A REVIEW ON AUTISM SPECTRUM DISORDER DIAGNOSIS USING TASK-BASED FUNCTIONAL MRI

Reem Haweel*
Computer Systems, Faculty of Computers and Information Sciences, Ain Shams University,
Cairo, Egypt
reem.t.haweel@cis.asu.edu.eg

Noha Seada
Computer Systems, Faculty of Computers and Information Sciences, Ain Shams University,
Cairo, Egypt
noha_sabour@cis.asu.edu.eg

Ayman El-Baz
Bioengineering department, Speed School,
University of Louisville,
Kentucky, USA
ayman.elbaz@louisville.edu

Said Ghoniemy
Computer Systems, Faculty of Computers and Information Sciences, Ain Shams University,
Cairo, Egypt
ghoniemy1@cis.asu.edu.eg

Received 2021-5-9; Revised 2021-7-5; Accepted 2021-7-6

Abstract: Autism spectrum disorder (ASD) is a neurodevelopmental disorder associated with impairments in social and lingual abilities. The current gold standard for diagnosis is the autism diagnostic observation schedule (ADOS) combined with expert clinical judgement. The actual cause for autism is still unknown. Early ASD diagnosis is critical for conducting personalized treatment plans and can lead to significant development enhancements. Machine learning techniques, especially deep learning, have been widely incorporated in attempts to develop objective computer-aided technologies to diagnose autism with brain imaging modalities. Task-based functional magnetic resonance imaging (TfMRI) is a brain imaging modality that reveals functional activity of the brain in response to different experiments to study the effects of a brain disease or disorder. This study provides a comprehensive review of research that deploys traditional machine learning and deep learning techniques in diagnosing ASD based on TfMRI. Classification results manifest that TfMRI holds early autism biomarkers and suggests future research to establish multi-modal studies that integrate TfMRI with structural, functional, clinical and genomic data with higher number of participating subjects.

* Corresponding author: Reem Haweel
Computer Systems, Faculty of Computers and Information Sciences, Ain Shams University, Cairo, Egypt
E-mail address: reem.t.haweel@cis.asu.edu.eg
Keywords: Autism, ASD, fMRI, task-based fMRI, diagnosis, computer aided diagnosis systems, machine learning, deep learning, neural networks.

1. Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by impairments in social communication combined with language delay and repetitive behavior [1, 2]. Many etiologies of autism are idiopathic, and affected individuals fall in a wide range of linguistic and intellectual defect severity [3]. The term “spectrum” in autism spectrum disorder refers to the wide severity range. Autism, Asperger’s syndrome, pervasive developmental disorder not otherwise specified and childhood disintegrative disorder belong to the ASD [4] and are conditions that previously were considered separate. Autism disorder is a severe form of ASD, whereas Asperger syndrome is a mild form. Common autistic behaviors include getting nervous by minor surrounding changes, having strong attachments to possessions, underestimation of danger, repeating phrases or movements, avoiding eye contact and having tendency to be alone [5, 6]. According to the centers of disease control (CDC) in 2020, approximately 1 in 54 children living in the United States are diagnosed with ASD [7]. Boys are four times more likely to have autism than girls. One in 34 boys and 1 in 144 girls are identified with autism. Susceptibility to ASD is not affected by ethnicity, race, or socioeconomic status. Although symptoms of autistic children may emerge at the first year [8], ASD is generally diagnosed between the age of three to five years, coinciding with the emergent of noticeable symptoms. However, few ASD children develop normally during the first year. The exact cause and cure for ASD are still unknown. Researchers suggest that genetic factors together with environmental surroundings during the child’s upbringing are the main factors for autism [9]. Early diagnosis and developing a personalized treatment plan are crucial for autistic children development. Educational, skill-oriented, behavioral and occupational therapies substantially improve autistic behavior when introduced at an early age. Early observation by parents contributes to the reduction of false positive rate and reduces unnecessary referrals [10]. However, the up-to-date conventional diagnostic methods for ASD rely mainly on several screening interviews starting at the age of one year or older. These interviews are performed by various professionals such as pediatricians, psychologists, speech language pathologists or psychiatrists [11]. The common standard for ASD diagnosis employs a behavioral instrument such as the autism diagnostic observation schedule (ADOS) [12], autism in toddlers and young children, the American psychiatric association’s diagnostic and statistical manual of mental disorders (DSM)-5 [13], autism diagnostic interview revised (ADI-R) [14,15], the diagnostic instrument for social communication disorders (DISCO), and the computer-generated developmental, dimensional, and diagnostic interview (3di) [16] together with history and expert clinical judgement. ADOS is the gold-standard for diagnosing and symptom quantification of ASD kids. It was published in 2009 by western psychological services (WPS). ADOS provides standardized methods under controlled conditions in performing direct screenings by trained professionals, allowing for allow for
reproducibility. There are four different ADOS modules that are designed to handle age and functional level differences, toddler module, module 1, module 2, module 3 and module 4. ADOS total score is a direct sum of diagnostic items raw scores. Previously, it has been an indicator for ASD severity. However, ADOS total scores could not address severity comparison in different ASD groups over time because of the influence by chronological age and language aptitude. Gotham et al. [12] proposed calibrated severity score (CSS) as a standardized comparison measure of ASD symptom severity independent of age and other conditions. CSS is less affected by developmental functionality over time, such as language ability and maternal education, than ADOS total scores.

The aforementioned exams and assessments coupled with parental observation provide a subjective evaluation. The final diagnosis depends on the individual experience of consulting physicians, the behavior of the patient at the time of assessment, the parent's judgement, and answers on various questionnaires. Therefore, ASD diagnosis instruments are subject to human perceptual and observational errors and cannot be relied upon for handling a quantified correct and accurate diagnosis. Currently, objective computer aided diagnosis (CAD) systems are gaining more attention for their necessity in understanding autism. Several techniques are used for developing CAD systems such as: genetic and blood analysis, magnetic resonance imaging (MRI), electroencephalography (EEG) based analysis, wearables and sensors, and computer vision techniques.

This survey is organized as follows: section two describes TfMRI acquisition and preprocessing. The third section reveals the contribution of machine learning in brain diagnosis. An overview on ASD datasets and main ML and DL classifiers incorporated in the reviewed papers are presented. Section 4 provides a detailed review on ASD diagnosis research using TfMRI. Section 5 discusses several aspects for comparison. Section 6 reveals some limitations and challenges in the reviewed ASD CAD systems. Finally, section 7 concludes the study and provides suggestions for future work.

2. Functional brain imaging

ASD is typically diagnosed at the age of three. However, ASD diagnosis can be investigated as early as 12-14 months old with the advent of brain image modalities such as structural magnetic resonance imaging (sMRI), functional magnetic resonance imaging (fMRI), and diffusion tensor imaging (DTI) [17]. The clinical diagnosis is stable by this time in development [18]. SMRI is a high-resolution brain scanning technique that enables professionals to examine in detail the anatomy of the brain. EEG is the basic modality for identifying functional activity of the brain. It records high temporal resolution (in order of milliseconds) electrical activity of the brain from the scalp. Adopting task-based fMRI (TfMRI) and resting fMRI characteristics can function as early biomarkers of autism [19] for their high spatial resolution (in the order of millimeters).

2.1. Functional MRI acquisition

FMRI is a technique for measuring the activity of the brain. It detects the changes in blood flow and measures the blood oxygen level dependent (BOLD) signals in response to various activities. When a neuron is emitted, the BOLD signal fluctuates in a typical response scenario called the hemodynamic response (HR). As shown in Figure. 1. First immediately after emission, the blood oxygen level decreases causing an initial dip. Then a flow of the blood increases towards the activated brain region, which takes 4 to 5 seconds until it reaches the peak. Then when no further activation is required, the
signal decreases again and undershoots, preparing to rise again to the baseline level. The resolution of the fMRI is usually lower than the resolution of the sMRI because of the fact than fMRI requires multiple scans over time. The resolution normally ranges between 2 and 4 mm, depending on the magnetic field strength of the scanner, e.g. 1.5T, 3T or 7T. The bright colors represent the gray matter, whereas the dark colors represent the white matter, which is the opposite of sMRI representation. There are two designs of the functional experiments: event-related or block designs to record T1fMRI scans, and the resting-state designs to record resting-state fMRI scans.

**Figure 1** The hemodynamic response signal

**event-related design:** The subjects in the scanner are administered to brief and random stimuli for a short period. Typically, there are visual, motor, or audible stimulus. Therefore, the BOLD response includes peaks manifesting short activities. 

**block design:** Multiple similar stimuli (not necessarily identical) are organized in blocks of 10-30 seconds, separated by blocks of baseline. Such a block design elevates the peak in the BOLD signal for a longer time, whereas the peak in the event-related design is just attained for a short period. This plateau introduces the advantage of simplifying the detection of the increase in activation. Figure 2 demonstrates the difference between both event-related and block design.

**resting-state design:** Images in the resting-state designs are acquired without any stimulation. The scanning procedure requires subjects to lay still and rest. This design is used to record the functional connectivity of the brain during the absence of external stimuli.

**Figure 2** Block and event-related design.

### 2.2. Task-based FMRI preprocessing
FMRI data are complex and require many preprocessing steps to be prepared for analysis and classification. Low-level analysis preprocesses fMRI images for each subject, independently. High-level analysis produces group inferences. The pipeline for performing preprocessing entails many common steps. The most employed software tools for preprocessing are fMRI's software library (FSL) [20], FMRI expert analysis tool (FEAT) [21], FreeSurfer [22] and SPM [23]. Such tools have set normal standard preprocessing steps that can be tuned based on preferences and individual cases. They provide efficient performance and reduce execution time. Standard steps for preprocessing mainly include [24]:

- Slice time correction by shifting the voxel's time series of each slice to align all slices within a volume to a reference time point, usually the time of the first slice.
- Motion correction to remove artifacts caused by subject's motion in the scanner. Rigid-body transformations with 6 or 12 DOF can be applied [25] to align volumes to a reference one. The residual motion remaining after motion correction can be added as a confound regressor if found to be substantial.
- Spatial smoothing to enhance noise caused by high frequencies by averaging adjacent voxels.
- High pass temporal filtering to correct for scanner drifts by removing low frequencies.
- Brain extraction using BET on both sMRI and fMRI data, to remove all non-brain data such as the scalp and cerebellum.
- Registration to standardize the fMRI brain images using fMRIB's linear image registration tool (FLIRT) [25] by applying two steps: 1) registration of the functional volumes to their high resolution anatomical sMRI scan, and 2) registration of the anatomical sMRI scan to MNI-152 space [26]. This allows for comparisons between groups and is used to divide the brain into areas and regions of interest (ROI). Widely used atlases are automated anatomical labeling (AAL) [27], Harvard-Oxford (HO) [28], Talaaraich and Tournoux (TT) [29], Craddock 200 (CC200) and Craddock 400 (CC400) [30].

3. Machine Learning for Brain Diagnosis

Machine learning (ML) has made it possible to develop intelligent and automated systems for several pattern recognition applications. The emergence of noninvasive or minimally invasive medical screening devices created massive informative data structures that allowed for the exploitation of ML for automated diagnosis. Medical diagnostic applications assist in accurate kidney, heart, prostate, retinal, and lung diagnosis and abnormality detection. One application is renal transplant functional assessment, utilizing different image modalities (e.g., ultrasound, computed tomography (CT), MRI, etc.) [31,32]. The clinical assessment of myocardial perfusion plays a major role in the diagnosis, management, and prognosis of ischemic heart disease patients [33,34]. Optical coherence tomography has become a powerful modality for the noninvasive diagnosis of various retinal abnormalities [35, 36]. Abnormalities of the lung could also be another promising area of research [37,38].

Machine learning and deep learning (DL), which is a subset of ML that involves deep networks, have played a very important rule in many neuroscience applications because they enable for an accurate detection and assessment of the brain functionality. High performance is attained specially with the availability of high-quality brain scans such as MRI, fMRI, CT, positive emission tomography (PET),
magnetoencephalography (MEG) and EEG. Brain image segmentation [39,40], and detection of brain tumors [41] are ML applications in neural imaging. Moreover, ML applications in psychiatric disorders include the early detection of Alzheimer [42,43], epilepsy [44,45], epileptic seizures detection [46,47], and Parkinson’s disease.[48,49]

To develop diagnosis system for ASD and comprehensive study of ASD brain mechanisms, a large dataset is essential. Single laboratories cannot afford and manage large-scale samples. The national database for autism research (NDAR) is considered the largest heterogeneous data repository of neuroscience and genomic data (“https://ndar.nih.gov”) funded by the US National Institutes of Health (NIH) [50,51]. Data created by ASD research centers and collected at NDAR contain: biomedical data, genetic and genomic findings, clinical assessments, imaging modalities, and quantitative behavioral assessments [52]. NDAR introduced a global unique patient identifier (GUID) which is a very important component that facilitates the replication of research finding and data heterogeneity [53]. There are 117,573 subjects (57,510 ASD subjects and 59,763 typically developed subjects) with data shared at NDAR. Data available for these subjects include 4000 EEGs, 2000 MRIs, more than 10,000 whole exome sequences, and 800 clinical measures.

After preprocessing, some classification studies feed the classifiers with raw preprocessed data. Other studies perform high-level analysis as a feature extraction step. Examples of high-level analysis include general linear model (GLM), sliding window, functional connectivity (CF) matrix, and signal processing techniques such as fast Fourier transformation (FFT), independent component analysis (ICA), and wavelet transforms. Figure 3 demonstrates the main pipeline steps for ASD diagnosis with TfMRI.

After preprocessing, some classification studies feed the classifiers with raw preprocessed data. Other studies perform high-level analysis as a feature extraction step. Examples of high-level analysis include general linear model (GLM), sliding window, functional connectivity (CF) matrix, and signal processing techniques such as fast Fourier transformation (FFT), independent component analysis (ICA), and wavelet transforms. Figure 3 demonstrates the main pipeline steps for ASD diagnosis with TfMRI.

4. ASD diagnosis using task-based fMRI

Recently, with the advent of ML and DL, brain image modalities have been a strong contender for understanding brain and behavior and in diagnosing different brain diseases. Despite the fact that
TfMRI can be considered as a potential autism biomarker, few attempts to deploy it for classification have been recorded; for its limited public availability compared to other scans. The commonly utilized and widely available image modalities used for classification with ML approaches are EEG [54, 55], structural MRI and resting state fMRI [56-58].

Guillaume Chanel in [59] applied multivariate pattern analysis on fMRI scans of two social stimuli experiments, static faces and dynamic bodies. Support vector machines (SVMs) and recursive feature elimination (RFE) were applied for classification on each experiment independently, then an average fusion was applied to get final classification decision. They recorded accuracies between 69% and 92.3% in classifying between 15 ASD and 14 TD subjects. They also derived activation maps based on the selected voxels after RFE. Finally, they correlated the phenotypic information with the classification results.

Another attempt to classify children with ASD from TfMRI scans is demonstrated in [60]. They examined various methods to train generalizable recurrent neural networks (RNN) with long short-term memory (LSTM) on the whole brain including a small medical dataset. Then an augmentation step was performed by resampling the time series of each brain region by calculating the average signal of sub voxel groups, with replacement. Finally, they included non-imaging variables to generate subject-specific initialization of the LSTM network. The classification accuracy for 42 subjects ranged from 51.8% to 69.8% correct diagnosis of 20 ASD vs. 19 TD subjects. They also predicted treatment outcome for 20 ASD individuals.

Xiaoxiao Li et al. [61] attempted to utilize both the temporal and spatial information exhibited in TfMRI. They applied a sliding window method that calculates the mean and standard deviation to extract the temporal statistics. To capture the spatial information, they developed a 3D convolutional neural networks (CNNs) that is fed with 2-channel images created by the sliding window. Finally, they interpreted middle layer outputs of their CNN model to identify useful spatial local information for classification. The dataset included 82 ASD children and 48 age-matched (p > 0.1) controls. The mean F-scores in the proposed framework increased by 8.5%. In continuation of their research, in [62] they added a stage after classification that analyzes ROI extracted feature's effect on the classifier. A frequency-normalized sampling method that created images with corrupted separate ROI in each image was developed. Afterwards, a comparison between the classifier performances on normal and corrupt input was analyzed. Another recent research by Xiaoxiao Li et al. [63] developed a two-stage pipeline for ASD biomarker identification. First, they tested the classification accuracy of an inductive graph neural networks (GNN) on TfMRI scans using different atlases. Second, feature importance of each brain region was evaluated by analyzing the developed GNN.

A study in [64] proposed a pipeline based on TfMRI for predicting treatment of social responsiveness scale. They extracted brain features using GLM and classified using the random forest (RF) classifier. They proposed two step feature selection approaches: first, a "shadow" approach selecting all relevant variables, second, a step wise approach selecting variables that are minimally optimum for prediction. Their study included 20 autistic children (mean age: 5.90 years, SD=1.07 years; 7 females, 13 males). For performance evaluation, they measured mean square error (MSE) and correlation between true and predicted SRS scores in 10 rounds of leave-one-subject-out cross-validation. In comparison to other classifiers, they achieved lower MSE of 125.5 ± 18.5 and higher correlation of 0.75 ± 0.05. The experiment used to scan TfMRI images in [60-64] presented coherent biological and scrambled point-light biological motion movie. It followed a block design of 24s duration for each block [65].

A recent research by [66] proposed a local and global diagnosis for ASD toddlers. They analyzed GLM extracted features from Brainnetome atlas (BNT) parcellated brain areas. The study classified 30 ASD toddlers against 30 TD toddlers using stacked non-negativity constraint auto-encoder (SNCAE) with an accuracy of 75.8%.
Another recent study graded the severity of autism into 3 groups [67, 68]. GLM analysis for low individual level analysis, to extract features, and high group level analysis, to infer statistical differences between groups and validation, were applied. They utilized different approaches to extract features from GLM analyzed whole brain areas. After applying several classifiers, best accuracy of 78% was achieved by RF. The most recent research in ASD diagnosis with TfMRI was proposed by [69]. They developed a comprehensive feature extraction framework for both temporal and spatial feature extraction. First, in the spatial dimensionality reduction steps, they selected significantly activated ROI after performing a group analysis. Second, they clustered each brain area using k-means and extracted from each cluster a varying number of voxels. For temporal feature reduction, they compressed the BOLD Signal using discrete wavelet transform (DWT) compression. They fed the reduced features to a 2D CNN classifier. After testing several parameters, they achieved a classification accuracy of 78%.
### Table 1: Summary of articles published in ASD diagnosis with TfMRI

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Subjects (ASD/TD)</th>
<th>Age(years)</th>
<th>Atlas</th>
<th>Task</th>
<th>Feature extraction</th>
<th>Classifier</th>
<th>Testing</th>
<th>Evaluation results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nicha C. Dvornek [60]</td>
<td>2018</td>
<td>21/19</td>
<td></td>
<td>ASD: 6.05 ± 1.24 TD: 6.42±1.29</td>
<td>AAL</td>
<td>Biopoint task</td>
<td>ROI, augmentation with bootstrapping averaging</td>
<td>RNN with LSTM</td>
<td>69.8% accuracy</td>
</tr>
<tr>
<td>Xiaoxiao Li cite [61]</td>
<td>2018</td>
<td>82/48</td>
<td>none</td>
<td>Data driven from whole brain</td>
<td>Biopoint task</td>
<td>Down sampling and sliding window</td>
<td>2CC3D</td>
<td>85% train, 7% validate 8% test</td>
<td>increased F-scores over 8.5%.</td>
</tr>
<tr>
<td>Xiaoxiao Li [62]</td>
<td>2018</td>
<td>82/48</td>
<td>none</td>
<td>AAL</td>
<td>Biopoint task</td>
<td>Down sampling and sliding window</td>
<td>2CC3D</td>
<td>85% train, 7% validate 8% test</td>
<td>87.10% accuracy</td>
</tr>
<tr>
<td>Xiaoxiao Li [63]</td>
<td>2019</td>
<td>75/43</td>
<td>none</td>
<td>Destrieux atlas and the Desikan-Killiany atlas</td>
<td>Biopoint task</td>
<td>Signal statistical measures, GLM, augmentation, Down sampling and sliding window</td>
<td>GNN</td>
<td>5-fold cross validation</td>
<td>0.76±0.06 accuracy</td>
</tr>
<tr>
<td>Juntang Zhuang [64]</td>
<td>2018</td>
<td>20/0</td>
<td></td>
<td>ASD 5.90 +/- 1.07</td>
<td>Biopoint task</td>
<td>t-statistic image , Social ROI selection, GLM, feature selection</td>
<td>RF</td>
<td>10 rounds of leave-one-out cross-validation, MSE:125.5 ± 18.5, correlation: 0.75 ± 0.05</td>
<td>75.8% accuracy</td>
</tr>
<tr>
<td>Reem Haweel [66]</td>
<td>2019</td>
<td>30/30</td>
<td>[1.1 to 3.1]</td>
<td>BNT</td>
<td>Response to speech</td>
<td>GLM</td>
<td>SNCAE</td>
<td>4-fold cross validation</td>
<td>75.8% accuracy</td>
</tr>
<tr>
<td>Reem Haweel [67,68]</td>
<td>2019</td>
<td>39 (13 Mild, 13 Moderate, 13 Severe)</td>
<td>1.7 +/- 0.42 [1-2.2]</td>
<td>BNT</td>
<td>Response to speech</td>
<td>GLM</td>
<td>RFE, RF</td>
<td>10-fold</td>
<td>72-78% accuracy</td>
</tr>
</tbody>
</table>
5. Discussion

In recent years, TfMRI based ML and DL approaches have emerged and started to gain more attention, especially DL networks. Yet, there is no common protocol for designing experiments while acquiring TfMRI and in choosing scanning parameters. Therefore, it is inconvenient to provide a direct and objective comparison between included studies. However, some useful inferences and general observations can be derived. Moreover, suggestions for future work are also discussed.

Table 1 provides a detailed representation of main characteristics of each included research. The total number of subjects at each study ranged from 20 to 130, which is considered a limited number. The high performance of DL depends on an adequate number of subjects, which is limited in TfMRI studies. However, convenient accuracies above 75% in [59] [62,63] [68-69] have been achieved. These high accuracies suggest TfMRI to hold promise as an ASD biomarker and a valuable contributor in identifying ASD.

The main aim for incorporating brain imaging for diagnosis is to develop an early identifier for autism. The research in [59] performed their fMRI task on adults. [60, 64] included young children above 4 years old in their study. Inference based on brain images of children and adults can be generalized to young children and toddlers to a limited extent. Therefore, studies in [67-69] established their analysis and classification on toddlers with ages in the range of 1 to 3 years old to pursue more reliable and stable diagnosis.

Since fMRI is a high dimensional 4D data with too many data points, feature extraction and reduction techniques are very crucial before developing a classifier, especially traditional ML classifiers. Although DL networks perform automated feature extraction, it is still important to reduce the dimensionality. This is due to the fact that networks perform inefficiently with higher dimensions as in the 4D fMRI data, especially, when the number of extracted features is much less than the number of input samples, which is the common in medical applications. Fortunately, the spatial dimension representing the whole brain contains many redundant data shared in each brain area. Therefore, many studies perform ROI selection [60][64][69]. Either when ROI selection is performed or not, features are selected from each brain area by average, histogram, k-means clustering or down sampling with bootstrapping. The temporal dimension can be reduced without losing information. Approaches performed for temporal reduction include GLM, sliding window and DWT. Another advantage in feature engineering is to introduce a possibility augmentation as in [60][63].

It is desired to produce informative inference from the performance of the classifiers and correlate it with literature, as performed in several included studies. Another opposite approach designs their classifiers based on prior statistical inference. They begin their pipeline with selecting ROI based on literature or GLM group analysis as a feature reduction step [64][69].

Some studies included datasets from public shareable data repository with a license such as NDAR [67-69]. Other studies collected data from their labs. The number of available public TfMRI data for autistic children is still limited in both cases.

CAD systems aim to classify autistic subjects against typically developed peers. Although such diagnosis is useful, it is not sufficient for identifying differences among subjects and locating them across the wide autism spectrum. Haweel et al. [67] has developed an autism severity analysis and grading approach to identify each ASD toddler on the autism spectrum as mild, moderate or severe based on the ADOS CSS score.
6. Challenges

The incorporated TfMRI studies have a relatively small sample size and lack a real control for variability that might exist for performed task and age across a population. These small sample sizes could also fail to provide detection of small differences across groups or more complex developmental influences.

Previous studies suggest a multi-modal approach that may lead to more accurate classification [70]. These modalities may include ERP/EEG, MEG, fNIRs, and DTI/Rest/Structural MRI [71-76]. Such attempts need to combine evidence of MRI modalities, genetics, and early clinical behavioral assessment in larger sample sizes where the prevalence of ASD subgroups could be accurately quantitated. Such classification would also need to include non ASD populations such as larger sample size of TD group, global developmental delay group, and language delayed groups. Therefore, there is a critical need for more public databases with multi-modals.

7. Conclusion

Current ASD diagnosis research is drawing more inference on autism brain image markers. This paper provided a survey on machine learning studies that aim to develop computer aided autism diagnosis by analyzing brain activation related to task-oriented experiments. A comprehensive investigation on sample sizes, ages, feature extraction techniques, ML networks and other performance evaluations have been provided.

In recent years, the number of researches utilizing DL for classification has increased. Out of the 9 referenced studies, 6 employed deep learning algorithms. The majority used CNNs for the popularity, efficient performance and favorable results compared to other networks. Autoencoders and RNN achieved promising results as well. More research to apply DL and ML is encouraged. Examples of such networks include generalized probabilistic neural network, neural dynamic classification, Polynomial Series functional-link artificial neural network (FLANN) [77, 78], Legendre FLANN [79] and deep reinforcement networks.

The performance of current research cannot contribute to a reliable clinical establishment. The reason for that is the limited number of subjects and limited generalizable heterogeneous research that trains and tests performance on data collected from various sites. However, it is expected in the next years that CAD systems will be available for clinicians, especially with the development of multi-modal based studies.

8. Appendix A

In order to analyze the performance of a proposed classifier, there are several used calculations that are either expected to be maximized or minimized. One of the main metrics for evaluation is the confusion matrix, which consists of four main items:

- **True Positive (TP)**: the number of cases classified as having the disease and actually they do have it.
- **True Negative (TN)**: the number of cases classified as not having the disease and actually they don't have it.
- **False Positive (FP)**: the number of cases classified as having the disease and actually they don't have it.
• **False Negative (FN):** the number of cases classified as not having the disease and actually they do have it.

There are different metrics that can be concluded from the confusion matrix:

- **Sensitivity (True Positive rate):** measures the percentage of positive classifications that correctly diagnose a disease.
  \[ \text{TNR} = \frac{TN}{N} = \frac{TN}{(TN + FP)} = 1 - \text{FNR} \]  

- **Specificity (True Negative rate):** measures the percentage of Negative classifications that correctly diagnose not having a disease.
  \[ \text{TPR} = \frac{TP}{P} = \frac{TP}{(TP + FN)} = 1 - \text{FNR} \]  

- **False Positive Rate (FPR):**
  \[ \text{FPR} = \frac{FP}{N} = \frac{FP}{(TN + FP)} = 1 - \text{TNR} \]  

- **False Negative Rate (FNR):**
  \[ \text{FNR} = \frac{FN}{P} = \frac{FN}{(TP + FN)} = 1 - \text{TPR} \]  

- **False Discovery Rate (FDR):**
  \[ \text{FDR} = \frac{FP}{P + TP} \]  

- **Accuracy (ACC):**
  \[ \text{ACC} = \frac{(TP + TN)}{(P + N)} = \frac{(TP + TN)}{(TP + TN + FP + FN)} = 1 - \text{TPR} \]  

- **Receiver Operating Characteristic (ROC) curve:** a graph that manifests the performance of the classifier by presenting the TPR versus the FPR at each classification threshold and plotting a curve connecting them. TPR is represented on the Y-axis and FPR is presented on the X-axis.

- **Area Under the Curve (AUC):** calculates the area under the ROC curve. It aggregates the classification performance values of all thresholds into a value between 0 and 1. More correct classifications lead to higher AUC value.

### Table 2 List of abbreviation

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASD</td>
<td>Autism spectrum disorder</td>
</tr>
<tr>
<td>ADOS</td>
<td>autism diagnostic observation schedule</td>
</tr>
<tr>
<td>TiMRI</td>
<td>Task-based functional magnetic resonance imaging</td>
</tr>
<tr>
<td>CDC</td>
<td>centers of disease control</td>
</tr>
<tr>
<td>DSM-5</td>
<td>diagnostic and statistical manual of mental disorders</td>
</tr>
<tr>
<td>ADI-R</td>
<td>autism diagnostic interview revised</td>
</tr>
<tr>
<td>DISCO</td>
<td>diagnostic instrument for social communication disorders</td>
</tr>
<tr>
<td>3di</td>
<td>developmental, dimensional, and diagnostic interview</td>
</tr>
<tr>
<td>WPS</td>
<td>Western Psychological Services</td>
</tr>
<tr>
<td>CSS</td>
<td>calibrated severity score</td>
</tr>
<tr>
<td>CAD</td>
<td>computer aided diagnosis</td>
</tr>
<tr>
<td>MRI</td>
<td>magnetic resonance imaging</td>
</tr>
<tr>
<td>EEG</td>
<td>electroencephalography</td>
</tr>
<tr>
<td>sMRI</td>
<td>structural magnetic resonance imaging</td>
</tr>
<tr>
<td>fMRI</td>
<td>functional magnetic resonance imaging</td>
</tr>
<tr>
<td>DTI</td>
<td>diffusion tensor imaging</td>
</tr>
<tr>
<td>BOLD</td>
<td>blood oxygen level dependent</td>
</tr>
<tr>
<td>AAL</td>
<td>automated anatomical labeling</td>
</tr>
<tr>
<td>HO</td>
<td>Harvard-Oxford</td>
</tr>
<tr>
<td>TT</td>
<td>Talaraich and Tournoux</td>
</tr>
<tr>
<td>CC200</td>
<td>Craddock 200</td>
</tr>
<tr>
<td>CC400</td>
<td>Craddock 400</td>
</tr>
<tr>
<td>FFT</td>
<td>fast Fourier transformation</td>
</tr>
<tr>
<td>ICA</td>
<td>independent component analysis</td>
</tr>
<tr>
<td>SVM</td>
<td>Support vector machines</td>
</tr>
<tr>
<td>RFE</td>
<td>recursive feature elimination</td>
</tr>
<tr>
<td>RNN</td>
<td>recurrent neural networks</td>
</tr>
<tr>
<td>LSTM</td>
<td>long short-term memory</td>
</tr>
<tr>
<td>CNN</td>
<td>convolutional neural networks</td>
</tr>
<tr>
<td>GNN</td>
<td>graph neural networks</td>
</tr>
<tr>
<td>GLM</td>
<td>general linear model</td>
</tr>
<tr>
<td>FC</td>
<td>functional connectivity</td>
</tr>
<tr>
<td>RF</td>
<td>random forest</td>
</tr>
<tr>
<td>MSE</td>
<td>mean square error</td>
</tr>
</tbody>
</table>
A REVIEW ON AUTISM SPECTRUM DISORDER DIAGNOSIS USING TASK-BASED FUNCTIONAL MRI

<table>
<thead>
<tr>
<th>HR</th>
<th>BNT</th>
<th>Brainnetome atlas</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML</td>
<td>Machine learning</td>
<td>SNCAE stacked non-negativity constraint autoencoder</td>
</tr>
<tr>
<td>CT</td>
<td>computed tomography</td>
<td>DWT discrete wavelet transform</td>
</tr>
<tr>
<td>DL</td>
<td>deep learning</td>
<td>TP True Positive</td>
</tr>
<tr>
<td>PET</td>
<td>positive emission tomography</td>
<td>TN True Negative</td>
</tr>
<tr>
<td>MEG</td>
<td>magnetoencephalography</td>
<td>FP False Positive</td>
</tr>
<tr>
<td>NDAR</td>
<td>national database for autism research</td>
<td>FN False Negative</td>
</tr>
<tr>
<td>NIH</td>
<td>US National Institutes of Health</td>
<td>FPR False Positive Rate</td>
</tr>
<tr>
<td>GUID</td>
<td>global unique patient identifier</td>
<td>FNR False Negative Rate</td>
</tr>
<tr>
<td>FSL</td>
<td>fMRI's software library</td>
<td>FDR False Discovery Rate</td>
</tr>
<tr>
<td>FEAT</td>
<td>FMRI expert analysis tool</td>
<td>ACC Accuracy</td>
</tr>
<tr>
<td>FLIRT</td>
<td>fMRIB's linear image registration tool</td>
<td>ROC Receiver Operating Characteristic</td>
</tr>
<tr>
<td>ROI</td>
<td>region of interest</td>
<td>AUC Area Under the Curve</td>
</tr>
<tr>
<td>2CC3D</td>
<td>2 channel 3D CNN</td>
<td></td>
</tr>
</tbody>
</table>

References

23. J. Ashburner, Computational anatomy with the spm software, Magnetic resonance imaging, 27(8), (2009) 1163-1174.


