUC: 97.0	5
Sharif and Khan (2022)	Classify of ASD and TD using corpus callosum and intracranial brain volume	MRI	CNN	ABIDE I	Softmax	N/A	Acc: 66.0; Sen: N/A; Spe: N/A	5
Wang et al. (2022)	Diagnose ASD based on multi-atlas GCN	MRI	GCN	ABIDE	Softmax	CE	Acc: 75.8; Sen: 79.2; Spe: 71.53	10
Torres et al. (2022)	Classify facial emotions using EEG signals.	EEG	CNN	Primary	Softmax	N/A	Acc: 86.0; Sen: N/A, Spe: N/A	LOO
Zhang et al. (2022a)	Identify ASD based based on MLP	MRI	VAE-MLP	ABIDE I	Softmax	CE	Acc: 78.1; Sen: 77.8; Spe: 78.3	10
Jiang et al. (2022b)	Classify ASD and TD using Spatio-temporal feature	MRI	3D CNN-GRU	ABIDE I	Sigmoid	CE	Acc: 72.4; Sen:74.3; Spe: 79.2	N/A
Hao (2022)	Diagnose ASD by exploring higher order correlation and AE	MRI	AE	ABIDE I	NN	MSE	Acc: 71.8 Sen: 70.8; Spe: 65.9	10
Kang et al. (2022)	Identify ASD and TD based on multi-view ensemble learning	MRI	LSTM+DAE	ABIDE	Sigmoid	N/A	Acc: 72.0; Sen: N/A; Spe: N/A	LOO
Guo et al. (2022)	Diagnose ASD using 3D ResNet-18	MRI	3D ResNet-18	Primary	N/A	N/A	Acc: 84.4; Sen: 85.0; Spe: 84.0	N/A
Zhang et al. (2022b)	Diagnose ASD based on F-score selection method using fMRI	MRI	AE	ABIDE	N/A	N/A	Acc: 70.9; Sen: N/A; Spe: N/A	N/A
Wen et al. (2022)	classify ASD and TD using multi-view GCN	MRI	GCN	ABIDE	N/A	Proposed	Acc: 69.3; Sen: N/A; AUC: 69.0	10
Devika et al. (2022)	Classify ASD and TD using GAN	MRI	GAN	ABIDE II, ADHD-200	N/A	Proposed	Acc: 97.8; Sen: N/A; Spe: N/A	N/A
Li et al. (2022)	Diagnose ASD using deep learning framework from MRI data	MRI	3D ResNet-Inception	ABIDE, Primary	N/A	N/A	Acc: N/A; Sen: 86.0; Spe: 62.0; AUC: 85.6 Acc: N/A; Sen: 88.0; Spe: 75.0; AUC: 78.7	N/A
Cai et al. (2022)	Diagnose ASD using DL framework	Face	ResNet-50	Public	N/A	N/A	Acc: 95.0; Sen: 92.5; Spe: 96.4	3
Yang et al. (2022)	Classify ASD and TD using various classifiers with Altas	MRI	DNN	ABIDE I	Softmax	CE	Acc: 68.4; Sen: 62.7; Spe: 73.6	5
Kashef (2022)	Identify ASD using enhanced CNN	MRI	CNN	ABIDE I	Softmax	N/A	Acc: 80.0; Sen: N/A; Spe: N/A	10
Elakkiya and Dejey (2022)	Classify ASD and TD using RBM e-Gaussian Process	MRI	Bernoulli RBM	ABIDE I	N/A	N/A	MSE: 20.0	2
Pandian et al. (2022)	Detects stimming behavior of children to help ASD diagnosis by developing RGBPOSE-SLOWFAST	Skeleton	ResNet-34	SSBD, Primary	N/A	N/A	Acc: 98.0; Sen: N/A; Spe: N/A Acc: 86.0; Sen: N/A; Spe: N/A	N/A
Lu et al. (2022)	Classify ASD and TD using fuzzy inference system and DBN	MRI	DBN	ABIDE I	Softmax	CE	Acc: 68.6; Sen: 67.1; Spe: 70.0	5
Lakkapragada et al. (2022)	Detect hand-flapping to analyze ASD	Multi-modal	MobileNetV2-LSTM	SSBD	Sigmoid	CE	Acc: 85.0; Sen: 80.4; F1: 84.0	5
Ahmed et al. (2022a)	Classify ASD and TD by analyzing the scan path of individuals eye	Eye Gaze	FFNN, ANN	Figshare	Softmax	N/A	Acc: 99.8; Sen: 99.5; Spe: 100 Acc: 99.8; Sen: 100; Spe: 99.7	N/A
Mujeeb Rahman and Subashini (2022)	Distinguish ASD and TD from static features of face images	Face	Xception, EfficientNetB1	AFID	Sigmoid	CE	Acc: 90.0; Sen: 88.4; Spe: 91.6; AUC: 96.6 Acc: 89.6; Sen: 86.0; Spe: 94.0; AUC: 95.0	N/A

Continued on next page

Table 3 – *Continued from previous page*

Author (Year)	Focus	Modality	Method	Datasets	Activation Function	Loss Function	Result [%]	K-Fold
Alsaade and Alzharani (2022)	Classify ASD and TD based DL methods using facial features	Face	Xception, VGG-19,	AFID	Softmax	N/A	Acc: 91.0; Sen: 88.0; Spe: 94.0 Acc: 80.0; Sen: 78.0; Spe: 83.0	N/A
Ali et al. (2022)	Recognize autistic behaviors using a multi-modal fusion framework	Multi-modal	3D CNN	Primary, SSBD	N/A	N/A	Acc: 86.0; Sen: N/A; Spe: N/A; F1: 88.8 Acc: 75.6; Sen: N/A; Spe: N/A; F1: 90.5	5
Baygin et al. (2021)	Detect ASD using EEG signals	EEG	MobileNet+Shuff-leNet+SqueezeNet	Primary	N/A	N/A	Acc: 96.4; Sen: 97.7; Spe: 93.1	10
Liang et al. (2021a)	Classify self-stimulatory behaviors of ASD using Temporal Coherency Deep Network	Multi-modal	AlexNet	SSBD	N/A	Proposed	Acc: 98.3; Sen: N/A; Spe: N/A	5
Wei et al. (2021)	Identify ASD based on spatiotemporal features of eye movement	Eye Gaze	CNN- LSTM	Saliency4ASD	Sigmoid	CE	Acc: 61.4; Sen: 68.5; Spe: 54.6	N/A
Hosseini et al. (2021)	Classify ASD and TD based on facial images, and DL methods	Face	MobileNet	AFID	Softmax	N/A	Acc: 94.6; Sen: N/A; Spe: N/A	N/A
Akter et al. (2021)	Identify ASD by transfer-learning-based method from face images	Face	MobileNet, DenseNet-121	AFID	N/A	N/A	Acc: 92.1; Sen: N/A; Spe: N/A Acc: 83.6; Sen: 83.6; Spe: 83.6	10
Almuqhim and Saeed (2021)	Classify ASD and TD by developing ASD-SAENet	MRI	SAE	ABIDE I	Softmax	CE	Acc: 70.8; Sen: 62.2; Spe: 79.1	10
Cao et al. (2021)	Identify ASD using deep GCN	MRI	GCN	ABIDE I	Softmax	N/A	Acc: 73.7; Sen: N/A; AUC: 75.0; F1: 69.6	10
Gao et al. (2021)	Identify ASD based on morphological covariance brain networks	MRI	ResNet	ABIDE I	N/A	CE	Acc: 71.8; Sen: 81.2; Spe: 68.7	10
Al-Hiyali et al. (2021b)	Diagnose ASD using temporal dynamic features of fMRI	MRI	DenseNet-201, ResNet-101	ABIDE I	N/A	N/A	Acc: 85.9; Sen: 79.3, Spe: 92.6 Acc: 84.4; Sen: 73.4; Spe: 82.4	N/A
Liang et al. (2021b)	Classify ASD and TD by combining CNN and prototype learning framework	MRI	CNN	ABIDE I	N/A	Proposed	Acc: 77.3, Sen: 78.0; Spe: 77.8	10
Al-Hiyali et al. (2021a)	Identify ASD subtypes using CNN and dynamic FC features	MRI	CNN	ABIDE	Softmax	N/A	Acc: 89.8; Sen: 90.1 Spe: N/A (Binary class) Acc: 82.1; Sen: N/A; Spe: N/A (Multi class)	20
Kiruthigha and Jaganathan (2021)	Identify ASD using GCN	MRI	CNN+GCN+VAE	ABIDE	N/A	N/A	Acc: 62.6; Sen: N/A; Spe: N/A	N/A
Liu et al. (2021)	Identify ASD using multi-regional rs-fMRI data	MRI	LSTM-AE	ABIDE	Softmax	CE	Acc: 71.3; Sen: N/A; Spe: N/A; Pre: 70.5	10
Kojovic et al. (2021)	Detect ASD by extracting skeletal key points during social interaction	Skeleton	CNN-LSTM	Primary	Softmax	CE	Acc: 80.9; Sen: 85.4; Spe: N/A; Pre: 78.4	N/A
Banire et al. (2021)	Recognize attention of ASD children based on facial expression	Face	CNN	Primary	N/A	N/A	Acc: 89.4; Sen: N/A; Spe: N/A; AUC: 85.6;	N/A
Bayram et al. (2021)	Detect ASD using various DL-methods through rs-fMRI data	MRI	RNN, BiLSTM	ABIDE I	Sigmoid	N/A	Acc: 74.7; Sen: 72.9; Spe: 76.2; Acc: 74.5; Sen: 72.2; Spe: 76.5;	10
Lu and Perkowski (2021)	Diagnose ASD using transfer learning-based methods	Face	CNN	East Asian*	N/A	N/A	Acc: 95.0; Sen: N/A; Spe: N/A; F1: 95.0	10
Subah et al. (2021)	Detect ASD from functional connectivity features of rs-fMRI	MRI	DNN	ABIDE I	Sigmoid	CE	Acc: 88.0; Sen: 90.0; F1: 87.0; AUC: 96.0	5
Saranya and Anandan (2021)	Detect ASD from human gaits using multi-modal features with DL	Multi-modal	CNN	FER2013, CASIA KDEF Lundqvist et al. (1998)	Softmax	RMSE	Acc: 96.5; Sen: 94.5; Spe: 95.0	10
Liaqat et al. (2021)	Classify ASD and TD using ResNets from gaze data	Eye Gaze	ResNet-18, ResNet-50	Saliency4ASD	N/A	CE	Acc: 61.4; Sen: 73.0; Spe: 50.0; AUC: 66.0 Acc: 62.1; Sen: 71.0; Spe: 54.0; AUC: 67.0	N/A
De Belen et al. (2021)	Diagnose ASD by developing DNN-based model from eye-tracking data	Eye Gaze	CNN-LSTM	Primary	N/A	Proposed	Acc: 68.0-100; Sen: 57.0-100; Spe: 65.0-100	LOO
Tawhid et al. (2021)	Classify ASD and TD based on spectrogram image of EEG	EEG	CNN	KAU [4]	Softmax	N/A	Acc: 99.1; Spe: 99.0; Sen: 99.1	N/A
Ganesh et al. (2021)	Classify ASD and TD based on facial thermal imaging	Face	ResNet-50, CNN	Primary	Softmax	N/A	Acc: 90.0; Sen: 87.0; Spe: 92.0; Acc: 96.0; Sen: 100; Spe: 93.0;	N/A
Cilia et al. (2021)	Screening ASD using the eye scan path and correlating between autism severity	Eye Gaze	CNN	Primary	N/A	N/A	Acc: 90.0; Sen: 83.0; Pre: 80.0; AUC: 90.0	3
Wu et al. (2021)	Diagnose ASD based on behaviour feature using ResNet	Face	ResNet-18	Primary	N/A	CE	Acc: (Smile): 70.0; (Look face): 68.0; (Look object): 67.0; (Vocalization): 53.0	N/A
Yin et al. (2021)	Diagnose ASD using AE-based method with neuroimage	MRI	AE-DNN	ABIDE I	Softmax	N/A	Acc: 79.2; Sen: N/A; Spe:N/A; AUC: 82.4	10
Xia et al. (2020)	Identify ASD using eye-tracking data	Eye Gaze	CNN	Primary	N/A	N/A	Acc: 93.1 Sen: 94.6; Spe: 92.0	N/A
Berardini et al. (2020)	Detect whether an ASD child washes their hands or not	Multi-modal	VGG-16	Primary	Softmax	BCE	Acc: 91.0; Sen: N/A; Spe: N/A	N/A
Fabiano et al. (2020)	Classify the risk of ASD as low, medium, and high	Eye Gaze	DNN	ETS-E[5]	Softmax	N/A	Acc: 85.1; Spe: N/A Sen: N/A (Raw Gaze) Acc: 92.5; Spe: N/A; Sen: N/A (Gaze Patterns)	10
Haputhanthri et al. (2020)	Classify ASD and TD using thermographic and EEG data	Multi-modal	MLP	Primary	Sigmoid	N/A	Acc: 94.0; Sen: N/A; Spe: N/A	LOO
Tamilarasi and Shanmugam (2020)	Classify ASD and TD based on thermal face images	Face	ResNet-50	Primary	N/A	N/A	Acc: 89.2; Sen: N/A; Spe: N/A	N/A

Continued on next page

[4] https://malhaddad.kau.edu.sa/Pages-BCI-Datasets.aspx
[5] https://nda.nih.gov/

Table 3 – *Continued from previous page*

Author (Year)	Focus	Modality	Method	Datasets	Activation Function	Loss Function	Result [%]	K-Fold
Javed and Park (2020)	Identify the risk of ASD using behavioral data	Multi-modal	CNN	Primary	Softmax	N/A	Acc: 88.4; Sen: 88.5; Pre: 89.1	N/A
Sewani and Kashef (2020)	Classify ASD and TD using CNN with rs-fMRI data	MRI	AE-CNN	ABIDE I	Sigmoid	N/A	Acc: 84.0; Sen: 80.0; Spe: 75.3; AUC: 78.0	10
Ke and Yang (2020)	Classify ASD and TD and investigate biomarker	MRI	RAM	ABIDE I	N/A	Proposed	Acc: 87.4; Sen: 93.7; Spe: 69.9	5
Huang et al. (2020)	Identify ASD using Deep Belief Network (DBN)	MRI	DBN	ABIDE I	Softmax	CE	Acc: 76.4; Sen: N/A; Spe: N/A	10
Du et al. (2020)	Classify ASD and Schizophrenia using 3D CNN	MRI	3D CNN	ABIDE I +Primary	N/A	N/A	Acc: 87.0; Sen: N/A; Spe: N/A	5
Thomas et al. (2020)	Classify ASD and TD using 3D CNN with temporal statistics of rs-MRI	MRI	3D CNN	ABIDE I + ABIDE II	Sigmoid	CE	Acc: 64.0; Sen: N/A; Spe: N/A; F1: 66.0;	5
D'Souza et al. (2020)	Predicting spectrum-level deficits for autism using a generative model	MRI	LSTM-ANN	Primary	N/A	Proposed	Median Absolute Error: 13.5	5
Wang et al. (2020b)	Identify ASD using multiple atlases	MRI	AE	ABIDE	Softmax	Proposed	Acc: 74.5; Sen: 80.6; Spe: 66.7; AUC: 80.2	10
Tang et al. (2020)	Diagnose ASD by combining MLP and ResNet	MRI	MLP+CNN	ABIDE I	Softmax	CE	Acc: 74.0; Rec: 94.9; Pre: 69.9; F1: 80.5	N/A
Awatramani and Hasteer (2020)	Educate children with ASD to identify human emotions	Face	CNN	FER-2013	Softmax	N/A	Acc: 67.5; Sen: N/A; Spe: N/A	N/A
Li et al. (2020)	Diagnose ASD using LSTM from raw video data	Eye Gaze	LSTM	Primary	N/A	CE	Acc: 92.6; Sen: 91.9; Spe: 93.4	10
Niu et al. (2020)	Classify ASD and TD using multichannel deep attention NN	MRI	MLP	ABIDE I	Sigmoid	CE	Acc: 73.2; Sen: 74.5; Spe: 71.7; F1: 73.6	10
Ahmed et al. (2020)	Classify ASD and TD by generating single-volume images from the whole brain using a CNN-based model	MRI	Xception+CNN, ResNet+CNN	ABIDE I	Sigmoid	CE	Acc: 87.0; Pre: 86.8; Sen: 86.0; F1: 86.0; Acc: 86.0; Pre: 85.6; Sen: 84.5; F1: 85.1;	N/A
Sherkatghanad et al. (2020)	Detect ASD automatically using CNN with ABIDE I	MRI	CNN	ABIDE I	MLP	N/A	Acc: 70.2; Sen: 77.0; Spe: 61.0	10
Yang et al. (2020)	Classify ASD and TD using DNN with rs-fMRI Data	MRI	DNN	ABIDE I	Softmax	CE	Acc: 75.2; Sen: 74.0; Pre: 76.8	5
Fang et al. (2019)	Saliency prediction for ASD using CNN with eye gaze data	Eye Gaze	CNN	Saliency4ASD	Sigmoid	MSE	AUC_Judd : 76.8; AUC_Borji: 78.9	6
Duan et al. (2019a)	Analyze the visual attention of ASD when looking at a face	Multi-modal	CNN	Primary	N/A	N/A	AUC_Judd: 84.8; AUC_Borji: 82.3	10
Wei et al. (2019)	Analyze ASD using a CNN-based saliency prediction model	Eye Gaze	CNN	Saliency4ASD	N/A	Porposed	AUC_Judd: 81.8	N/A
Rathore et al. (2019)	Classify ASD and TD using DNN with tropological features of fMRI	MRI	DNN	ABIDE I	Softmax	CE	Acc: 69.2; Sen: N/A; Spe: N/A	5
Mellema et al. (2019)	Diagnose ASD using LSTM from rs-fMRI data	MRI	FFNN, LSTM	IMPAC Toro et al. (2018)	N/A	BCE	Acc: N/A; Sen: N/A; Spe: N/A; AUC: 80.0 Acc: N/A; Sen: N/A; Spe: N/A; AUC: 77.6	3
Mostafa et al. (2019b)	Diagnose ASD by employing an AE-based approach	MRI	AE	ABIDE I	N/A	N/A	Acc: 79.2; Sen: N/A; Spe: N/A; AUC: 82.4	10
Mostafa et al. (2019a)	Classify ASD and TD using brain network and eigenvalue	MRI	NN	ABIDE I	N/A	N/A	Acc: 71.7; Sen: N/A; Spe: N/A; AUC: 78.7	5
Aghdam et al. (2019)	Diagnose ASD using a mixture of experts CNN	MRI	CNN	ABIDE I, ABIDE II	N/A	CE	Acc: 72.7; Sen: 71.2; Spe: 73.4 Acc: 70.0; Sen: 58.2; Spe: 80.4	10
Pugazhenthi et al. (2019)	Identify ASD using a CNN-based approach from MRI	MRI	AlexNet	ABIDE	Softmax	CE	Acc: 82.6; Sen: N/A; Spe: N/A	N/A
Wang et al. (2019a)	Identify ASD using stacked AE from MRI data	MRI	SAE	ABIDE I	Softmax	N/A	Acc: 93.5; Sen: 92.5; Spe: 94.5	Multi
Tao and Shyu (2019)	Classify ASD and TD from the scanpath of the observer gaze	Eye Gaze	CNN-LSTM	Saliency4ASD	N/A	CE	Acc: 57.9; Rec: 59.2; Pre: 56.2	N/A
Wang et al. (2019b)	Assessing ASD by gesture and mutual gaze data	Multi-modal	VGG-16	Oxford hand, Egohands	N/A	N/A	N/A	N/A
Rani (2019)	Detect the emotion of autistic children from the face image	Face	ANN	Primary	N/A	N/A	Acc: 70.0; Sen: N/A; Spe: N/A	N/A
Leo et al. (2019)	Quantitative assessment of facial expression for ASD and TD	Face	CNN	Primary	N/A	N/A	Acc: N/A; Rec: 85.0; Pre: 88.0; F1: 86.0	N/A
Li et al. (2019a)	Classify ASD and TD from facial expressions, action units, arousal, and valence	Face	CNN	AffectNet, EmotioNet	N/A	CE	Acc: N/A; Sen: 76.0; Spe: 69.0; F1: 76.0;	LOO
Kong et al. (2019)	Classify ASD and TD by generating individual brain network features with the DNN classifier	MRI	AE	ABIDE I	Softmax	MSE	Acc: 90.3; Sen: 84.3; Spe: 95.8; AUC: 97.3	10
Chen and Zhao (2019)	Classify ASD and TD by photo taking and eye tracking task	Multi-modal	CNN-LSTM	Primary, Saliency4ASD	Sigmoid	CE	Acc: 99.0; Sen: 100; Spe: 98.0; AUC: 100 Acc: 93.0; Sen: 93.0; Spe: 93.0; AUC: 98.0	LOO
Leo et al. (2018a)	Analyze facial expression production automatically using CNN	Face	CNN	Primary	N/A	N/A	Action units 6: 1.0; Action units 12: 1.7 (ASD) Action units 6: 2.4; Action units 12 : 3.0 (TD)	N/A
Leo et al. (2018b)	Assess the capability of ASD children to produce facial expression	Face	CNN	CK+ Lucey et al. (2010)	N/A	N/A	Acc: Neutral: 90.0; Happiness: 99.0; Sadness: 67.0; Fear: 84.0 Anger: 73.0	N/A
Tang et al. (2018)	Detect ASD with the help of smile detection for infants	Face	CNN	GENKI-4K Littlewort et al. (2011), CelebA Liu et al. (2015)	Softmax	CE	Acc: 94.5; Sen: N/A; Spe: N/A Acc: 92.6; Sen: N/A; Spe: N/A	4
Heinsfeld et al. (2018)	Classify ASD and TD based on their neural patterns of functional connectivity using DAE	MRI	DAE	ABIDE I	Softmax	MSE	Acc: 70.0; Sen: 74.0; Spc: 63.0	10
Chrysouli et al. (2018)	Recogne affective state using two-stream CNN from eye gaze	Eye Gaze	Two stream CNN	Primary	Sigmoid	MSE	Acc: 95.3; Sen: N/A; Spe: N/A (two class) Acc: 92.7; Sen: N/A; Spe: N/A (three class)	5
Han et al. (2018)	Computing emotional expression of children with ASD using feature transfer-based approach	Face	CNN	FERET Phillips et al. (2000)+ CK+ Lucey et al. (2010)+ Primary	N/A	MMD	Acc: 79.4; Sen: N/A; Spe: N/A	N/A
Marinoiu et al. (2018)	Recognize action and emotion recognition during Robot-assisted therapy of children with Autism	Skeleton	CNN, RNN	DE-ENIGMA	N/A	N/A	Acc: (Kinect): 53.1; (DMHS-SMPL-T): 47.9 Acc: (Kinect): 37.8; (DMHS-SMPL-T): 36.2	LOO

Continued on next page

Table 3 – *Continued from previous page*

Author (Year)	**Focus**	**Modality**	**Method**	**Datasets**	**Activation Function**	**Loss Function**	**Result [%]**	**K-Fold**
Li et al. (2018c)	Classify ASD and TD using a multi-stage method from fMRI and interprets the saliency feature	MRI	CNN	Primary, ABIDE I	Sigmoid	N/A	Acc: 87.1; Sen: N/A; Spe: N/A Acc: 85.3; Sen: N/A; Spe: N/A	N/A
Li et al. (2018a)	Classify ASD and TD using stacked sparse AE	MRI	SAE	ABIDE I	Softmax	MSE	Acc: 67.1; Sen: 65.7; Spe: 68.3	5
Zunino et al. (2018)	Diagnose ASD and TD by grasping gesture data using CNN-LSTM	Multi-modal	CNN-LSTM	Public	Softmax	N/A	Acc: 72.0; Sen: N/A; Spe: N/A (ASD) Acc: 77.0; Sen: N/A; Spe: N/A (TD)	LOO
Liao and Lu (2018)	Classify ASD and control based on DL and community structure on rs-fMRI	MRI	DAE	ABIDE I	N/A	Proposed	Acc: 54.4 Sen: N/A; Spe: N/A	LOO
Jiang and Zhao (2017)	Diagnose ASD using DNN with eye-tracking data	Eye Gaze	CNN	Primary	N/A	CE	Acc: 92.0; Sen: 93.0; Spe: 92.0; AUC: 92.0	LOO
Chong et al. (2017)	Detect eye contact during adult-child social interactions and head pose	Eye Gaze	Pi-CNN	Primary	N/A	N/A	Acc: N/A; Pre: 76.0; Sen: 80.0; AUC: 79.0	5
Patnam et al. (2017)	Recognize meltdown actions of ASD	Face	R-CNN	Primary	N/A	N/A	Acc: 92.0 Sen: N/A; Spe: N/A	N/A
Shukla et al. (2017)	Detect developmental disorders using AlexNet from face image	Face	AlexNet	Primary	N/A	N/A	Acc: 98.8; Sen: N/A; Spe: N/A	5
Ismail et al. (2017)	Detect ASD using shape variation structure of MRI using DL-based approach	MRI	AE	ABIDE I, NDAR/Pitt Hall et al. (2012),	Softmax	log-likelihood	Acc: 92.8; Sen: N/A; Spe: N/A Acc: 96.8; Sen: N/A; Spe: N/A	N/A

Denoising Autoencoder (DAE) (Lu et al., 2013) is an extension of a simple AE to help hidden layers learn more robust filters and reduce the risk of overfitting. Heinsfeld et al. (2018) used a DAE-based method to train a model for the classification of ASD. The method includes two stacked DAEs employed for extracting low-dimensional data. The input-output layers had 19,900 dimensions, reducing it into a bottleneck of 1,000 units, and finally, they employed the softmax layer. They achieved an accuracy of 70.0% over the ABIDE I dataset. Similarly, in (Liao and Lu, 2018) explored the DAE for ASD classification and achieved an accuracy of 54.4% accuracy on ABIDE I dataset.

6.5. *Recurrent Neural Network*

Recurrent Neural Networks (RNN) (Medsker and Jain, 2001) are a form of NN in which the previous layer's output is given as input to the current layer. The network remembers its previous input due to internal memory, allowing it to predict the following input. The ability of RNNs to model sequential data, handle variable-length inputs, and capture long-term dependencies makes them a powerful tool in the field of deep learning. RNN can be employed to analyze ASD in the literature for autism research. A simple RNN architecture to analyze ASD is illustrated in Fig. 12. Bayram et al. (2021) classified ASD and TD using RNN models based on rs-fMRI data. Their models are made up of three layers. The first is an RNN layer with a scaled exponential linear unit activation function, followed by a dropout layer, and finally, an FC layer with sigmoid activation. They demonstrated on ABIDE I dataset and achieved an accuracy of 74.7%. In another study, Ke and Yang (2020) proposed a Recurrent Attention Model (RAM), a combination of RNN and reinforcement learning algorithm. Later, they added the Gaussian sampling method into the RAM and achieved an accuracy of 87.4% on ABIDE dataset. Marinoiu et al. (2018) constructed hierarchical bidirectional RNNs for action classification. They employed five skeleton subcomponents: torso, left arm, right arm, therapist left arm, and therapist right arm as the input to the network. They demonstrated their model on 3D skeleton features obtained with Kinect (Shotton et al., 2011) and achieved an accuracy of 37.8% for action recognition.

6.6. *Long Short Term Memory*

Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997) is a variant of recurrent neural networks (RNNs) that has already been discussed in Sec. 5, along with illustrated in Fig. 8. It can be employed for feature extraction as well as the classification of time series data. Here we explore the classifications for ASD and TD by employing LSTM as feature extraction, already discussed in the previous section. Li et al. (2020) used a three-layer LSTM where each layer has 64 hidden units for classifying ASD and TD from video. The accumulative histogram of eye features is converted into consecutive time series and fed into LSTM for classification. They evaluate their model with their privately collected dataset and find a specificity of 93.4%. Similarly, the authors in (D'Souza et al., 2020) proposed a framework that decomposes the complementary information from rs-fMRI connectivity and diffusion tensor imaging tractography to extract predictive biomarkers. The deep part of their framework is based on LSTM along with ANN. They demonstrated their framework on a privately collected dataset of 57 subjects with ASD.

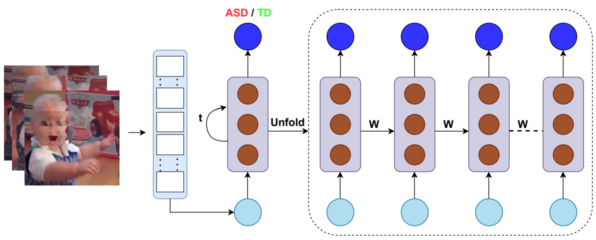

Fig. 12: An example of a simple recurrent neural network (RNN) structure. Left: compressed and right: unfolded RNN architecture.

6.7. *CNN-LSTM*

The CNN-LSTM architecture was created primarily for sequence prediction issues with spatial inputs, such as images or videos. The extracted features of the CNN layer are usually integrated with LSTMs. A CNN-LSTM can be constructed by first adding CNN layers, then LSTM is employed, and finally, a dense layer is employed to get the output. Because of this combination, the model can learn complicated patterns and correlations in sequential data. A basic structure of CNN-LSTM architecture is illustrated in Fig. 9. CNN-LSTM can be explored in the literature for autism research. For example, Kojovic et al. (2021) employed the CNN-LSTM model to classify individuals with ASD and TD. The VGG-16 pre-trained CNN was utilized to extract high-dimensional features from individual video clips, and the extracted feature flattened and attached to the input into the LSTM unit. Finally, the softmax activation was used to classify ASD and TD. They demonstrated that it achieved of F1-score of 81.0% over its own collected dataset.

Furthermore, Studied in (Tao and Shyu, 2019) proposed a system named SP-ASDNet, including CNN along with LSTM; the CNN model took a sequence of image patches of the saliency map as input and generated a visual feature vector with 1024 dimensions while two-layer LSTM classifies ASD and TD by employing the extracted feature. They achieved an accuracy of 57.9% with batch normalization on the Saliency4ASD dataset. In addition, Zunino et al. (2018) classify ASD and TD from video gesture data. First, the video data is provided into the CNN and generated frame-wise features of 7x7x1024 size. Then, 128-dimensional LSTM layers with softmax activation were employed to get the results of ASD 72.0% and TD 77.0%. Similarly, Lakkapragada et al. (2022) develop a CNN-LSTM-based model which helps find abnormalities of a hand flipping to aid in detecting ASD. They employed MobileNetV2 (Howard et al., 2017) for feature extraction, followed by an LSTM for classification. The LSTM produces 64-dimensional output followed by an FC layer with a sigmoid activation function. They achieved an accuracy of 85.0% with the SSBD dataset for classification.

7. Deep Leaning-based Rehabilitation

Rehabilitation means returning a person to a regular life through training and therapy. Numerous deep learning techniques can be employed to develop mobile applications, cloud-based software, devices, robots, etc., in autism rehabilitation. The DL-based approaches to rehabilitate individuals with ASD are summarized in Table 4. Some studies developed rehabilitation tools by employing facial features. For example, Haque and Valles (2019) developed a mobile IOS app by exploring deep convolutional NN to teach ASD children how to recognize facial emotions. It operates to snap a photo, which is then converted into a variety of emoji so the autistic child can convey their feelings. In addition, Ahmed et al. (2022b) developed a web application to assess children's state (e.g., ASD or TD) based on the facial image. They built their model using pre-trained CNN networks: MobileNet, Xception, and InceptionV3.

On the other hand, Rudovic et al. (2018) estimate the engagement levels using the ResNet with five FC layers for ASD children using facial images from a video dataset (Rudovic et al., 2017) collected from different cultural ASD children during robot-assisted therapy. Similarly, studied in (Li et al., 2019b) developed an assisted therapeutic system to predict the emotion of ASD

Table 4: Summary of articles published for the rehabilitation of ASD using DL-based methods with the image(s) or video.

Author (Year)	Rehabilitation Focus	Method
Singh et al. (2023)	Developed a socially designed robot to assist ASD children	SSD YOLO v3
Ahmed et al. (2022b)	Develop web-based app to identify ASD from face	MobileNet Xception Inception
Salhi et al. (2022)	Employed a humanoid robot to assist ASD therapy.	CNN
Zhang et al. (2020)	Improve social skill of ASD children using robots	CNN-LSTM
Sun et al. (2020)	Automatic action recognition of ASD	LSTM
Haque and Valles (2019)	Develop IOS app to teach children with asd	CNN
Elbattah et al. (2019)	rehabilitate ASD using DL and clustering algorithm	AE
Wu et al. (2019)	rehabilitate ASD using saliency prediction	ResNet
Li et al. (2019b)	Develop ASD social skill by providing games	CNN-SVR
Rudovic et al. (2018)	Estimate engagement of ASD during robot therapy	R-CNN ResNet

children using five interactive subgames. They employed CNN with a Support Vector Regression model based on reinforcement learning to predict emotions. Singh et al. (2023) developed a cost-effective, social, and educational robot named 'Tinku' to assist ASD children. The robot is able to teach ASD children such as brushing, storytelling, and table manners to improve their regular activities. For developing the robots they used Yolo v3 and single shot detector (SSD) methods.

Furthermore, eye-tracking-based methods were developed to rehabilitate ASD children. For example, Elbattah et al. (2019) explored unsupervised clustering algorithms with a deep learning-based approach. They visualized eye-tracking fixations and saccades and forwarded them to the AE for feature learning. Finally, the K-Means were employed for clustering with k=2 and 3. Similarly, Wu et al. (2019) predict ASD using two DL-based approaches from gaze data to help ASD rehabilitation. The first is a generative model of synthetic saccade patterns, and the other uses ResNet-18.

In addition, Zhang et al. (2020) developed a dense image captioning method to improve the cognitive ability and social communication of ASD children using robots. It consists of CNN and NLP, where CNN extracted features from images, and then these features were fed into the NLP model. Moreover, Sun et al. (2020) develop an automatic stereotyped motor movement detection system from video data to rehabilitate ASD. They integrate the spatial attentional bilinear 3D convolutional network with LSTM.

8. Discussion and Challenges Ahead

8.1. Discussion

This article aims to conduct a systematic review of deep learning using image-based autism spectrum disorder analysis and demonstrate the utility of deep learning in autism research. A total of 130 articles that used deep learning-based methods were reviewed. In the field of ASD detection, classification, and diagnosis, innumerable articles have been published using the extracted feature from image(s) or video data, as summarized in Table 3, along with rehabilitation which is outlined in Table 4.

A variety of DL-based approaches have been proposed or employed existing off-the-self pretraining networks for analyzing ASD research. The number of DL networks used for ASD research is illustrated in Fig. 13. It showed that CNN is found to be the most widely explored network compared to other methods (i.e., 69 articles out of 130), while autoencoder becomes the second best choice (see Fig. 13). CNN-LSTM and LSTM/RNN are also explored. From this information, we are unable to justify clearly about the most dominant model though CNN seems the leading one. The reason can be better understood from Figure 14. In this Figure, we demonstrate a presentation of some well-known datasets for autism and deep learning research in terms of accuracy. We only presented the top 4 or 5 results on four important datasets. We notice that each dataset reached more than 93% accuracy. It implies that we need to explore smarter approaches for much higher accuracies. Therefore, we need to develop a platform where we can try all major datasets by prominent and dominant methods or models. Through some rigorous efforts, we may find a better understanding. However, it will be a much more challenging task to make a single platform and do the needed analysis. In recent years, the number of DL-based articles has increased exponentially due to their good performance in the field of ASD research, as shown in Fig. 1. Among the various classification methods employed in the DL-based approach, it showed that the Softmax algorithm is one of the best and most widely used algorithms (Table 3).

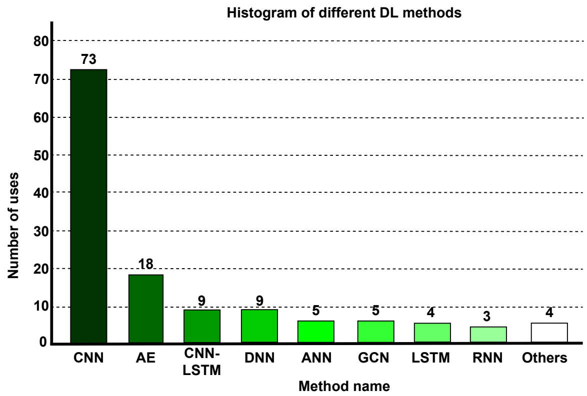

Fig. 13: Histogram of different DL-based approaches considered in this systematic review. Here, the best method is considered if multiple methods are employed in the same article.

8.2. Challenges in this Domain

With the growing data science and deep learning trend, a very large-scale dataset is always needed to solve a problem efficiently. Recently, many sophisticated DL-based methods have been developed for detection and classification, which require a massive number of samples for training because more data are essential than an algorithm. Therefore, large-scale benchmark datasets are needed in this domain. As presented above (Table 1), the existing publicly available datasets are small, with few subjects. Its challenging to manage more subjects with autism due to ethical clearance, social issues in some regions, and the complexity and diversities of autism levels. Furthermore, competitions/challenges can be arranged to find superior as well as diversified methods based on some challenging datasets; eventually, these activities can help to grow the research domain. While collecting the dataset, expert therapists do not provide scores during data collection. In some cases, the autism levels or scores are taken from the children's schools, which may not be the recent data. Therefore, it would be interesting to involve autism therapists during the data collection process.

In addition, most of the datasets are uni-modal; hence, multi-modality issues are not much explored. Multi-modal data can provide superior information, as explained in Sec. 4.1.4. For example, the DREAM dataset (Billing et al., 2020) has skeleton data of the upper body but no video or face-related data (even though they collected video data). As skeleton data from the Kinect sensor are noisy and only provide for a few joints, the results from these data could be more suitable. Video data can indeed hinder privacy, but to gain better results and evaluations, video data, face informa-

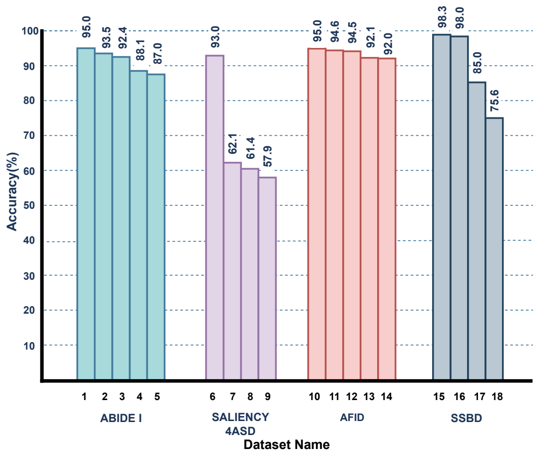

Fig. 14: A visual representation of popular datasets for autism research according to accuracy, based on the deep learning approaches considered in this systematic review. Numbers in the abscissa refer 1: (Othmani et al., 2023), 2: (Wang et al., 2019a), 3: (Bhandage et al., 2023), 4: (Wadhera et al., 2023), 5: (Ahmed et al., 2020), 6: (Chen and Zhao, 2019), 7: (Wei et al., 2021), 8: (Liaqat et al., 2021), 9: (Tao and Shyu, 2019), 10: (Alam et al., 2022), 11: (Hosseini et al., 2021), 12: (Cao et al., 2023) 13: (Akter et al., 2021), 14: (Alkahtani et al., 2023) 15: (Liang et al., 2021a), 16: (Pandian et al., 2022), 17: (Lakkapragada et al., 2022), 18: (Ali et al., 2022). Here, top four/five methods are considered per dataset. The method is considered for ABIDE I dataset, if more than 1,000 samples are employed in the experiment.

tion, eye gaze, head directions, etc., are essential. Moreover, it would be interesting to construct a large-scale multi-modal dataset if several research consortiums can be formed through research workshops or the like.

ASD has various scoring mechanisms or tools, e.g., Autism Diagnostic Interview-Revised (ADI-R) score (Lord et al., 1994), Childhood Autism Rating Scale (CARS) (Schopler et al., 1980), Diagnostic Interview for Social and Communication Disorders, etc. How to merge various scoring mechanisms along with the existing publicly-available datasets is a concern. Note that most of the research groups working on computer vision or sensor-based autism study by machine learning and/or deep learning – are not direct experts on autism. The research teams may occasionally need to manually collect and label data. Because of this, inaccurate annotations and predictions may result from the group's lack of knowledge. Moreover, Researchers mostly explore data and the corresponding scores that are provided with the dataset. Therefore, it is sometimes challenging to comprehend the discussion and findings of the results. Furthermore, Computer-vision based technology for the diagnosis of autism spectrum disorder is a comparatively new field so it is not fully industrial yet. But there is some growing system that can show promising output in the industry e.g., Wearable and mobile technologies (Koumpouros and Kafazis, 2019), Virtual Reality (Zhang et al., 2022c), and Mobile apps (Ahmed et al., 2022b) etc.

There is no work considering edge devices. IoT sensors and edge devices are widely used and almost remain unnoticed by subjects having autism. Some subjects may not like or even damage direct cameras or sensors while playing.

So, smart sensors or edge devices with proven results can be effective in this domain. So far, there is no proven deep-learning method for this domain. Therefore, we need to keep exploring this research and improve the research community.

9. Conclusions

The scope and promise of activity analysis to automatically identify autism using deep learning-based approaches using images and or videos as input were thoroughly investigated and analyzed, as well as summarized in this systematic review. Using the PRISMA procedure, 130 articles were chosen whose primary object was to develop a deep learning tool that is faster, cheaper, and more accurate.

Among these studies, some of the public and private datasets useful to the researcher are extensively discussed and summarized in a table. Deep learning (DL)-based approaches were broadly discussed and summarized in a table, which is employed to extract features from eye gaze, face, MRI, and fMRI image data, eventually performing diagnosis using those extracted features. It is noted that CNN-based approaches are often utilized for feature selection and classification in the research of autism. Moreover, pre-trained CNN models that have been trained with a large number of images can also be employed for autism analysis and makes the researcher's job easier. Autoencoder, LSTM, and RNN were also explored in this review. Furthermore, DL-based approaches for autism rehabilitation were also explored and summarized in this extensive review.

Currently, research is being conducted to identify ASD automatically using DL, although it is still in its early stages. Nonetheless, recent advancements indicate that it is near at hand. Although most automated detection approaches can gather data under unconstrained situations, the results are substantially related to established human ASD procedures. Some limitations might have an impact on overall performance. For example, deep learning can misclassify if large amounts of data are not provided during the training. Again, biases in the data, such as gender or ethnic bias, can affect the model's performance and generalizability. Moreover, ASD is a complicated and heterogeneous disorder with wide variation in appearance, symptoms, severity among individuals, and cognitive features that co-occur with other conditions which may also affect the clinical sessions. Therefore, a large-scale, unbiased dataset with a more generalized model is needed to research autism using deep learning. However, deep learning provides an excellent opportunity for academics to further their study of ASD. We hope this review will be a helpful resource for anyone interested in ASD and deep learning.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

MZU: Conceptualization, writing - review and editing. MZU, AS, and MNM: Study literature review, analysis, and manuscript drafting. MIP and FA: Review and editing. MARA: Conceptualization, review and editing. All authors have read and approved the manuscript.

Acknowledgment

This work is partially supported by the ICT division, Government of the people's Republic of Bangladesh, No: (1280101-120008431-3631108).

References

Aghdam, M.A., Sharifi, A., Pedram, M.M., 2019. Diagnosis of autism spectrum disorders in young children based on resting-state functional magnetic resonance imaging data using convolutional neural networks. Journal of digital imaging 32, 899–918.

Ahmed, I.A., Senan, E.M., Rassem, T.H., Ali, M.A., Shatnawi, H.S.A., Alwazer, S.M., Alshahrani, M., 2022a. Eye tracking-based diagnosis and early detection of autism spectrum disorder using machine learning and deep learning techniques. Electronics 11, 530.

Ahmed, M.R., Zhang, Y., Liu, Y., Liao, H., 2020. Single volume image generator and deep learning-based asd classification. IEEE Journal of Biomedical and Health Informatics 24, 3044–3054.

Ahmed, Z.A., Aldhyani, T.H., Jadhav, M.E., Alzahrani, M.Y., Alzahrani, M.E., Althobaiti, M.M., Alassery, F., Alshaflut, A., Alzahrani, N.M., Al-Madani, A.M., 2022b. Facial features detection system to identify children with autism spectrum disorder: Deep learning models. Computational and Mathematical Methods in Medicine 2022.

Akter, T., Ali, M.H., Khan, M.I., Satu, M.S., Uddin, M.J., Alyami, S.A., Ali, S., Azad, A., Moni, M.A., 2021. Improved transfer-learning-based facial recognition framework to detect autistic children at an early stage. Brain Sciences 11, 734.

Al-Hiyali, M.I., Yahya, N., Faye, I., Hussein, A.F., 2021a. Identification of autism subtypes based on wavelet coherence of bold fmri signals using convolutional neural network. Sensors 21, 5256.

Al-Hiyali, M.I., Yahya, N., Faye, I., Khan, Z., Alsaih, K., 2021b. Classification of bold fmri signals using wavelet transform and transfer learning for detection of autism spectrum disorder, in: 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), IEEE. pp. 94–98.

Alam, M.S., Rashid, M.M., Roy, R., Faizabadi, A.R., Gupta, K.D., Ahsan, M.M., 2022. Empirical study of autism spectrum disorder diagnosis using facial images by improved transfer learning approach. Bioengineering 9, 710.

Ali, A., Negin, F., Bremond, F., Thümmler, S., 2022. Video-based behavior understanding of children for objective diagnosis of autism, in: VISAPP 2022-International Conference on Computer Vision Theory and Applications.

Alkahtani, H., Aldhyani, T.H., Alzahrani, M.Y., 2023. Deep learning algorithms to identify autism spectrum disorder in children-based facial landmarks. Applied Sciences 13, 4855.

Almuqhim, F., Saeed, F., 2021. Asd-saenet: a sparse autoencoder, and deep-neural network model for detecting autism spectrum disorder (asd) using fmri data. Frontiers in Computational Neuroscience 15, 654315.

Alsaade, F.W., Alzahrani, M.S., 2022. Classification and detection of autism spectrum disorder based on deep learning algorithms. Computational Intelligence and Neuroscience 2022.

Anwar, S.M., Majid, M., Qayyum, A., Awais, M., Alnowami, M., Khan, M.K., 2018. Medical image analysis using convolutional neural networks: a review. Journal of medical systems 42, 1–13.

Atyabi, A., Shic, F., Jiang, J., Foster, C.E., Barney, E., Kim, M., Li, B., Ventola, P., Chen, C.H., 2022. Stratification of children with autism spectrum disorder through fusion of temporal information in eye-gaze scan-paths. ACM Transactions on Knowledge Discovery from Data (TKDD) .

Atyabi, A., Shic, F., Jiang, J., Foster, C.E., Barney, E., Kim, M., Li, B., Ventola, P., Chen, C.H., 2023. Stratification of children with autism spectrum disorder through fusion of temporal information in eye-gaze scan-paths. ACM Transactions on Knowledge Discovery from Data 17, 1–20.

Awatramani, J., Hasteer, N., 2020. Facial expression recognition using deep learning for children with autism spectrum disorder, in: 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), IEEE. pp. 35–39.

Baltrusaitis, T., Robinson, P., Morency, L.P., 2013. Constrained local neural fields for robust facial landmark detection in the wild, in: Proceedings of the IEEE international conference on computer vision workshops, pp. 354–361.

Baltrušaitis, T., Robinson, P., Morency, L.P., 2016. Openface: an open source facial behavior analysis toolkit, in: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE. pp. 1–10.

Baltrusaitis, T., Zadeh, A., Lim, Y.C., Morency, L.P., 2018. Openface 2.0: Facial behavior analysis toolkit, in: 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018), IEEE. pp. 59–66.

Banire, B., Al Thani, D., Qaraqe, M., Mansoor, B., 2021. Face-based attention recognition model for children with autism spectrum disorder. Journal of Healthcare Informatics Research 5, 420–445.

Baygin, M., Dogan, S., Tuncer, T., Barua, P.D., Faust, O., Arunkumar, N., Abdulhay, E.W., Palmer, E.E., Acharya, U.R., 2021. Automated asd detection using hybrid deep lightweight features extracted from eeg signals. Computers in Biology and Medicine 134, 104548.

Bayram, M.A., İlyas, Ö., Temurtaş, F., 2021. Deep learning methods for autism spectrum disorder diagnosis based on fmri images. Sakarya University Journal of Computer and Information Sciences 4, 142–155.

Bellec, P., Rosa-Neto, P., Lyttelton, O.C., Benali, H., Evans, A.C., 2010. Multi-level bootstrap analysis of stable clusters in resting-state fmri. Neuroimage 51, 1126–1139.

Berardini, D., Migliorelli, L., Moccia, S., Naldini, M., De Angelis, G., Frontoni, E., 2020. Evaluating the autonomy of children with autism spectrum disorder in washing hands: a deep-learning approach, in: 2020 IEEE Symposium on Computers and Communications (ISCC), IEEE. pp. 1–7.

Bhandage, V., Muppidi, S., Maram, B., et al., 2023. Autism spectrum disorder classification using adam war strategy optimization enabled deep belief network. Biomedical Signal Processing and Control 86, 104914.

Billing, E., Belpaeme, T., Cai, H., Cao, H.L., Ciocan, A., Costescu, C., David, D., Homewood, R., Hernandez Garcia, D., Gómez Esteban, P., et al., 2020. The dream dataset: Supporting a data-driven study of autism spectrum disorder and robot enhanced therapy. PloS one 15, e0236939.

Borji, A., Sihite, D.N., Itti, L., 2012. Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. IEEE Transactions on Image Processing 22, 55–69.

Cai, M., Li, M., Xiong, Z., Zhao, P., Li, E., Tang, J., 2022. An advanced deep learning framework for video-based diagnosis of asd, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 434–444.

Cao, M., Yang, M., Qin, C., Zhu, X., Chen, Y., Wang, J., Liu, T., 2021. Using deepgcn to identify the autism spectrum disorder from multi-site resting-state data. Biomedical Signal Processing and Control 70, 103015.

Cao, X., Ye, W., Sizikova, E., Bai, X., Coffee, M., Zeng, H., Cao, J., 2023. Vitasd: Robust vision transformer baselines for autism spectrum disorder facial diagnosis, in: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE. pp. 1–5.

Cao, Z., Simon, T., Wei, S.E., Sheikh, Y., 2017. Realtime multi-person 2d pose estimation using part affinity fields, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7291–7299.

Carette, R., Elbattah, M., Dequen, G., Guérin, J.L., Cilia, F., 2018. Visualization of eye-tracking patterns in autism spectrum disorder: method and dataset, in: 2018 Thirteenth International Conference on Digital Information Management (ICDIM), IEEE. pp. 248–253.

Chang, Z., Di Martino, J.M., Aiello, R., Baker, J., Carpenter, K., Compton, S., Davis, N., Eichner, B., Espinosa, S., Flowers, J., et al., 2021. Computational methods to measure patterns of gaze in toddlers with autism spectrum disorder. JAMA pediatrics 175, 827–836.

Chen, S., Zhao, Q., 2019. Attention-based autism spectrum disorder screening with privileged modality, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1181–1190.

Chen, Y.H., Chen, Q., Kong, L., Liu, G., 2022. Early detection of autism spectrum disorder in young children with machine learning using medical claims data. BMJ Health & Care Informatics 29, e100544.

Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1251–1258.

Chong, E., Chanda, K., Ye, Z., Southerland, A., Ruiz, N., Jones, R.M., Rozga, A., Rehg, J.M., 2017. Detecting gaze towards eyes in natural social interactions and its use in child assessment. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1, 1–20.

Chrysouli, C., Vretos, N., Daras, P., 2018. Affective state recognition based on eye gaze analysis using two–stream convolutional networks, in: 2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP), IEEE. pp. 1–6.

Cilia, F., Carette, R., Elbattah, M., Dequen, G., Guérin, J.L., Bosche, J., Vandromme, L., Le Driant, B., et al., 2021. Computer-aided screening of autism spectrum disorder: eye-tracking study using data visualization and deep learning. JMIR Human Factors 8, e27706.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine learning 20, 273–297.

Craddock, R.C., James, G.A., Holtzheimer III, P.E., Hu, X.P., Mayberg, H.S., 2012. A whole brain fmri atlas generated via spatially constrained spectral clustering. Human brain mapping 33, 1914–1928.

Daugman, J.G., 1993. High confidence visual recognition of persons by a test of statistical independence. IEEE transactions on pattern analysis and machine intelligence 15, 1148–1161.

De Belen, R.A.J., Bednarz, T., Sowmya, A., 2021. Eyexplain autism: interactive system for eye tracking data analysis and deep neural network interpretation for autism spectrum disorder diagnosis, in: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–7.

De Belen, R.A.J., Bednarz, T., Sowmya, A., Del Favero, D., 2020. Computer vision in autism spectrum disorder research: a systematic review of published studies from 2009 to 2019. Translational psychiatry 10, 1–20.

Del Coco, M., Leo, M., Carcagni, P., Spagnolo, P., Luigi Mazzeo, P., Bernava, M., Marino, F., Pioggia, G., Distante, C., 2017. A computer vision based approach for understanding emotional involvements in children with autism spectrum disorders, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 1401–1407.

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., et al., 2006. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. Neuroimage 31, 968–980.

Devika, K., Mahapatra, D., Subramanian, R., Oruganti, V.R.M., 2022. Dense attentive gan-based one-class model for detection of autism and adhd. Journal of King Saud University-Computer and Information Sciences .

Di Martino, A., O'connor, D., Chen, B., Alaerts, K., Anderson, J.S., Assaf, M., Balsters, J.H., Baxter, L., Beggiato, A., Bernaerts, S., et al., 2017. Enhancing studies of the connectome in autism using the autism brain imaging data exchange ii. Scientific data 4, 1–15.

Di Martino, A., Yan, C.G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., Anderson, J.S., Assaf, M., Bookheimer, S.Y., Dapretto, M., et al., 2014. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular psychiatry 19, 659–667.

Dosenbach, N.U., Nardos, B., Cohen, A.L., Fair, D.A., Power, J.D., Church, J.A., Nelson, S.M., Wig, G.S., Vogel, A.C., Lessov-Schlaggar, C.N., et al., 2010. Prediction of individual brain maturity using fmri. Science 329, 1358–1361.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 .

Du, Y., Li, B., Hou, Y., Calhoun, V.D., 2020. A deep learning fusion model for brain disorder classification: Application to distinguishing schizophrenia and autism spectrum disorder, in: Proceedings of

the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, pp. 1–7.
Duan, H., Min, X., Fang, Y., Fan, L., Yang, X., Zhai, G., 2019a. Visual attention analysis and prediction on human faces for children with autism spectrum disorder. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 15, 1–23.
Duan, H., Zhai, G., Min, X., Che, Z., Fang, Y., Yang, X., Gutiérrez, J., Callet, P.L., 2019b. A dataset of eye movements for the children with autism spectrum disorder, in: Proceedings of the 10th ACM Multimedia Systems Conference, pp. 255–260.
D'Souza, N.S., Nebel, M.B., Crocetti, D., Wymbs, N., Robinson, J., Mostofsky, S., Venkataraman, A., 2020. A deep-generative hybrid model to integrate multimodal and dynamic connectivity for predicting spectrum-level deficits in autism, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 437–447.
Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., Zilles, K., 2005. A new spm toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 25, 1325–1335.
Elakkiya, M.K., Dejey, 2022. Rbm-gp with novel kernels coupled deep learning model for autismscreening br. ENGINEERING APPLICATIONS OF ARTIFICIAL INTELLIGENCE 114.
Elbattah, M., Carette, R., Dequen, G., Guérin, J.L., Cilia, F., 2019. Learning clusters in autism spectrum disorder: Image-based clustering of eye-tracking scanpaths with deep autoencoder, in: 2019 41st Annual international conference of the IEEE engineering in medicine and biology society (EMBC), IEEE. pp. 1417–1420.
Fabiano, D., Canavan, S., Agazzi, H., Hinduja, S., Goldgof, D., 2020. Gaze-based classification of autism spectrum disorder. Pattern Recognition Letters 135, 204–212.
Fang, Y., Huang, H., Wan, B., Zuo, Y., 2019. Visual attention modeling for autism spectrum disorder by semantic features, in: 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), IEEE. pp. 625–628.
Galliver, M., Gowling, E., Farr, W., Gain, A., Male, I., 2017. Cost of assessing a child for possible autism spectrum disorder? an observational study of current practice in child development centres in the uk. BMJ paediatrics open 1.
Ganesh, K., Umapathy, S., Thanaraj Krishnan, P., 2021. Deep learning techniques for automated detection of autism spectrum disorder based on thermal imaging. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine 235, 1113–1127.
Gao, J., Chen, M., Li, Y., Gao, Y., Li, Y., Cai, S., Wang, J., 2021. Multisite autism spectrum disorder classification using convolutional neural network classifier and individual morphological brain networks. Frontiers in Neuroscience 14, 629630.
Gillberg, C., Gillberg, C., Råstam, M., Wentz, E., 2001. The asperger syndrome (and high-functioning autism) diagnostic interview (asdi): a preliminary study of a new structured clinical interview. Autism 5, 57–66.
Girshick, R., Donahue, J., Darrell, T., Malik, J., 2015. Region-based convolutional networks for accurate object detection and segmentation. IEEE transactions on pattern analysis and machine intelligence 38, 142–158.
Groff, E., 1995. Laban movement analysis: Charting the ineffable domain of human movement. Journal of Physical Education, Recreation & Dance 66, 27–30.
Großekathöfer, U., Manyakov, N.V., Mihajlović, V., Pandina, G., Skalkin, A., Ness, S., Bangerter, A., Goodwin, M.S., 2017. Automated detection of stereotypical motor movements in autism spectrum disorder using recurrence quantification analysis. Frontiers in neuroinformatics 11, 9.
Guo, X., Wang, J., Wang, X., Liu, W., Yu, H., Xu, L., Li, H., Wu, J., Dong, M., Tan, W., et al., 2022. Diagnosing autism spectrum disorder in children using conventional mri and apparent diffusion coefficient based deep learning algorithms. European Radiology 32, 761–770.
Hall, D., Huerta, M.F., McAuliffe, M.J., Farber, G.K., 2012. Sharing heterogeneous data: the national database for autism research. Neuroinformatics 10, 331–339.
Han, J., Jiang, G., Ouyang, G., Li, X., 2022. A multimodal approach for identifying autism spectrum disorders in children. IEEE Transactions on Neural Systems and Rehabilitation Engineering 30, 2003–2011.
Han, J., Li, X., Xie, L., Liu, J., Wang, F., Wang, Z., 2018. Affective computing of childern with authism based on feature transfer, in: 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), IEEE. pp. 845–849.
Hao, X., A.Q.L.J.M.H.G.Y.Y.M..Q.J., 2022. Exploring high-order correlations with deep-broad learning for autism spectrum disorder diagnosis. Frontiers in neuroscience , 12.
Haputhanthri, D., Brihadiswaran, G., Gunathilaka, S., Meedeniya, D., Jayarathna, S., Jaime, M., Harshaw, C., 2020. Integration of facial thermography in eeg-based classification of asd. International Journal of Automation and Computing 17, 837–854.
Haque, M.I.U., Valles, D., 2019. Facial expression recognition using dcnn and development of an ios app for children with asd to enhance communication abilities, in: 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), IEEE. pp. 0476–0482.
Heinsfeld, A.S., Franco, A.R., Craddock, R.C., Buchweitz, A., Meneguzzi, F., 2018. Identification of autism spectrum disorder using deep learning and the abide dataset. NeuroImage: Clinical 17, 16–23.
Hinton, G.E., 2009. Deep belief networks. Scholarpedia 4, 5947.
Hluchyj, M.G., Karol, M.J., 1991. Shuffle net: An application of generalized perfect shuffles to multihop lightwave networks. Journal of Lightwave Technology 9, 1386–1397.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural computation 9, 1735–1780.
Hossain, M.D., Kabir, M.A., Anwar, A., Islam, M.Z., 2021. Detecting autism spectrum disorder using machine learning techniques. Health Information Science and Systems 9, 1–13.
Hosseini, M.P., Beary, M., Hadsell, A., Messersmith, R., Soltanian-Zadeh, H., 2021. Deep learning for autism diagnosis and facial analysis in children. Frontiers in Computational Neuroscience 15.
Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 .
Huang, X., Shen, C., Boix, X., Zhao, Q., 2015. Salicon: Reducing the semantic gap in saliency prediction by adapting deep neural networks, in: Proceedings of the IEEE international conference on computer vision, pp. 262–270.
Huang, Z.A., Zhu, Z., Yau, C.H., Tan, K.C., 2020. Identifying autism spectrum disorder from resting-state fmri using deep belief network. IEEE Transactions on neural networks and learning systems 32, 2847–2861.
Hyde, K.K., Novack, M.N., LaHaye, N., Parlett-Pelleriti, C., Anden, R., Dixon, D.R., Linstead, E., 2019. Applications of supervised machine learning in autism spectrum disorder research: a review. Review Journal of Autism and Developmental Disorders 6, 128–146.
Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K., 2016. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size. arXiv preprint arXiv:1602.07360 .

Ismail, M., Barnes, G., Nitzken, M., Switala, A., Shalaby, A., Hosseini-Asl, E., Casanova, M., Keynton, R., Khalil, A., El-Baz, A., 2017. A new deep-learning approach for early detection of shape variations in autism using structural mri, in: 2017 IEEE International Conference on Image Processing (ICIP), IEEE. pp. 1057–1061.

Javed, H., Park, C.H., 2020. Behavior-based risk detection of autism spectrum disorder through child-robot interaction, in: Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, pp. 275–277.

Jiang, B., Chen, S., Wang, B., Luo, B., 2022a. Mglnn: Semi-supervised learning via multiple graph cooperative learning neural networks. Neural Networks 153, 204–214.

Jiang, M., Zhao, Q., 2017. Learning visual attention to identify people with autism spectrum disorder, in: Proceedings of the ieee international conference on computer vision, pp. 3267–3276.

Jiang, W., Liu, S., Zhang, H., Sun, X., Wang, S., Zhao, J., Yan, J., 2022b. Cnng: A convolutional neural networks with gated recurrent units for asd classification. Frontiers in Aging Neuroscience , 723.

Judd, T., Ehinger, K., Durand, F., Torralba, A., 2009. Learning to predict where humans look, in: 2009 IEEE 12th iccv, IEEE. pp. 2106–2113.

Kang, L., Chen, J., Huang, J., Jiang, J., 2022. Autism spectrum disorder recognition based on multi-view ensemble learning with multi-site fmri. Cognitive Neurodynamics , 1–11.

Kanhirakadavath, M.R., Chandran, M.S.M., 2022. Investigation of eye-tracking scan path as a biomarker for autism screening using machine learning algorithms. Diagnostics 12, 518.

Kanner, L., et al., 1943. Autistic disturbances of affective contact. Nervous child 2, 217–250.

Kashef, R., 2022. Ecnn: Enhanced convolutional neural network for efficient diagnosis of autism spectrum disorder. Cognitive Systems Research 71, 41–49.

Ke, F., Yang, R., 2020. Classification and biomarker exploration of autism spectrum disorders based on recurrent attention model. IEEE Access 8, 216298–216307.

Khodatars, M., Shoeibi, A., Sadeghi, D., Ghaasemi, N., Jafari, M., Moridian, P., Khadem, A., Alizadehsani, R., Zare, A., Kong, Y., et al., 2021. Deep learning for neuroimaging-based diagnosis and rehabilitation of autism spectrum disorder: a review. Computers in Biology and Medicine 139, 104949.

Kingma, D.P., Welling, M., 2013. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 .

Kiruthigha, M., Jaganathan, S., 2021. Graph convolutional model to diagnose autism spectrum disorder using rs-fmri data, in: 2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP), IEEE. pp. 1–5.

Kojovic, N., Natraj, S., Mohanty, S.P., Maillart, T., Schaer, M., 2021. Using 2d video-based pose estimation for automated prediction of autism spectrum disorders in young children. Scientific Reports 11, 1–10.

Kong, Y., Gao, J., Xu, Y., Pan, Y., Wang, J., Liu, J., 2019. Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier. Neurocomputing 324, 63–68.

Koumpouros, Y., Kafazis, T., 2019. Wearables and mobile technologies in autism spectrum disorder interventions: A systematic literature review. Research in Autism Spectrum Disorders 66, 101405.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. Imagenet classification with deep convolutional neural networks. Communications of the ACM 60, 84–90.

Kunda, M., Zhou, S., Gong, G., Lu, H., 2020. Improving multi-site autism classification based on site-dependence minimisation and second-order functional connectivity. bioRxiv .

Lakkapragada, A., Kline, A., Mutlu, O.C., Paskov, K., Chrisman, B., Stockham, N., Washington, P., Wall, D.P., et al., 2022. The classification of abnormal hand movement to aid in autism detection: Machine learning study. JMIR Biomedical Engineering 7, e33771.

Lecavalier, L., 2005. An evaluation of the gilliam autism rating scale. Journal of autism and developmental disorders 35, 795–805.

Leo, M., Carcagnì, P., Del Coco, M., Spagnolo, P., Mazzeo, P.L., Celeste, G., Distante, C., Lecciso, F., Levante, A., Rosato, A.C., et al., 2018a. Towards the automatic assessment of abilities to produce facial expressions: The case study of children with asd, in: 20th Italian National Conference on Photonic Technologies (Fotonica 2018), IET. pp. 1–4.

Leo, M., Carcagnì, P., Distante, C., Mazzeo, P.L., Spagnolo, P., Levante, A., Petrocchi, S., Lecciso, F., 2019. Computational analysis of deep visual data for quantifying facial expression production. Applied Sciences 9, 4542.

Leo, M., Carcagnì, P., Distante, C., Spagnolo, P., Mazzeo, P.L., Rosato, A.C., Petrocchi, S., Pellegrino, C., Levante, A., De Lumè, F., et al., 2018b. Computational assessment of facial expression production in asd children. Sensors 18, 3993.

Li, B., Mehta, S., Aneja, D., Foster, C., Ventola, P., Shic, F., Shapiro, L., 2019a. A facial affect analysis system for autism spectrum disorder, in: 2019 IEEE International Conference on Image Processing (ICIP), IEEE.

Li, H., Parikh, N.A., He, L., 2018a. A novel transfer learning approach to enhance deep neural network classification of brain functional connectomes. Frontiers in neuroscience , 491.

Li, J., Zhong, Y., Han, J., Ouyang, G., Li, X., Liu, H., 2020. Classifying asd children with lstm based on raw videos. Neurocomputing 390, 226–238.

Li, M., Li, X., Xie, L., Liu, J., Wang, F., Wang, Z., 2019b. Assisted therapeutic system based on reinforcement learning for children with autism. Computer Assisted Surgery 24, 94–104.

Li, S., Tang, Z., Jin, N., Yang, Q., Liu, G., Liu, T., Hu, J., Liu, S., Wang, P., Hao, J., et al., 2022. Uncovering brain differences in preschoolers and young adolescents with autism spectrum disorder using deep learning. International Journal of Neural Systems , 2250044–2250044.

Li, X., Dvornek, N.C., Papademetris, X., Zhuang, J., Staib, L.H., Ventola, P., Duncan, J.S., 2018b. 2-channel convolutional 3d deep neural network (2cc3d) for fmri analysis: Asd classification and feature learning, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE. pp. 1252–1255.

Li, X., Dvornek, N.C., Zhuang, J., Ventola, P., Duncan, J.S., 2018c. Brain biomarker interpretation in asd using deep learning and fmri, in: International conference on medical image computing and computer-assisted intervention, Springer. pp. 206–214.

Liang, S., Sabri, A.Q.M., Alnajjar, F., Loo, C.K., 2021a. Autism spectrum self-stimulatory behaviors classification using explainable temporal coherency deep features and svm classifier. IEEE Access 9, 34264–34275.

Liang, Y., Liu, B., Zhang, H., 2021b. A convolutional neural network combined with prototype learning framework for brain functional network classification of autism spectrum disorder. IEEE Transactions on Neural Systems and Rehabilitation Engineering 29, 2193–2202.

Liao, D., Lu, H., 2018. Classify autism and control based on deep learning and community structure on resting-state fmri, in: 2018 Tenth International Conference on Advanced Computational Intelligence (ICACI), IEEE. pp. 289–294.

Liaqat, S., Wu, C., Duggirala, P.R., Cheung, S.c.S., Chuah, C.N., Ozonoff, S., Young, G., 2021. Predicting asd diagnosis in children with synthetic and image-based eye gaze data. Signal Processing: Image Communication 94, 116198.

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., Bartlett, M., 2011. The computer expression recognition toolbox (cert), in: 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG), IEEE. pp. 298–305.

Liu, Y., Xu, L., Yu, J., Li, J., Yu, X., 2021. Identification of autism spectrum disorder using multi-regional resting-state data through an attention learning approach. Biomedical Signal Processing and Control 69, 102833.

Liu, Z., Luo, P., Wang, X., Tang, X., 2015. Deep learning face attributes in the wild, in: Proceedings of the IEEE international conference on computer vision.

Lord, C., Elsabbagh, M., Baird, G., Veenstra-Vanderweele, J., 2018. Seminar autism spectrum disorder. Lancet 392, 508–20.

Lord, C., Risi, S., DiLavore, P.S., Shulman, C., Thurm, A., Pickles, A., 2006. Autism from 2 to 9 years of age. Archives of general psychiatry 63, 694–701.

Lord, C., Rutter, M., Le Couteur, A., 1994. Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. Journal of autism and developmental disorders 24, 659–685.

Lu, A., Perkowski, M., 2021. Deep learning approach for screening autism spectrum disorder in children with facial images and analysis of ethnoracial factors in model development and application. Brain Sciences 11, 1446.

Lu, X., Tsao, Y., Matsuda, S., Hori, C., 2013. Speech enhancement based on deep denoising autoencoder., in: Interspeech, pp. 436–440.

Lu, Z., Wang, J., Mao, R., Lu, M., Shi, J., 2022. Jointly composite feature learning and autism spectrum disorder classification using deep multi-output takagi-sugeno-kang fuzzy inference systems. IEEE/ACM Transactions on Computational Biology and Bioinformatics .

Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I., 2010. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression, in: 2010 ieee computer society conference on computer vision and pattern recognition-workshops, IEEE. pp. 94–101.

Lundqvist, D., Flykt, A., Öhman, A., 1998. Karolinska directed emotional faces. Cognition and Emotion .

Maenner, M.J., Shaw, K.A., Bakian, A.V., Bilder, D.A., Durkin, M.S., Esler, A., Furnier, S.M., Hallas, L., Hall-Lande, J., Hudson, A., et al., 2021. Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2018. MMWR Surveillance Summaries 70, 1.

Marinoiu, E., Zanfir, M., Olaru, V., Sminchisescu, C., 2018. 3d human sensing, action and emotion recognition in robot assisted therapy of children with autism, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2158–2167.

Mayor-Torres, J.M., Medina-DeVilliers, S., Clarkson, T., Lerner, M.D., Riccardi, G., 2021. Evaluation of interpretability for deep learning algorithms in eeg emotion recognition: A case study in autism. arXiv preprint arXiv:2111.13208 .

Medsker, L.R., Jain, L., 2001. Recurrent neural networks. Design and Applications 5, 64–67.

Mellema, C., Treacher, A., Nguyen, K., Montillo, A., 2019. Multiple deep learning architectures achieve superior performance diagnosing autism spectrum disorder using features previously extracted from structural and functional mri, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), IEEE. pp. 1891–1895.

Milano, N., Simeoli, R., Rega, A., Marocco, D., 2023. A deep learning latent variable model to identify children with autism through motor abnormalities. Frontiers in Psychology 14.

Minissi, M.E., Chicchi Giglioli, I.A., Mantovani, F., Alcaniz Raya, M., 2021. Assessment of the autism spectrum disorder based on machine learning and social visual attention: A systematic review. Journal of Autism and Developmental Disorders , 1–16.

Mishra, M., Pati, U.C., 2023. A classification framework for autism spectrum disorder detection using smri: Optimizer based ensemble of deep convolution neural network with on-the-fly data augmentation. Biomedical Signal Processing and Control 84, 104686.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., Group*, P., 2009. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. Annals of internal medicine 151, 264–269.

Mostafa, S., Tang, L., Wu, F.X., 2019a. Diagnosis of autism spectrum disorder based on eigenvalues of brain networks. Ieee Access 7, 128474–128486.

Mostafa, S., Yin, W., Wu, F.X., 2019b. Autoencoder based methods for diagnosis of autism spectrum disorder, in: International Conference on Computational Advances in Bio and Medical Sciences, Springer. pp. 39–51.

Mujeeb Rahman, K., Subashini, M.M., 2022. Identification of autism in children using static facial features and deep neural networks. Brain Sciences 12, 94.

Ng, A., et al., 2011. Sparse autoencoder. CS294A Lecture notes 72, 1–19.

Niu, K., Guo, J., Pan, Y., Gao, X., Peng, X., Li, N., Li, H., 2020. Multichannel deep attention neural networks for the classification of autism spectrum disorder using neuroimaging and personal characteristic data. Complexity 2020.

Nogay, H.S., Adeli, H., 2023. Diagnostic of autism spectrum disorder based on structural brain mri images using, grid search optimization, and convolutional neural networks. Biomedical Signal Processing and Control 79, 104234.

Noorbakhsh-Sabet, N., Zand, R., Zhang, Y., Abedi, V., 2019. Artificial intelligence transforms the future of health care. The American journal of medicine 132, 795–801.

Othmani, A., Bizet, T., Pellerin, T., Hamdi, B., Bock, M.A., Dev, S., 2023. Significant cc400 functional brain parcellations based lenet5 convolutional neural network for autism spectrum disorder detection, in: Recent Trends in Image Processing and Pattern Recognition: 5th International Conference, RTIP2R 2022, Kingsville, TX, USA, December 1-2, 2022, Revised Selected Papers, Springer. pp. 34–45.

Pan, J., Ferrer, C.C., McGuinness, K., O'Connor, N.E., Torres, J., Sayrol, E., Giro-i Nieto, X., 2017. Salgan: Visual saliency prediction with generative adversarial networks. arXiv preprint arXiv:1701.01081 .

Pandian, D., Rajagopalan, S.S., Jayagopi, D., et al., 2022. Detecting a child's stimming behaviours for autism spectrum disorder diagnosis using rgbpose-slowfast network, in: 2022 IEEE International Conference on Image Processing (ICIP), IEEE. pp. 3356–3360.

Park, K.W., Cho, S.B., 2023. A residual graph convolutional network with spatio-temporal features for autism classification from fmri brain images. Applied Soft Computing 142, 110363.

Parkhi, O.M., Vedaldi, A., Zisserman, A., 2015. Deep face recognition .

Parlett-Pelleriti, C.M., Stevens, E., Dixon, D., Linstead, E.J., 2022. Applications of unsupervised machine learning in autism spectrum disorder research: a review. Review Journal of Autism and Developmental Disorders , 1–16.

Patnam, V.S.P., George, F.T., George, K., Verma, A., 2017. Deep learning based recognition of meltdown in autistic kids, in: 2017 IEEE International Conference on Healthcare Informatics (ICHI), IEEE. pp. 391–396.

Pavithra, R., Abirami, S., Krithika, S., Sabitha, S., Tharanidharan, P., 2023. Identification of autism spectrum disorder from functional mri using deep learning, in: Advances in Information Communication Technology and Computing: Proceedings of AICTC 2022.

Springer, pp. 277–284.
PEH, D., 2001. Ro duda, pe hart, and dg stork, pattern classification.
Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J., 2000. The feret evaluation methodology for face-recognition algorithms. IEEE Transactions on pattern analysis and machine intelligence 22, 1090–1104.
Pickles, A., Le Couteur, A., Leadbitter, K., Salomone, E., Cole-Fletcher, R., Tobin, H., Gammer, I., Lowry, J., Vamvakas, G., Byford, S., et al., 2016. Parent-mediated social communication therapy for young children with autism (pact): long-term follow-up of a randomised controlled trial. The Lancet 388, 2501–2509.
Piosenka, G., 2021. Detect autism from a facial image.
Popa, A.I., Zanfir, M., Sminchisescu, C., 2017. Deep multitask architecture for integrated 2d and 3d human sensing, in: Proceedings of the ieee conference on computer vision and pattern recognition, pp. 6289–6298.
Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., et al., 2011. Functional network organization of the human brain. Neuron 72, 665–678.
Prakash, V.G., Kohli, M., Kohli, S., Prathosh, A., Wadhera, T., Das, D., Panigrahi, D., Kommu, J.V.S., 2023. Computer vision-based assessment of autistic children: Analyzing interactions, emotions, human pose, and life skills. IEEE Access .
Pugazhenthi, B., Senapathy, G., Pavithra, M., 2019. Identification of autism in mr brain images using deep learning networks, in: 2019 International Conference on Smart Structures and Systems (ICSSS), IEEE. pp. 1–7.
Rabbi, M.F., Zohra, F.T., Hossain, F., Akhi, N.N., Khan, S., Mahbub, K., Biswas, M., 2023. Autism spectrum disorder detection using transfer learning with vgg 19, inception v3 and densenet 201, in: Recent Trends in Image Processing and Pattern Recognition: 5th International Conference, RTIP2R 2022, Kingsville, TX, USA, December 1-2, 2022, Revised Selected Papers, Springer. pp. 190–204.
Rahman, S., Ahmed, S.F., Shahid, O., Arrafi, M.A., Ahad, M., 2021. Automated detection approaches to autism spectrum disorder based on human activity analysis: A review. Cognitive Computation , 1–28.
Rajagopalan, S., Dhall, A., Goecke, R., 2013. Self-stimulatory behaviours in the wild for autism diagnosis, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 755–761.
Rani, P., 2019. Emotion detection of autistic children using image processing, in: 2019 Fifth International Conference on Image Information Processing (ICIIP), IEEE. pp. 532–535.
Rathore, A., Palande, S., Anderson, J.S., Zielinski, B.A., Fletcher, P.T., Wang, B., 2019. Autism classification using topological features and deep learning: A cautionary tale, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 736–744.
Rehg, J., Abowd, G., Rozga, A., Romero, M., Clements, M., Sclaroff, S., Essa, I., Ousley, O., Li, Y., Kim, C., et al., 2013. Decoding children's social behavior, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3414–3421.
Roy, A.M., Bhaduri, J., 2023. A computer vision enabled damage detection model with improved yolov5 based on transformer prediction head. arXiv preprint arXiv:2303.04275 .
Rudovic, O., Lee, J., Mascarell-Maricic, L., Schuller, B.W., Picard, R.W., 2017. Measuring engagement in robot-assisted autism therapy: a cross-cultural study. Frontiers in Robotics and AI 4, 36.
Rudovic, O., Utsumi, Y., Lee, J., Hernandez, J., Ferrer, E.C., Schuller, B., Picard, R.W., 2018. Culturenet: a deep learning approach for engagement intensity estimation from face images of children with autism, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE. pp. 339–346.
Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1985. Learning internal representations by error propagation. Technical Report. California Univ San Diego La Jolla Inst for Cognitive Science.
Rutter, M., Le Couteur, A., Lord, C., et al., 2003. Autism diagnostic interview-revised. Los Angeles, CA: Western Psychological Services 29, 30.
Sabegh, A.M., Samadzadehaghdam, N., Seyedarabi, H., Ghadiri, T., 2023. Automatic detection of autism spectrum disorder based on fmri images using a novel convolutional neural network. Research on Biomedical Engineering , 1–7.
Salhi, I., Qbadou, M., Gouraguine, S., Mansouri, K., Lytridis, C., Kaburlasos, V., 2022. Towards robot-assisted therapy for children with autism—the ontological knowledge models and reinforcement learning-based algorithms. Frontiers in Robotics and AI 9.
Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4510–4520.
Sapiro, G., Hashemi, J., Dawson, G., 2019. Computer vision and behavioral phenotyping: an autism case study. Current Opinion in Biomedical Engineering 9, 14–20.
Saranya, A., Anandan, R., 2021. Figs-deaf: an novel implementation of hybrid deep learning algorithm to predict autism spectrum disorders using facial fused gait features. Distributed and Parallel Databases , 1–26.
Schopler, E., Reichler, R.J., DeVellis, R.F., Daly, K., 1980. Toward objective classification of childhood autism: Childhood autism rating scale (cars). Journal of autism and developmental disorders .
Sewani, H., Kashef, R., 2020. An autoencoder-based deep learning classifier for efficient diagnosis of autism. Children 7, 182.
Sharif, H., Khan, R.A., 2022. A novel machine learning based framework for detection of autism spectrum disorder (asd). Applied Artificial Intelligence 36, 2004655.
Shen, J., Ainger, E., Alcorn, A., Dimitrijevic, S.B., Baird, A., Chevalier, P., Cummins, N., Li, J., Marchi, E., Marinoiu, E., et al., 2018. Autism data goes big: A publicly-accessible multi-modal database of child interactions for behavioural and machine learning research, in: International Society for Autism Research Annual Meeting.
Sherkatghanad, Z., Akhondzadeh, M., Salari, S., Zomorodi-Moghadam, M., Abdar, M., Acharya, U.R., Khosrowabadi, R., Salari, V., 2020. Automated detection of autism spectrum disorder using a convolutional neural network. Frontiers in neuroscience 13, 1325.
Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., Blake, A., 2011. Real-time human pose recognition in parts from single depth images, in: CVPR 2011, Ieee. pp. 1297–1304.
Shukla, P., Gupta, T., Saini, A., Singh, P., Balasubramanian, R., 2017. A deep learning frame-work for recognizing developmental disorders, in: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE. pp. 705–714.
Simonyan, K., Zisserman, A., 2014a. Two-stream convolutional networks for action recognition in videos. Advances in neural information processing systems 27.
Simonyan, K., Zisserman, A., 2014b. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .
Singh, A., Raj, K., Kumar, T., Verma, S., Roy, A.M., 2023. Deep learning-based cost-effective and responsive robot for autism treatment. Drones 7, 81.
Skuse, D., Warrington, R., Bishop, D., Chowdhury, U., Lau, J., Mandy, W., Place, M., 2004. The developmental, dimensional and diagnostic interview (3di): a novel computerized assessment for

autism spectrum disorders. Journal of the American Academy of Child & Adolescent Psychiatry 43, 548–558.
Smith, B.A., Yin, Q., Feiner, S.K., Nayar, S.K., 2013. Gaze locking: passive eye contact detection for human-object interaction, in: Proceedings of the 26th annual ACM symposium on User interface software and technology, pp. 271–280.
Song, D.Y., Topriceanu, C.C., Ilie-Ablachim, D.C., Kinali, M., Bisdas, S., 2021. Machine learning with neuroimaging data to identify autism spectrum disorder: a systematic review and meta-analysis. Neuroradiology 63, 2057–2072.
Subah, F.Z., Deb, K., Dhar, P.K., Koshiba, T., 2021. A deep learning approach to predict autism spectrum disorder using multisite resting-state fmri. Applied Sciences 11, 3636.
Sun, K., Li, L., Li, L., He, N., Zhu, J., 2020. Spatial attentional bilinear 3d convolutional network for video-based autism spectrum disorder detection, in: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE. pp. 3387–3391.
Talairach, J., 1988. Co-planar stereotaxic atlas of the human brain-3-dimensional proportional system. An approach to cerebral imaging .
Tamilarasi, F.C., Shanmugam, J., 2020. Convolutional neural network based autism classification, in: 2020 5th International Conference on Communication and Electronics Systems (ICCES), IEEE. pp. 1208–1212.
Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks, in: International conference on machine learning, PMLR. pp. 6105–6114.
Tanaka, J.W., Sung, A., 2016. The "eye avoidance" hypothesis of autism face processing. Journal of autism and developmental disorders 46, 1538–1552.
Tang, C., Zheng, W., Zong, Y., Cui, Z., Qiu, N., Yan, S., Ke, X., 2018. Automatic smile detection of infants in mother-infant interaction via cnn-based feature learning, in: Proceedings of the Joint Workshop of the 4th Workshop on Affective Social Multimedia Computing and first Multi-Modal Affective Computing of Large-Scale Multimedia Data, pp. 35–40.
Tang, M., Kumar, P., Chen, H., Shrivastava, A., 2020. Deep multimodal learning for the diagnosis of autism spectrum disorder. Journal of Imaging 6, 47.
Tanguay, P.E., Robertson, J., Derrick, A., 1998. A dimensional classification of autism spectrum disorder by social communication domains. Journal of the American Academy of Child & Adolescent Psychiatry 37, 271–277.
Tao, Y., Shyu, M.L., 2019. Sp-asdnet: Cnn-lstm based asd classification model using observer scanpaths, in: 2019 IEEE International conference on multimedia & expo workshops (ICMEW), IEEE. pp. 641–646.
Tawhid, M.N.A., Siuly, S., Wang, H., Whittaker, F., Wang, K., Zhang, Y., 2021. A spectrogram image based intelligent technique for automatic detection of autism spectrum disorder from eeg. Plos one 16, e0253094.
Thabtah, F., 2019. An accessible and efficient autism screening method for behavioural data and predictive analyses. Health informatics journal 25, 1739–1755.
Thomas, R.M., Gallo, S., Cerliani, L., Zhutovsky, P., El-Gazzar, A., Van Wingen, G., 2020. Classifying autism spectrum disorder using the temporal statistics of resting-state functional mri data with 3d convolutional neural networks. Frontiers in psychiatry 11, 440.
Toro, R., Traut, N., Beggatio, A., Heuer, K., Varoquaux, G., et al., 2018. Impac: Imaging-psychiatry challenge: predicting autism, a data challenge on autism spectrum disorder detection. Online Challenge .
De la Torre Frade, F., Chu, W.S., Xiong, X., Carrasco, F.V., Ding, X., Cohn, J., 2015. Intraface, in: Automatic face and gesture recognition.
Torres, J.M.M., Clarkson, T., Hauschild, K.M., Luhmann, C.C., Lerner, M.D., Riccardi, G., 2022. Facial emotions are accurately encoded in the neural signal of those with autism spectrum disorder: A deep learning approach. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging 7, 688–695.
Tran, D., Wang, H., Torresani, L., Ray, J., LeCun, Y., Paluri, M., 2018. A closer look at spatiotemporal convolutions for action recognition, in: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 6450–6459.
Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. Neuroimage 15, 273–289.
Uddin, M., Muramatsu, D., Kimura, T., Makihara, Y., Yagi, Y., et al., 2017. Multiq: single sensor-based multi-quality multi-modal large-scale biometric score database and its performance evaluation. IPSJ Transactions on Computer Vision and Applications 9, 1–25.
Uddin, M., Muramatsu, D., Takemura, N., Ahad, M., Rahman, A., Yagi, Y., et al., 2019. Spatio-temporal silhouette sequence reconstruction for gait recognition against occlusion. IPSJ Transactions on Computer Vision and Applications 11, 1–18.
Uddin, M., Ngo, T.T., Makihara, Y., Takemura, N., Li, X., Muramatsu, D., Yagi, Y., et al., 2018. The ou-isir large population gait database with real-life carried object and its performance evaluation. IPSJ Transactions on Computer Vision and Applications 10, 1–11.
Varoquaux, G., Gramfort, A., Pedregosa, F., Michel, V., Thirion, B., 2011. Multi-subject dictionary learning to segment an atlas of brain spontaneous activity, in: Biennial International Conference on information processing in medical imaging, Springer. pp. 562–573.
Wadhera, T., Mahmud, M., Brown, D.J., 2023. A deep concatenated convolutional neural network-based method to classify autism, in: Neural Information Processing: 29th International Conference, ICONIP 2022, Virtual Event, November 22–26, 2022, Proceedings, Part VII, Springer. pp. 446–458.
Wang, C., Xiao, Z., Wang, B., Wu, J., 2019a. Identification of autism based on svm-rfe and stacked sparse auto-encoder. Ieee Access 7, 118030–118036.
Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., et al., 2020a. Deep high-resolution representation learning for visual recognition. IEEE transactions on pattern analysis and machine intelligence 43, 3349–3364.
Wang, S., Jiang, M., Duchesne, X.M., Laugeson, E.A., Kennedy, D.P., Adolphs, R., Zhao, Q., 2015. Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking. Neuron 88, 604–616.
Wang, W., Gang, J., 2018. Application of convolutional neural network in natural language processing, in: 2018 International Conference on Information Systems and Computer Aided Education (ICISCAE), IEEE. pp. 64–70.
Wang, W., Shen, J., Guo, F., Cheng, M.M., Borji, A., 2018. Revisiting video saliency: A large-scale benchmark and a new model, in: Proceedings of the IEEE Conference on computer vision and pattern recognition, pp. 4894–4903.
Wang, Y., Liu, J., Xiang, Y., Wang, J., Chen, Q., Chong, J., 2022. Mage: automatic diagnosis of autism spectrum disorders using multi-atlas graph convolutional networks and ensemble learning. Neurocomputing 469, 346–353.
Wang, Y., Wang, J., Wu, F.X., Hayrat, R., Liu, J., 2020b. Aimafe: Autism spectrum disorder identification with multi-atlas deep feature representation and ensemble learning. Journal of neuroscience methods 343, 108840.
Wang, Z., Xu, K., Liu, H., 2019b. Screening early children with

autism spectrum disorder via expressing needs with index finger pointing, in: Proceedings of the 13th International Conference on Distributed Smart Cameras, pp. 1–6.
Wei, W., Liu, Z., Huang, L., Nebout, A., Le Meur, O., 2019. Saliency prediction via multi-level features and deep supervision for children with autism spectrum disorder, in: 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), IEEE. pp. 621–624.
Wei, W., Liu, Z., Huang, L., Wang, Z., Chen, W., Zhang, T., Wang, J., Xu, L., 2021. Identify autism spectrum disorder via dynamic filter and deep spatiotemporal feature extraction. Signal Processing: Image Communication 94, 116195.
Wen, G., Cao, P., Bao, H., Yang, W., Zheng, T., Zaiane, O., 2022. Mvs-gcn: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis. Computers in Biology and Medicine 142, 105239.
Wilkinson, K.M., 1998. Profiles of language and communication skills in autism. Mental retardation and developmental disabilities research reviews 4, 73–79.
Wloka, C., Kotseruba, I., Tsotsos, J.K., 2017. Saccade sequence prediction: Beyond static saliency maps. arXiv preprint arXiv:1711.10959 .
Wojke, N., Bewley, A., Paulus, D., 2017. Simple online and realtime tracking with a deep association metric, in: 2017 IEEE international conference on image processing (ICIP), IEEE. pp. 3645–3649.
Wu, C., Liaqat, S., Cheung, S.c., Chuah, C.N., Ozonoff, S., 2019. Predicting autism diagnosis using image with fixations and synthetic saccade patterns, in: 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), IEEE. pp. 647–650.
Wu, C., Liaqat, S., Helvaci, H., Chcung, S.c.S., Chuah, C.N., Ozonoff, S., Young, G., 2021. Machine learning based autism spectrum disorder detection from videos, in: 2020 IEEE International Conference on E-health Networking, Application & Services (HEALTHCOM), IEEE. pp. 1–6.
Xia, C., Chen, K., Li, K., Li, H., 2020. Identification of autism spectrum disorder via an eye-tracking based representation learning model, in: 2020 7th International Conference on Bioinformatics Research and Applications, pp. 59–65.
Xu, J., Jiang, M., Wang, S., Kankanhalli, M.S., Zhao, Q., 2014. Predicting human gaze beyond pixels. Journal of vision 14, 28–28.
Yang, X., Schrader, P.T., Zhang, N., 2020. A deep neural network study of the abide repository on autism spectrum classification. International Journal of Advanced Computer Science and Applications 11.
Yang, X., Zhang, N., Schrader, P., 2022. A study of brain networks for autism spectrum disorder classification using resting-state functional connectivity. Machine Learning with Applications 8, 100290.
Ye, Z., Li, Y., Liu, Y., Bridges, C., Rozga, A., Rehg, J.M., 2015. Detecting bids for eye contact using a wearable camera, in: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), IEEE. pp. 1–8.
Yin, W., Mostafa, S., Wu, F.X., 2021. Diagnosis of autism spectrum disorder based on functional brain networks with deep learning. Journal of Computational Biology 28, 146–165.
Yoo, H.J., 2015. Deep convolution neural networks in computer vision: a review. IEIE Transactions on Smart Processing and Computing 4, 35–43.
Zhang, B., Zhou, L., Song, S., Chen, L., Jiang, Z., Zhang, J., 2020. Image captioning in chinese and its application for children with autism spectrum disorder, in: Proceedings of the 2020 12th International Conference on Machine Learning and Computing, pp. 426–432.
Zhang, F., Wei, Y., Liu, J., Wang, Y., Xi, W., Pan, Y., 2022a. Identification of autism spectrum disorder based on a novel feature selection method and variational autoencoder. arXiv preprint arXiv:2204.03654 .
Zhang, J., Feng, F., Han, T., Gong, X., Duan, F., 2022b. Detection of autism spectrum disorder using fmri functional connectivity with feature selection and deep learning. Cognitive Computation , 1–12.
Zhang, M., Ding, H., Naumceska, M., Zhang, Y., 2022c. Virtual reality technology as an educational and intervention tool for children with autism spectrum disorder: current perspectives and future directions. Behavioral Sciences 12, 138.
Zunino, A., Morerio, P., Cavallo, A., Ansuini, C., Podda, J., Battaglia, F., Veneselli, E., Becchio, C., Murino, V., 2018. Video gesture analysis for autism spectrum disorder detection, in: 2018 24th international conference on pattern recognition (ICPR), IEEE. pp. 3421–3426.