Deep Learning for Neuroimaging-based Diagnosis and Rehabilitation of Autism Spectrum Disorder: A Review

Marjane Khodatars, Afshin Shoeibi, Delaram Sadeghi, Navid Ghassemi, Mahboobeh Jafari, Parisa Moridian, Ali Khadem, Roohallah Alizadehsani, Assef Zare, Yinan Kong, Abbas Khosravi, Saeid Nahavandi, Sadiq Hussain, U. Rajendra Acharya, Michael Berk

Abstract—Accurate diagnosis of Autism Spectrum Disorder (ASD) followed by effective rehabilitation is essential for the management of this disorder. Artificial intelligence (AI) techniques can aid physicians to apply automatic diagnosis and rehabilitation procedures. AI techniques comprise traditional machine learning (ML) approaches and deep learning (DL) techniques. Conventional ML methods employ various feature extraction and classification techniques, but in DL, the process of feature extraction and classification is accomplished intelligently and integrally. DL methods for diagnosis of ASD have been focused on neuroimaging-based approaches. Neuroimaging techniques are non-invasive disease markers potentially useful for ASD diagnosis. Structural and functional neuroimaging techniques provide physicians substantial information about the structure (anatomy and structural connectivity) and function (activity and functional connectivity) of the brain. Due to the intricate structure and function of the brain, proposing optimum procedures for ASD diagnosis with neuroimaging data without exploiting powerful AI techniques like DL may be challenging. In this paper, studies conducted with the aid of DL networks to distinguish ASD are investigated. Rehabilitation tools provided for supporting ASD patients utilizing DL networks are also assessed. Finally, we will present important challenges in the automated detection and rehabilitation of ASD and propose some future works.

Index Terms—Autism Spectrum Disorder, Diagnosis, Rehabilitation, Deep Learning, Neuroimaging, Neuroscience.

I. INTRODUCTION

ASD is a disorder of the nervous system that affects the brain and results in difficulties in speech, social interaction and communication deficits, repetitive behaviors, and delays in motor abilities [1]. This disease can generally be distinguished with extant diagnostic protocols from the age of three years onwards. Autism influences many parts of the brain. This disorder also involves a genetic influence via the gene interactions or polymorphisms [2], [3]. One in 70 children worldwide is affected by autism. In 2018, the prevalence of ASD was estimated to occur in 168 out of 10,000 children in the United States, one of the highest prevalence rates worldwide. Autism is significantly more common in boys than in girls. In the United States, about 3.63 percent of boys aged 3 to 17 years have autism spectrum disorder, compared with approximately 1.25 percent of girls [4].

Diagnosing ASD is difficult because there is no pathophysiological marker, relying instead just on psychological criteria [5]. Psychological tools can identify individual behaviors, levels of social interaction, and consequently facilitate early diagnosis. Behavioral evaluations embrace various instruments and questionnaires to assist the physicians to specify the particular type of delay in a child’s development, including clinical observations, medical history, autism diagnosis instructions, and growth and intelligence tests [6].

Several investigations for the diagnosis of ASD have recently been conducted on neuroimaging data (structural and functional).

Analyzing anatomy and structural connections of brain areas with structural neuroimaging is an essential tool for studying structural disorders of the brain in ASD. The principal tools for structural brain imaging are magnetic resonance imaging (MRI) techniques [7], [8], [9]. Cerebral anatomy is investigated by structural MRI (sMRI) images and anatomical connections are assessed by diffusion tensor imaging MRI (DTI-MR) [10]. Investigating the activity and functional connections of brain areas using functional neuroimaging can also be used for studying ASD. Brain functional diagnostic
tools are older approaches than structural methods for studying ASD. The most basic modality of functional neuroimaging is electroencephalography (EEG), which records the electrical activity of the brain from the scalp with a high temporal resolution (in milliseconds order) [11]. Studies have shown that employing EEG signals to diagnose ASD have been useful [12], [13], [14]. Functional MRI (fMRI) is one of the most promising imaging modalities in functional brain disorders, used as task-based (T-fMRI) or resting-state (rs-fMRI) [15], [16]. fMRI-based techniques have a high spatial resolution (in the order of millimeters) but a low temporal resolution due to slow response of the hemodynamic system of the brain as well as fMRI imaging time constraints and is not ideal for recording the fast dynamics of brain activities. In addition, these techniques have a high sensitivity to motion artifacts. It should be stressed that in consonance with studies, three less prevalent modalities of electrocorticography (ECoG) [17], functional near-infrared spectroscopy (fNIRS) [18], and Magnetoencephalography (MEG) [19] can also attain reasonable performance in ASD diagnosis. An appropriate approach is to utilize machine-learning techniques alongside functional and structural data to collaborate with physicians in the process of accurately assessing ASD. In the field of ASD, applying machine learning methods generally entail two categories of traditional methods [20] and DL methods [21]. As opposed to traditional methods, much less work has been done on DL methods to explore ASD or design rehabilitation tools.

This study reviews ASD assessment methods and patients’ rehabilitation with DL networks. The outline of this paper is as follows. Section 2 is search strategy. Section 3 concisely presents the DL networks employed in the field of ASD. In section 4, existing computer-aided diagnosis systems (CADS) that use brain functional and structural data are reviewed. In section 5, DL-based rehabilitation tools for supporting ASD patients are introduced. Section 6 discusses the reviewed papers. Section 7 reveals the challenges of ASD diagnosis and rehabilitation with DL. Finally, the paper concludes and suggests future work in section 8.

II. SEARCH STRATEGY

In this review, IEEE Xplore, ScienceDirect, SpringerLink, ACM, as well as other conferences or journals were used to acquire papers on ASD diagnosis and rehabilitation using DL methods. Further, the keywords “ASD”, “Autism Spectrum Disorder” and “Deep Learning” were used to select the papers. The papers are analyzed till June 03th, 2020 by the authors (AK, SN). Figure 1 depicts the number of considered papers using DL methods for the automated detection and rehabilitation of ASD each year.

III. DEEP LEARNING TECHNIQUES FOR ASD DIAGNOSIS AND REHABILITATION

Nowadays, DL algorithms are used in many areas of medicine including structural and functional neuroimaging. The application of DL in neural imaging ranges from brain MR image segmentation [22], to detection of brain lesions such as tumors [23], diagnosis of brain functional disorders such as ASD [24], and production of artificial structural or functional brain images [25]. Machine learning techniques are categorized into three fundamental categories of learning: supervised learning [26], unsupervised learning [27], and reinforcement learning [28], and a variety of DL networks are provided for each type. So far, most studies applied to identify ASD using DL have been based on supervised or unsupervised approaches. Figure 2 illustrates generally employed types of DL networks with supervised or unsupervised learning to study ASD.

IV. CADS-BASED DEEP LEARNING TECHNIQUES FOR ASD DIAGNOSIS BY NEUROIMAGING DATA

A traditional artificial intelligence (AI)-based CADS encompasses several stages of data acquisition, data preprocessing, feature extraction, and classification [29], [30], [31], [32]. In [33], [34], [35], existing traditional algorithms for diagnosing ASD have been reviewed. In contrast to traditional methods, in DL-based CADS, feature extraction, and classification are performed intelligently within the model. Also, due to the structure of DL networks, using large dataset to train DL networks and recognize intricate patterns in datasets is incumbent. The components of DL-based CADS for ASD detection are shown in Figure 3. It can be noted from the figure that, large and free databases are first introduced to diagnose ASD. In the second step, various types of preprocessing techniques are used on functional and structural data to be scrutinized. Finally, the DL networks are applied on the preprocessed data.
A. Neuroimaging ASD Datasets

Datasets are the heart of any CADS development and the capability of CADS depends primarily on the affluence of the input data. To diagnose ASD, several brain functional and structural datasets are available. The most complete free dataset available is ABIDE [36] dataset with two subsets: ABIDE-I and ABIDE-II, which encompasses sMRI, rs-fMRI, and phenotypic data. ABIDE-I involves data from 17 international sites, yielding a total of 1112 datasets, including 539 individuals with ASD and 573 healthy individuals (ages 6-7). In accordance with HIPAA guidelines and 1000 FCP / INDI protocols, these data are anonymized. In contrast, ABIDE-II contains data from 19 international sites, with a total of 1114 datasets from 521 individuals with ASD and 593 healthy individuals (ages 5-64). Also, preprocessed images of the ABIDE-I series called PCP [37] can be freely downloaded by the researchers. The second recently released ASD diagnostic database is called NDAR, which comprises various modalities, and more information is provided in [38].

B. Preprocessing Techniques

Neuroimaging data (especially functional ones) is of relatively complicated structure, and if not pre-processed properly, it may affect the final diagnosis. Preprocessing of this data typically entails multiple common steps performed by different software as standard. Indeed, occasionally prepared pipelines are applied on the dataset to yield pre-processed data for future researches. In the following section, preprocessing steps are briefly explained for fMRI data.

1) Standard (Low-level) fMRI preprocessing steps: Low-level pre-processing of fMRI images normally has fixed number of steps applied on the data, and prepared toolboxes are usually used to reduce execution time and yield better accuracy. Some of these reputable toolboxes contain FMRIB software libraries (FSL) [49], BET [40], FreeSurfer [41], and SPM [42]. Also, important and vital fMRI preprocessing incorporates brain extraction, spatial smoothing, temporal filtering, motion correction, slice timing correction, intensity normalization, and registration to standard atlas, which are summarized as follows:

**BRAIN EXTRACTION:** the goal is to remove the skull and cerebellum from the fMRI image and maintain the brain tissue [43], [44], [45].

**SPATIAL SMOOTHING:** involves averaging the adjacent voxels signal. This process is persuasive on account of neighbor-

**INTENSITY NORMALIZATION:** at this stage, the average intensity of fMRI signals are rescaled to compensate for global deviations within and between the recording sessions [43], [44], [45].

**REGISTRATION TO A STANDARD ATLAS:** The human brain entails hundreds of cortical and subcortical areas with various structures and functions, each of which is very time-consuming and complex to study. To overcome the problem, brain atlases are employed to partition brain images into a manifold of atlases, including Automated Anatomical Labeling (AAL) [47], Eickhoff-Zilles (EZ) [48], Harvard-Oxford (HO) [49], Talairach and Tournoux (TT) [50], Dosenbach 160 [51], Craddock 200 (CC200) [52] and Craddock 400 (CC400) [53] and more information is provided in [54]. Table I provides complete information on preprocessing tools, atlases, and some other preprocessing information.

2) Pipeline Methods: Pipelines present preprocessed images of ABIDE databases. They embrace generic preprocessing procedures. Employing pipelines, distinct methods can be compared with each other. In ABIDE datasets, preprocessing is performed by four pipeline techniques: neuroimaging analysis kit (NIAK) [55], data processing assistant for rs-fMRI (DPARSF) [56], the configurable pipeline for the analysis of connectomes (CPAC) [57], or connectome computation system (CCS) [58]. The preprocessing steps carried out by the various pipelines are comparatively analogous. The chief differences are in the particular algorithms for each step, the software simulations, and the parameters applied. Details of each pipeline technique are provided in [54]. Table I
demonstrates the pipeline techniques used in autism detection exploiting DL.

3) **High-level preprocessing Steps**: High-level techniques for pre-processing brain data are important, and using them accompanying preliminary pre-processing methods can enhance the accuracy of ASD recognition. These methods are applied after the standard pre-processing of functional and structural brain data. These include sliding window (SW) [24], data augmentation (DA) [59], functional connectivity matrix (FCM) estimation [60], [61] and applying fast Fourier transformation (FFT) [62]. Furthermore, some of the researches utilized feature extraction [63] techniques and some also use feature selection methods. Precise information on reviewed studies is indicated in detail in Table I.

### C. Deep Neural Networks

Deep learning in various medical applications, including the diagnosis of ASD, has become extremely popular in recent years. In this section of the paper, the types of Deep Learning networks used in ASD detection are examined, which include CNN, RNN, AE, DBN, CNN-RNN, and CNN-AE models.

1) **Convolutional Neural Networks (CNNs)**: In this section, the types of popular convolutional networks used in ASD diagnosis are surveyed. These networks involve 1D-CNN, 2D-CNN, 3D-CNN models, and a variety of pre-trained networks such as VGG.

1D AND 2D-CNN

There are many spatial dependencies present in the data and it is difficult to extract these hidden signatures from the data. Convolution network uses a structure alike to convolution filters to extract these features properly and contribute to the knowledge that features should be processed taking into account spatial dependencies; so the number of network parameters are significantly reduced. The principal application of these networks is in image processing and due to the two-dimensional (2D) image inputs, convolution layers form 2D structures, which is why these networks are called 2D convolutional neural network (2D-CNN). By using another type of data, one-dimensional signals, the convolution layers’ structure also resembles the data structure [64]. In convolution networks, assuming that various data sections do not require learning different filters, the number of parameters are markedly lessened and make it feasible to train these networks with smaller databases [21]. Figure 4 shows the block diagram of 2D-CNN used for ASD detection.

3D-CNN

By transforming the data into three dimensions, the convolution network will also be altered to a three-dimensional format (Figure 5). It should be noted that the manipulation of three dimensional CNN (3D-CNN) networks is less beneficial than 1D-CNN and 2D-CNN networks for diverse reasons. First, the data required to train these networks must be much larger which conventionally such datasets are not utilizable and methods such as pre-training, which are extensively exploited in 2D networks, cannot be used here. Another reason is that with more complicated structure of networks, it becomes much tougher to fix the number of layers, and network structure. The 3D activation map generated during the convolution of a 3D CNN is essential for analyzing data where volumetric or temporal context is crucial. This ability to analyze a series of frames or images in context has led to the use of 3D CNNs as tools for action detection and evaluation of medical imaging [65].

2) **Deep Belief Networks (DBNs)**: DBNs are not popular today as they used to be, and have been substituted by new models to perform various applications (e.g., autoencoders for unsupervised learning, generative adversarial networks (GAN) for generative modes [66], variational autoencoders (VAE) [67]). However, disregarding the restricted use of these networks in this era, their influence on the advancement of neural
networks cannot be overlooked. The use of these networks in this paper is related to the feature extraction without a supervisor or pre-training of networks. These networks serve as unsupervised, consisting of several layers after the input layer, which are shown in Figure 6. The training of these networks is done greedily and from bottom to top, in other words, each separate layer is trained and then the next layer is appended. After training, these networks are used as a feature extractor or the network weights are used as initial weights of a network for classification [21].

3) Autoencoders (AEs): Autoencoders (AEs) are more than 30 years old, and have undergone dramatic changes over the years to enhance their performance. But the overall structure of these networks has remained the same [21]. These networks consist of two parts: coder and decoder so that the first part of the input leads to coding in the latent space, and the decoder part endeavors to convert the code into preliminary data (Figure 7). Autoencoders are a special type of feedforward neural networks where the input is the same as the output. They compress the input into a lower-dimensional code and then reconstruct the output from this representation. The code is a compact “summary” or “compression” of the input, also called the latent-space representation. Various methods have been proposed to block the data memorization by the network, including sparse AE (SpAE) and denoising AE (DAE) [21]. Trained properly, the coder part of an Autoencoder can be used to extract features; creating an unsupervised feature extractor.

4) Recurrent Neural Networks (RNNs): In convolution networks, a kind of spatial dependencies in the data is addressed. But interdependencies between data are not confined to this model. For example in time-series, dependencies may be highly distant from each other, on the other hand, the long-term and variable length of these sequences results in that the ordinary networks do not perform well enough to process these data. To overcome these problems, RNNs can be used. LSTM structures are proposed to extract long term and short term dependencies in the data (Figure 8). Another well-known structure called GRU is developed after LSTM, and since then, most efforts have been made to enhance these two structures and make them resistant to challenges (e.g., GRU-D [68] is used to find the lost data).

5) CNN-RNN: The initial idea in these networks is to utilize convolution layers to amend the performance of RNNs so that the advantages of both networks can be used; CNN-RNN, on the one hand, can find temporal dependencies with the aid of RNN, and on the other hand, it can discover spatial dependencies in data with the help of convolution layers [69]. These networks are highly beneficial for analyzing time series with more than one dimension (such as video) [70] but further to the simpler matter, these networks also yield the analysis of three-dimensional data so that instead of a more complex design of a 3D-CNN, a 2D-CNN with an RNN is occasionally used. The superiority of this model is due to the feasibility of employing pre-trained models. Figure 9 demonstrates the CNN-RNN model.

6) CNN-AE: In the construction of these networks, the principal aim and prerequisite have been to decrease the number of parameters. As shown before, changing merely the network layers to convolution markedly lessens the number of parameters; combining AE with convolution structures also makes significant contribution. This helps to exploit higher dimensional data and extracts more information from the data without changing the size of the database. Similar structures, with or without some modifications, are widely deployed for image segmentation [71], and likewise unsupervised network can be applied for network pre-training or feature extraction. Figure 10 depicts the CNN-AE network used for ASD detection. Tables I and II provide the summary of papers published on detection and rehabilitation of ASD patients using DL,
V. DEEP LEARNING TECHNIQUES FOR ASD REHABILITATION

Rehabilitation tools are employed in multiple fields of medicine and their main purpose is to help the patients to recover after the treatment. Various and multiple rehabilitation tools using DL algorithms have been presented. Rehabilitation tools are used to help ASD patients using mobile, computer applications, robotic devices, cloud systems, and eye tracking, which will be discussed below. Also, the summary of papers published on rehabilitation of ASD patients using DL algorithms are shown in Table II.

A. Mobile and Software Applications

Facial expressions are a key mode of non-verbal communication in children with ASD and play a pivotal role in social interactions. Use of BCI systems provides insight into the user’s inner-emotional state. Valles et al. [72] conducted research focused on mobile software design to provide assistance to children with ASD. They aimed to design a smart iOS app based on facial images according to Figure 11. In this way, people’s faces at different angles and brightness are first photographed, and are turned into various emoji so that the autistic child can express his/her feelings and emotions. In this group’s investigation [72], Kaggle’s (The Facial Expression Recognition 2013) and KDEF (Kaggle’s FER2013 and Karolinska Directed Emotional Faces) databases were used to train the VGG-16. In addition, the LEAP system was adapted to train the model at the University of Texas. The research provided the highest rate accuracy of 86.44%. In another similar study, they achieved an accuracy of 78.32% [73].

B. Cloud Systems

Mohammadian et al. [74] proposed a new application of DL to facilitate automatic stereotypical motor movement (SMM) identification by applying multi-axis inertial measurement units (IMUs). They applied CNN to transform multi-sensor time series into feature space. An LSTM network was then combined with CNN to obtain the temporal patterns for SMM identification. Finally, they employed the classifier selection voting approach to combine an ensemble of the best base learners. After various experiments, the superiority of their proposed procedure over other base methods was proven. Figure 12 shows the real-time SMM detection system. First, IMUs, which are wearable sensors, are used for data collection; the data can then be analyzed locally or remotely (using Wi-Fi to transfer data to tablets, cell phones, medical center servers, etc.) to identify SMMs. If abnormal movements are detected, an alarm will be sent to a therapist or parents.

C. Eye Tracking

Wu et al. [75] proposed a model of DL saliency prediction for autistic children. They used DCN in their proposed paradigm, with a SM saliency map output. The fixation density map (FDM) was then processed by the single-side clipping (SSC) to optimize the proposed loss function as a true label along with the SM saliency map. Finally, they exploited an autism eye-tracking dataset to test the model. Their proposed model outperformed other base methods. Elbattah et al. [76] aimed to combine unsupervised clustering algorithms with deep learning to help ASD rehabilitation. The first step involved the visualization of the eye-tracking path, and the images captured from this step were fed to an autoencoder to learn the features. Using autoencoder features, clustering models are developed using the K-Means algorithm. Their method performed better than other state-of-art techniques.
| Work | Datasets | Neuroimaging Modalities | Number of Cases | Pipelines | Image Atlas | Preprocessing Toolbox | High level Preprocessing | Inputs DNN | DNN Toolbox | DNN | Number of Layers | Classifier | K fold | Performance Criteria (%) |
|------|----------|------------------------|-----------------|-----------|-------------|-----------------------|--------------------------|------------|-------------|-----|----------------|------------|--------|-------------------------|
| ABIDE I | rs-fMRI | T-fMRI | 15 ASD 22 HC | NA | AAL | DNN | Single Mean Heatmap Input | NA | 2CUCD | 17 | Majority Voting | NA | F1-Score = 89 |
| ABIDE I | rs-fMRI | T-fMRI | 15 ASD 22 HC | NA | AAL | DNN | Single Mean Heatmap Input | NA | 2CUCD | 14 | Sidgmore | NA | Acc = 97.32 |
| HCP Dataset in the H4HNI Project | T-fMRI | T-fMRI | 28 Subjects with 7 T1 and 2-6 T2-FlAIR Data | NA | AAL | DNN | Dictionary Learning and Sparse Coding | Functional RNs Maps | NA | 3D-CNN | 8 | Softmax | NA | Acc = 94.61 |
| ABIDE I | rs-fMRI | T-fMRI | 21 ASD 14 HC | NA | AAL | DNN | ROIs Time-Series | Keras LSTM | 7 | Sidgmore | 10 | Acc = 90.8 |
| ABIDE I | rs-fMRI | T-fMRI | 1711 ASD 15003 HC | NA | AAL | DNN | Word2: Transform and Different Techniques | FC51 | Keras | 14 | Softmax | NA | Ensemble AUC=0.92 | Ensemble Acc=85.19 |
| ABIDE I | rs-fMRI | T-fMRI | 81 ASD 45 HC | NA | AAL | DNN | Multi-NN Sequence | NA | 2CUCD | 16 | Sidgmore | NA | Acc = 87.1 |
| ABIDE I | rs-fMRI | T-fMRI | 61 ASD 35 HC | NA | AAL | DNN | Multi-NN Sequence | NA | 2CUCD | 16 | Sidgmore | NA | Acc = 85.3 |
| ABIDE I | rs-fMRI | T-fMRI | 521 ASD 374 HC | CPAC | All All ies | DNN | FCM | NA | FCM | 5 | Softmax | 10 | Acc = 73.3 |
| ABIDE I | rs-fMRI | T-fMRI | 505 ASD 530 HC | CPAC | All All ies | DNN | FCM | NA | FCM | 5 | Softmax | 10 | Acc = 90.8 |
| ABIDE I | rs-fMRI | T-fMRI | 372 subjects | NA | AAL | DNN | Raw Images | FCM | NA | FCM | 5 | Softmax | 10 | Acc = 90.8 |
| ABIDE I | rs-fMRI | T-fMRI | 13 ASD 22 HC | NA | AAL | DNN | Quant. NMR Statics Matrix | NA | DAE | NA | NA | NA | Acc = 54.49 |
| ABIDE I | rs-fMRI | T-fMRI | 11 ASD 16 HC | NA | NA | DNN | Preprocessed | NA | LeNet-5 | Standard | Softmax | NA | Acc = 89.99 | Spee = 100 |
| ABIDE I | rs-fMRI | T-fMRI | 54 ASD 55 HC | NCAK | AAL | DNN | Preprocess Selection | Whole-Brain FCs | NA | Multiple SAEs | 4 | Softmax regression | 5 | Acc = 86.56 |
| ABIDE II | rs-fMRI | T-fMRI | 54 ASD 62 HC | NA | AAL | DNN | Dimension Reduction | Images with 96 x 68, 79 x 69, and 70 x 70 Dimensions, Around the x, y, and z Axes | MCNN | 9 | Binary SR | 10 | Acc = 72.73 | Spee = 71.1 |
| ABIDE I + II | | | | | | | | | | | | | | |
| ABIDE I | rs-fMRI | Phenotypic Info | 542 ASD 625 HC | CPAC | All All ies | DNN | Creating Stochastic Parcels from 500000 brain seed points | NA | 3D-CNN | 6 | Various Methods | 10 | Acc = 72 |
| ABIDE I | rs-fMRI | T-fMRI | 485 ASD 507 HC | DPARSF | AAL | DNN | FCM | Edge Weights of Subject's Brain Graph | Keras | VAE | 3 | NA | NA | NA |
| ABIDE I | rs-fMRI | T-fMRI | 370 ASD 372 HC | CPAC | AAL | DNN | Mean Time Courses from ROIs | Keras LSTM | 5 | Sidgmore | 10 | Acc = 68.5 |
| ABIDE I | rs-fMRI | Phenotypic Info | 505 ASD 530 HC | CC200 | CPAC | DNN | Dimensional Spatial Standardization, Smoothing, Filtering, Removing Covariances, FCM, All-MKFC | 4003-Dimensional Eigenvector | NA | SAE | 3 | Clustering | NA | Acc = 86.61 | F1-Score = 69.2 |
| ABIDE I | rs-fMRI | T-fMRI | 42 ASD 42 HC | NA | AAL | DNN | Independent Components | Time Course of Each Subject | NA | SAE | 9 | Softmax | 21 | Acc = 87.21 | Spee = 89.49 | Spee = 85.75 |
| ABIDE I | rs-fMRI | T-fMRI | 146 ASD 146 HC | NA | AAL | DNN | iMRI ROI Time-Series, Functional Connectivity | Keras LSTM | 6 | Sidgmore | 10 | Acc = 74.8 |
| NY site | | | | | | | | | | | | | | |
| NY site | | | | | | | | | | | | | | |
| ABIDE I | rs-fMRI | T-fMRI | 408 ASD 401 HC | CPAC | AAL | DNN | 3 Different FCs Demographic Data | Keras DANN | 25 | Softmax | 10 | Acc = 75.2 | Spee = 74.5 | Spee = 71.7 |
| ABIDE I | rs-fMRI | T-fMRI | 46 ASD 46 HC | CPAC | AAL | DNN | 3 Different FCs Demographic Data | Keras DANN | 25 | Softmax | 10 | Acc = 75.2 | Spee = 74.5 | Spee = 71.7 |
| 11 ASD | | | | | | | | | | | | | | |
| 12 ASD | | | | | | | | | | | | | | |
| ABIDE I | rs-fMRI | 521 ASD 374 HC | NA | AAL | DNN | Mean Time-Series within each ROI | NA | 1D-CNN | 5 | Softmax | 10 | Acc = 90.8 |

TABLE I: Summary of articles published using DL methods for neuroimaging-based ASD detection.
Deep Glass Brain and Pipelines
Stat Map Images
573 HC
Classifier Image Generator

Upper Triangle Part of the Correlation Matrix

Auto-ASD-Network Proposed SVM 5

Upper Triangle Part of

Segmentation

DiagNet

Mean MHD=14.05

Spec=94.56

Average Acc= 70.5

Average Sen= 74

Mean Spec= 67

530 ASD, 530 HC

Sen=74

530 HC

Spec=63

573 HC

Mean MHD=14.05

FSL

Average Spec= 67

539 ASD, 539 HC

Sen=74

571 HC

Spec=63

449 HC

Sen=74

529 ASD

497 HC

Acc=96.88
| Dataset      | Type | Methodology | Preprocessing | Feature Extraction | Classification | Evaluation Metric     | Experiment Details |
|--------------|------|-------------|---------------|--------------------|----------------|-----------------------|-------------------|
| ABIDE-I      | rs-fMRI | 620 ASD   | CPAC          | FSL                | Geometric DA   | 5                     | F1-Score=64       |
|              |      | 2055 HC    |               |                    |                |                       |                   |
|              |      |            |               |                    |                |                       |                   |
| ABIDE-I      | rs-fMRI | 144 ASD  | CPAC          | FSL                | Geometric DA   | 5                     | F1-Score=79.2     |
|              |      | 110 HC     |               |                    |                |                       |                   |
|              |      |            |               |                    |                |                       |                   |
| ABIDE-I      | rs-fMRI | 405 ASD  | CPAC          | FSL                | Geometric DA   | 10                    | F1-score=78.35    |
|              |      | 466 HC     |               |                    |                |                       |                   |
|              |      |            |               |                    |                |                       |                   |
| ABIDE-I      | rs-fMRI | 505 ASD  | CPAC          | FSL                | Geometric DA   | 10                    | F1-Score=71.0     |
|              |      | 530 HC     |               |                    |                |                       |                   |
|              |      |            |               |                    |                |                       |                   |
| ABIDE-I      | rs-fMRI | 270 ASD  | CPAC          | BrainNetome Atlas  | Geometric DA   | 10                    | F1-Score=64       |
|              |      | 305 HC     |               |                    |                |                       |                   |
|              |      |            |               |                    |                |                       |                   |
| ABIDE-I      | rs-fMRI | 25 ASD   | CPAC          | BrainNetome Atlas  | Geometric DA   | 10                    | F1-Score=64       |
|              |      | 22 HC      |               |                    |                |                       |                   |
|              |      |            |               |                    |                |                       |                   |
| Different   |      |            |               |                    |                |                       |                   |
| Datasets    |      |            |               |                    |                |                       |                   |
| Work | Datasets | Type of Applications | Number of Cases | Preprocessing | Inputs | DNN | DNN Toolbox | DNNs | Number of Layers | Classifier | K fold | Performance Criteria (%) |
|------|----------|---------------------|----------------|--------------|--------|-----|-------------|------|----------------|------------|--------|--------------------------|
| [13] | OSIE     | —                   | 20 ASD 19 HC   | HFMs, Neural Scene Images | Caffe  | VGGNet | 50          | Softmax | 13           |            | 85            | Spec80, Spec89         |
| [14] | KDEF     | Facial Expression Recognition | 70 Individuals | DA | RGB Images (562x362) | Keras | DCNN | 44          | Softmax | NA            | Acc78.32     |
| [15] | Clinical Acquisition System | Detecting Audio Regimes That Directly Estimate ASD Social Affect scores | 33 ASD | MFCC Spectrograms | 32 Spectrograms | NA | Keras | 44          | Softmax | NA            | Acc94.7       |
| [16] | Kaggle FER2013 | Facial Expression Recognition | NA | No | 48x48-Pixel Images | Keras | TensorFlow | 44          | Softmax | NA            | Acc96.44      |
| [17] | SALICON  | ASD Classification  | 14 ASD 14 HC   | SalGAN Model, Feature Extraction | Sequence of Image Patches | NA | SP-ASDNet | 11          | NA            | NA          | Acc57.00 | Rec50.21, Pre56.26 |
| [18] | BigFauxX | Facial Expression Recognition | 196 Subjects | SW, Merge in the Chained Dimensions, DA | 5-channel Sub-Sequence Stacks within a Specific Time Window | Keras | TimeConvNet | PreTrain Net | Softmax | NA            | Acc97.9        |
| [19] | Different Datasets | Suitable Courseware for Children with ASD | NA | Different Inputs | NA | Different Nets | NA          | NA            | NA          | NA            | NA          |
| [20] | Camera Images | Estimating Visual Attention in Robot-Assisted Therapy | 6 ASD and ID | Face Detection (Video-Source), Feature Extraction (HOG Descriptors) | 5 Facial Landmarks - 36 HOG Descriptors | NA | R-CNN | VGG-16 | K-NN | 10            | Acc88.2 | Pre80.33, Spec89.0, Spec97.3 |
| [21] | Sensor Data | Automatic SMM detection | 6 ASD 5 HC | Resampling, Filtering, SW | Time-Series of Multiple Sensors | Keras | CNN-LSTM | 13          | Majority Voting | NA          | NA            | NA          |
| [22] | KOMMA   | Facial Expression Recognition | 55 subjects | Segmentation, Different Features, Z-scores | Greatly Forward Feature Selection | NA | CNN | 9          | SVM | NA            | Acc96       |
| [23] | Story-Telling Narrative Corpus | ASD Classification | 51 ASD 36 HC | DA, Chinese/WordVec | 32-Dimensional Word Vector | NA | LSTM | 1          | Coherence Representation of LSTM Forget Gate | NA          | Acc92       |
| [24] | Eo-Dataset (video dataset) | ASD Classification using Eye Tracking | 136 ASD 138 HC | TLD Method, Accumulative Histogram Computation | Angle Histogram, Length Histogram and Fused Histogram, Keras | LSTM | 4          | NA          | Acc95.6 | Spec95.4  |
| [25] | MFT1003 | Predicting Visual Attention of Children with ASD | 300 Images | NA | Raw Images | NA | DCNN | 26          | NA            | NA          | Acc97.6 | Spec97.4 |
| [26] | Scan Path Data | Including Location and Duration | 14 ASD 14 HC | DA Methods | Image, Data Points | Pytorch | ResNet18 | Standard | Softmax | NA            | Acc90.13  | Spec97.1 |
| [27] | UCI Machine Learning Repository | ASD Classification | 704 | Different Methods | Preprocessed Data | NA | CNN | 7          | NA            | NA          | Acc90.53 | Spec97.90, Spec100  |
| [28] | Eye Tracking Scaphai | ASD Classification | 20 ASD 30 HC | Visualization of Eye-Tracking Scaphai Scaling Down, PCA | 100x100 Image | Keras, Sickle-learn | AE | 8          | K-Means Clustering | NA | Silhouette score60 |
| [29] | Video Data | Engagement Estimation of Children with ASD During a Robot-Assisted Autism Therapy | 30 children | Capped Face Images (256 *256) | Keras | TensorFlow Backend | CultureNet | R-CNN + ResNet50 + 512 FC layers | Softmax | NA | ICE54.35, Spec53.11, Spec81.17, Spec11.17 |
| [30] | YouTube ASD Dataset | Modeling Typical and Atypical Behaviors in ASD Children | 68 video Clips | Different Methods | Sequences of Individual Frames at a Rate of 30 fps | Keras, Caffe | DCNN | NA          | DT | 5          | Acc97.03 | Arg Recall79, Arg Acc71.51 |
| [31] | Video Dataset | Behavioral Data Extracted from Video Analysis of Child-Robot Interactions | 5 ASD 7 HC | Segmentation, Upper Body tracking, Lahub Movement Analysis to Drive Weight, Different features | 3 Movement Features with 68 Facial Key-Points | NA | CNN | 10          | Softmax | NA            | Acc95.46 | Pre80.12, Recall80.53 |
| [32] | Video Dataset | Developing Automatic SMM Detection Systems | 6 ASD | Resampling, Filtering, SW, Data Balancing, Normalizing | Time-Series of Multiple Accelerometer Sensors | Deeppy Library | CNN | 8          | SVM | NA            | F1-score95 |
| [33] | ASD Screening | Autism Screening | 515 ASD 188 HC | Cleaning Missing Values and Outliers, Visualization, Identity Mapping | The Embedded Categorical Variables are Concatenated with Numerical Features as New Feature Vectors | NA | DENN | 4          | Sigmoid | NA            | Acc90.00 | Spec97.40, Spec100  |
| [34] | ASD Screening Datasets | Classification of Adults with ASD | — | Handling of Missing Values, Variable Reduction, Normalization, and Label Encoding | Normalized Variables | Keras | DNN | 7          | Sigmoid | NA            | Acc99.00 | Spec97.49, Spec100  |
VI. DISCUSSION

In this study, we performed a comprehensive overview of the investigations conducted in the scope of ASD diagnostic CAD systems as well as DL-based rehabilitation tools for ASD patients. In the field of ASD diagnosis, numerous papers have been published using functional and structural neuroimaging data as well as rehabilitation tools, as summarized in Table III in the appendix. A variety of DL toolboxes have been proposed for implementing deep networks. In Tables I and II the types of DL toolboxes utilized for each study are depicted, and the total number of their usage is demonstrated in Figure 13. The Keras toolbox is used in the majority of the studies due to its simplicity. Keras offers a consistent high-level application programming interface (APIs) to build the models more straightforward, and by using powerful backends such as TensorFlow, its performance is sound. Additionally, due to all pre-trained models and available codes on platforms such as GitHub, Keras is quite popular among researchers.

Number of DL networks used for the ASD detection in the reviewed works is shown in Figure 14. Among the various DL architectures, CNN is found to be the most popular one as it has achieved more promising results compared to other deep methodologies. The autoencoder, as well as RNN, have yielded favorable results. It can be noted that in recent years, the number of DL-based papers has increased exponentially due to their sound performance and also the availability of vast and thorough datasets.

The number of various classification algorithms used in DL networks are shown in Figure 15. One of the best and most widely used is the Softmax algorithm (Tables I and II). It is most popular since it is differentiable in the entire domain and computationally less expensive.

VII. CHALLENGES

Some of the most substantial challenges in ASD diagnosis scope using DL-based techniques are addressed in this section, which comprise database and algorithmic problems. There are only two-class brain structural and functional datasets (ASD and healthy) available in the public domain. Hence, researchers are not able to broaden their investigation to all sub-types of ASD. Two of the cheapest and most pragmatic functional neuro-screening modalities for diagnosis of ASD are EEG, and fNIRS. But unfortunately the deficiency of freely available datasets has resulted in little research in this area. Another obstacle is that multi-modality databases such as EEG-fMRI are not available to researchers to evaluate the effectiveness of incorporating information of different imaging modalities to detect ASD. Although fMRI and sMRI data are ubiquitous in the ABIDE dataset, the results of merging these structural and functional data for ASD diagnosis using DL have not yet been investigated. Another problem grappling the researchers is designing the DL-based rehabilitation systems with hardware resources. Nowadays, assistive tools such as Google Colab are available to researchers to improve the processing power; however, the problems still prevail when implementing these systems in real-world scenarios.
VIII. Conclusion and Future Works

ASD is typically characterized by social disorders, communication deficits, and stereotypical behaviors. Numerous computer-aided diagnosis systems and rehabilitation tools have been developed to assist patients with autism disorders. In this survey, research on ASD diagnosis applying DL and functional and structural neuroimaging data were first assessed. The researchers have taken advantage of deep CNNs, RNNs, AEs, and CNN-RNN networks to improve the performance of their systems. Boosting the accuracy of the system, the capability of generalizing and adapting to differing data and real-world challenges, as well as reducing the hardware power requirements to the extent that the final system can be utilized by all are the principal challenges of these systems. To enhance the accuracy and performance of CADs for ASD detection in the future, deep reinforcement networks (RL) or GANs can be exploited. Scarcity of data is always an aparamount problem in the medical field that can be resolved relatively with the help of these deep GANs. Also, as another direction for future works, handcrafted features can be extracted from data and fed to DL networks in addition to raw data; this can help increasing performance by adding the potential of traditional methods to DL-based models.

Many researchers have proposed various DL-based rehabilitation tools to aid the ASD patients. Designing a reliable, accurate, and wearable low power consumption DL algorithm based device is the future tool for ASD patients. An achievable rehabilitation tool is to wear smart glasses to help the children with ASD. These glasses with the built-in cameras will acquire the images from the different directions of environment. Then the DL algorithm processes these images and produces meaningful images for the ASD children to better communicate with their surroundings.

APPENDIX A
STATISTICAL METRICS

This section demonstrates the equations for the calculation of each evaluation metric. In these equations, True positive (TP) is the correct classification of the positive class, True negative (TN) is the correct classification of the negative class, False positive (FP) is the incorrect prediction of the positives, False negative (FN) is the incorrect prediction of the negatives.

\[
\text{Accuracy (Acc)} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}
\]

\[
\text{Specificity (Spec)} = \frac{TN}{TN + FP} \tag{2}
\]

\[
\text{Sensitivity (Sen)} = \frac{TP}{TP + FN} \tag{3}
\]

\[
\text{Precision (Prec)} = \frac{TP}{TP + FP} \tag{4}
\]

\[
F1 - \text{Score} = 2 \times \frac{\text{Prec} \times \text{Sens}}{\text{Prec} + \text{Sens}} \tag{5}
\]

The receiver operating characteristic curve (ROC-curve) depicts the performance of the proposed model at all classification thresholds. It is the graph of true positive rate vs. false positive rate (TPR vs. FPR). Equations for calculation of TPR and FPR are presented below.

\[
\text{TPR} = \frac{TP}{TP + FN} \tag{6}
\]

\[
\text{FPR} = \frac{FP}{FP + TN} \tag{7}
\]

AREA UNDER THE ROC CURVE (AUC)

AUC presents the area under the ROC-curve from (0, 0) to (1, 1). It provides the aggregate measure of all possible classification thresholds. AUC has a range from 0 to 1. A 100% wrong classification will have AUC value of 0.0, while a 100% correct classified version will have the AUC value of 1.0. It has two folded advantages. One is that it is scale-invariant, which implies how well the model is predicted rather than checking the absolute values. The second advantage is that it is classification threshold-invariant as it will verify the performance of the model irrespective of the threshold being selected.

APPENDIX B

Table III shows details of Deep Nets in all the papers reviewed in this study.

ACKNOWLEDGMENT

MB is supported by a NHMRC Senior Principal Research Fellowship (1059660 and 1156072). MB has received Grant/Research Support from the NIH, Cooperative Research Centre, Simons Autism Foundation, Cancer Council of Victoria, Stanley Medical Research Foundation, Medical Benefits Fund, National Health and Medical Research Council, Medical Research Futures Fund, Beyond Blue, Rotary Health, A2 milk company, Meat and Livestock Board, Woolworths, Avant and the Harry Windsor Foundation, has been a speaker for Astra Zeneca, Lundbeck, Merck, Pfizer, and served as a consultant to Allergan, Astra Zeneca, Biodyntex, Bionomics, Collaborative Medicinal Development, Lundbeck Merck, Pfizer and Servier - all unrelated to this work.
| Work | Network | Details for Deep Networks | Dropout | Classifier | Optimizer | Loss Function |
|------|---------|--------------------------|---------|------------|-----------|---------------|
| 1 | 2CNN | CNN Layers (5) + Pooling Layers (4) + FC Layers (1) | 0.1 | Softmax | NA | BCE |
| 2 | 2CNN | CNN Layers (5) + Pooling Layers (4) + FC Layers (1) | 0.1 | Softmax | NA | BCE |
| 3 | 3D-CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | NA | BCE |
| 4 | LSTM | LSTM Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | NA | BCE |
| 5 | SGD | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | NA | BCE |
| 6 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | NA | BCE |
| 7 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | NA | BCE |
| 8 | AE | Standard AE with Tanh Activation | NA | SLP | NA | BCE |
| 9 | CNN | Standard CNN with 3 Layer CNN | 0.1 | Softmax | Adam | NA |
| 10 | MCN | CNN Layers (3) + ReLU Activation + Pooling Layers (3) + FC Layers (1) | 0.1 | Softmax | Adam | BCE |
| 11 | SDN | CNN Layers (2) + ELU Activation + Pooling Layers (2) + FC Layers (1) | 0.1 | Softmax | Adam | BCE |
| 12 | ACM | CNN Layers (2) + ELU Activation + Pooling Layers (1) + FC Layers (1) | 0.1 | Softmax | Adam | BCE |
| 13 | LSTM | LSTM Layers (2) + Pooling Layers (2) + FC Layers (1) | 0.1 | Softmax | Adam | BCE |
| 14 | Multichannel DASW | 5 MLP (1) [spatial layer and 4 feature layers] + Self-Attention (3) + Fusion (3) + Aggregate Layer + Dense Layer (3) | 0.1 | Softmax | NA | CE |
| 15 | SAE | 5 SAE Layers | 0.1 | Softmax | Adam | NA |
| 16 | 3D-CNN | CNN Layers (1) + Pooling Layers (1) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 17 | CNN | CNN Layers (5) + Pooling Layers (4) + BN Layers (2) + FC Layers (2) | 0.1 | Softmax | Adam | NA |
| 18 | ID-CAD-CNN | Encoder (4 layers) + Decoder (4 layers) + CNN Layers (2) + pooling layers (2) + FC Layers (2) | 0.1 | Softmax | Adam | NA |
| 19 | CNN | Standard ENet Architecture | 0.1 | Softmax | Adam | NA |
| 20 | ADMM | Proposed ADMM | 0.1 | Softmax | Adam | NA |
| 21 | MCN | CNN Layers (1) + Pooling Layers (1) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 22 | SAE | CNN Layers (3) + Pooling Layers (3) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 23 | CNN | CNN Layers (3) + Pooling Layers (3) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 24 | SS-AE | CNN Layers (3) + Pooling Layers (3) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 25 | LSTM | LSTM Layers (2) + Pooling Layers (2) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 26 | Ensemble of 5 SAEs and MLPs | 5 [AE (3) + MLP (2) + Softmax] | 0.1 | Softmax | Adam | NA |
| 27 | Multi-Skilled CNN | CNN Layers (3) + ReLU Activation + Pooling Layers (2) + FC Layers (3) | 0.1 | Softmax | Adam | NA |
| 28 | MSA | MSA | 0.1 | Softmax | Adam | NA |
| 29 | DDCN | Proposed DDCN with 10 Blocks and ReLU Activation | 0.1 | Softmax | Adam | NA |
| 30 | 3D-CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (5) | 0.1 | Softmax | Adam | NA |
| 31 | Transformer-U | Proposed Transformer-U | 0.1 | Softmax | Adam | NA |
| 32 | 3D-CNN | CNN Layers (1) + Pooling Layers (1) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 33 | CNN-GRU | CNN Layers (4) + GRU Layers (2) + ReLU Activation + Pooling Layers (2) + FC Layers (2) | 0.1 | Softmax | Adam | NA |
| 34 | CNN-LSTM | CNN Layers (3) + LSTM Layers (1) | 0.1 | Softmax | Adam | NA |
| 35 | CNN-UNet | CNN Layers (5) + ReLU Activation + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 36 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 37 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 38 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 39 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 40 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 41 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 42 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 43 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 44 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 45 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 46 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 47 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 48 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 49 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 50 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 51 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 52 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 53 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 54 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 55 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 56 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 57 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 58 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 59 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
| 60 | CNN | CNN Layers (5) + Pooling Layers (5) + FC Layers (1) | 0.1 | Softmax | Adam | NA |
C. Wang, Z. Xiao, B. Wang, and J. Wu, “Identification of autism based
on svm-rfe and stacked sparse auto-encoder,” Neural Computating,
vol. 77, no. 17, pp. 22809–22820, 2018.

M. N. Dvornek, X. Li, J. Zhang, and J. S. Duncan, “Jointly discrimina-
tive and generative recurrent neural networks for learning from fmri,”
in International Workshop on Machine Learning in Medical Imaging.
Springer, 2019, pp. 382–390.

K. Niu, J. Guo, Y. Pan, X. Gao, X. Peng, N. Li, and H. Li, “Multichannel
deeplinear networksdneural networks for the classification of autism spec-
trum disorder using neuroimaging and personal characteristic data,”
Complexity, vol. 2020, 2020.

H. Li, N. A. Parikh, and L. He, “A novel transfer learning approach
to enhance deep neural network classification of brain functional
connectomes,” Frontiers in neuroscience, vol. 12, p. 491, 2018.

A. El Gazzar, L. Cerlani, G. van Wingen, and R. M. Thomas, “Simple
r-convolutional networks for resting-state fmri based classification in
autism,” in 2019 International Joint Conference on Neural Networks
(IJCNN). IEEE, 2019, pp. 1–6.

M. R. Ahmed, Y. Zhang, Y. Liu, and H. Liao, “Single volume image
generator and deep learning-based classification,” IEEE Journal of
Biomedical and Health Informatics, 2020.

Y. Zhao, H. Dai, W. Zhang, F. Ge, and T. Liu, “Two-stage spatial tem-
poral deep learning framework for functional brain network modeling,
in 2019 IEEE 16th International Symposium on Biomedical Imaging
(ISBI 2019). IEEE, 2019, pp. 1576–1580.

B. Pugazhenthi, G. Senapathy, and M. Pavithra, “Identification of
autism in mri brain images using deep learning networks,” in 2019
International Conference on Smart Structures and Systems (ICSSS).
IEEE, 2019, pp. 1–7.

T. Eslami, J. S. Raiker, and F. Saeed, “Explainable and scalable machine-learning algorithms for detection of autism spectrum disorder using fmri data,” arXiv preprint arXiv:2003.01541, 2020.

K. Sairam, J. Naren, G. Vithya, and S. Srivatsan, “Computer aided system for autism spectrum disorder using deep learning methods,” International Journal of Psychosocial Rehabilitation, vol. 23, no. 01, 2019.

J. Dolz, C. Desrosiers, and I. B. Ayed, “3d fully convolutional networks for subcortical segmentation in mri: A large-scale study,” Neuroimage, vol. 170, pp. 456–470, 2018.

C. Wang, Z. Xiao, B. Wang, and J. Wu, “Identification of autism based
on svm-rfe and stacked sparse auto-encoder,” IEEE Access, vol. 7,
p. 118063, 2019.

Y. Zhao, F. Ge, S. Zhang, and T. Liu, “3d deep convolutional neural network revealed the value of brain network overlap in differentiating autism spectrum disorder from healthy controls,” in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2018, pp. 172–180.

Y. Zhao, F. Ge, S. Zhang, and T. Liu, “3d deep convolutional neural
network revealed the value of brain network overlap in differentiating autism spectrum disorder from healthy controls,” in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2018, pp. 172–180.

Y. Zhao, F. Ge, S. Zhang, and T. Liu, “3d deep convolutional neural
network revealed the value of brain network overlap in differentiating autism spectrum disorder from healthy controls,” in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2018, pp. 172–180.

Y. Zhao, F. Ge, S. Zhang, and T. Liu, “3d deep convolutional neural
network revealed the value of brain network overlap in differentiating autism spectrum disorder from healthy controls,” in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2018, pp. 172–180.

Y. Zhao, F. Ge, S. Zhang, and T. Liu, “3d deep convolutional neural
network revealed the value of brain network overlap in differentiating autism spectrum disorder from healthy controls,” in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2018, pp. 172–180.

Y. Zhao, F. Ge, S. Zhang, and T. Liu, “3d deep convolutional neural
network revealed the value of brain network overlap in differentiating autism spectrum disorder from healthy controls,” in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2018, pp. 172–180.

Y. Zhao, F. Ge, S. Zhang, and T. Liu, “3d deep convolutional neural
network revealed the value of brain network overlap in differentiating autism spectrum disorder from healthy controls,” in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2018, pp. 172–180.

Y. Zhao, F. Ge, S. Zhang, and T. Liu, “3d deep convolutional neural
network revealed the value of brain network overlap in differentiating autism spectrum disorder from healthy controls,” in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2018, pp. 172–180.

Y. Zhao, F. Ge, S. Zhang, and T. Liu, “3d deep convolutional neural
network revealed the value of brain network overlap in differentiating autism spectrum disorder from healthy controls,” in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2018, pp. 172–180.
convolutional neural networks,” *Frontiers in Psychiatry*, vol. 11, p. 440, 2020.

[126] A. El-Gazzar, M. Quaak, L. Cerliani, P. Bloem, G. van Wingen, and R. M. Thomas, “A hybrid 3dcnn and 3dc-istm based model for 4d spatio-temporal fMRI data: An abide autism classification study,” in *OR 2.0 Context-Aware Operating Theaters and Machine Learning in Clinical Neuroimaging*. Springer, 2019, pp. 95–102.

[127] S. Mostafa, W. Yin, and F.-X. Wu, “Autoencoder based methods for diagnosis of autism spectrum disorder,” in *International Conference on Computational Advances in Bio and Medical Sciences*. Springer, 2019, pp. 39–51.

[128] Z. Jiao, H. Li, and Y. Fan, “Improving diagnosis of autism spectrum disorder and disentangling its heterogeneous functional connectivity patterns using capsule networks,” in 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, 2020, pp. 1331–1334.

[129] J. Xie, L. Wang, P. Webster, Y. Yao, J. Sun, S. Wang, and H. Zhou, “A two-stream end-to-end deep learning network for recognizing atypical visual attention in autism spectrum disorder;” *arXiv preprint arXiv:1911.11393*, 2019.

[130] S. Sadiq, M. Castellanos, J. Moffitt, M.-L. Shyu, L. Perry, and D. Messinger, “Deep learning based multimedia data mining for autism spectrum disorder (asd) diagnosis,” in 2019 International Conference on Data Mining Workshops (ICDMW). IEEE, 2019, pp. 847–854.

[131] Y. Tao and M.-L. Shyu, “Sp-ashnet: Cnn-istm based asd classification model using observer scanpaths,” in 2019 IEEE International Conference on Multimedia & Expo Workshops (ICM EW). IEEE, 2019, pp. 641–646.

[132] J. R. H. Lee and A. Wong, “Timeconvnets: A deep time windowed convolution neural network design for real-time video facial expression recognition,” in 2020 17th Conference on Computer and Robot Vision (CRV). IEEE, 2020, pp. 9–16.

[133] A. Vijayan, S. Janmasree, C. Keerthana, and L. B. Syla, “A framework for intelligent learning assistant platform based on cognitive computing for children with autism spectrum disorder,” in 2018 International CET Conference on Control, Communication, and Computing (IC4). IEEE, 2018, pp. 361–365.

[134] A. Di Nuovo, D. Conti, G. Trubia, S. Buono, and S. Di Nuovo, “Deep learning systems for estimating visual attention in robot-assisted therapy of children with autism and intellectual disability;” *Robotics*, vol. 7, no. 2, p. 25, 2018.

[135] N. M. Rad, S. M. Kia, C. Zarbo, T. van Laarhoven, G. Jurman, P. Venuti, E. Marchiori, and C. Furlanello, “Deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders,” *Signal Processing*, vol. 144, pp. 180–191, 2018.

[136] S. Jaiswal, M. F. Valstar, A. Gllott, and D. Daley, “Automatic detection of adhd and asd from expressive behaviour in rgbd data,” in 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017). IEEE, 2017, pp. 762–769.

[137] Y.-S. Liu, C.-P. Chen, S. S.-F. Gau, and C.-C. Lee, “Learning lexical coherence representation using lstm forget gate for children with autism spectrum disorder during story-telling,” in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018, pp. 6029–6033.

[138] J. Li, Y. Zhong, J. Han, G. Ouyang, X. Li, and H. Liu, “Classifying ad and children with lstm based on raw videos;” *Neurocomputing*, 2019.

[139] W. Wei, Z. Liu, L. Huang, A. Nebout, and O. Le Meur, “Saliency prediction via multi-level features and deep supervision for children with autism spectrum disorder,” in 2019 IEEE International Conference on Multimedia & Expo Workshops (ICM EW). IEEE, 2019, pp. 621–624.

[140] S. Raj and S. Masood, “Analysis and detection of autism spectrum disorder using machine learning techniques,” *Procedia Computer Science*, vol. 167, pp. 994–1004, 2020.

[141] O. Radkiewicz, Y. Usunier, J. Lee, J. Hernandez, E. C. Ferrer, B. Schuller, and R. W. Picard, “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, 2018, pp. 339–346.

[142] A. Cook, B. Mandal, D. Berry, and M. Johnson, “Towards automatic screening of typical and atypical behaviors in children with autism,” in 2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA). IEEE, 2019, pp. 504–510.

[143] H. Javed and C. H. Park, “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*, 2020, pp. 275–277.

[144] K. Sun, L. Li, L. Li, N. He, and J. Zhu, “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, 2020, pp. 3387–3391.

[145] H. Wang, L. Li, L. Chi, and Z. Zhao, “Autism screening using deep embedding representation,” in *International Conference on Computational Science*. Springer, 2019, pp. 160–173.

[146] M. F. Misman, A. A. Samah, F. A. Ezudin, H. A. Majid, Z. A. Shah, H. Hashim, and M. F. Harun, “Classification of adults with autism spectrum disorder using deep neural network,” in 2019 1st International Conference on Artificial Intelligence and Data Sciences (AIDAS). IEEE, 2019, pp. 29–34.