Article

Detection of ASD Children through Deep-Learning Application of fMRI

Min Feng [1,2,*] and **Juncai Xu** [3]

1   Nanjing Rehabilitation Medical Center, The Affiliated Brain Hospital, Nanjing Medical University, Nanjing 210029, China
2   School of Chinese Language and Literature, Nanjing Normal University, Nanjing 210024, China
3   School of Engineering, Case Western Reserve University, Cleveland, OH 44106, USA; juncai.xu@gmail.com
*   Correspondence: minfeng@njmu.edu.cn

Abstract: Autism spectrum disorder (ASD) necessitates prompt diagnostic scrutiny to enable immediate, targeted interventions. This study unveils an advanced convolutional-neural-network (CNN) algorithm that was meticulously engineered to examine resting-state functional magnetic resonance imaging (fMRI) for early ASD detection in pediatric cohorts. The CNN architecture amalgamates convolutional, pooling, batch-normalization, dropout, and fully connected layers, optimized for high-dimensional data interpretation. Rigorous preprocessing yielded 22,176 two-dimensional echo planar samples from 126 subjects (56 ASD, 70 controls) who were sourced from the Autism Brain Imaging Data Exchange (ABIDE I) repository. The model, trained on 17,740 samples across 50 epochs, demonstrated unparalleled diagnostic metrics—accuracy of 99.39%, recall of 98.80%, precision of 99.85%, and an F1 score of 99.32%—and thereby eclipsed extant computational methodologies. Feature map analyses substantiated the model's hierarchical feature extraction capabilities. This research elucidates a deep learning framework for computer-assisted ASD screening via fMRI, with transformative implications for early diagnosis and intervention.

Keywords: autism spectrum disorder; ASD screening; convolutional neural networks; deep learning; fMRI

1. Introduction

Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder that is characterized not only by impairments in social interaction and communication but also by restricted, repetitive patterns of behavior, interests, or activities, as outlined in DSM-5 criteria [1,2]. The imperative for early diagnostic endeavors is crucial for the initiation of targeted therapeutic interventions. However, traditional diagnostic paradigms hinge largely on subjective evaluations, often resulting in protracted and indeterminate diagnostic outcomes [3,4]. In contrast, functional magnetic resonance imaging (fMRI) has emerged as an invaluable instrument for the objective and meticulous diagnosis of ASD [5–8]. Beyond its diagnostic utility, fMRI is pertinent in personalized metabolic therapies in ASD, such as integrated PET/MRI systems, and offers the advantage of no exposure to radiation. To further substantiate the utility of fMRI in ASD diagnostics, it is worth noting that fMRI studies have revealed distinct neural connectivity patterns in children with and without ASD [9]. Specifically, increased local connectivity and decreased long-range connectivity have been observed in ASD subjects [10]. These findings have guided the selection of fMRI variables that were integrated into our deep-learning algorithm.

Building on these earlier observations, numerous scholarly investigations have rigorously probed the applicability of fMRI in the ASD diagnostic landscape, frequently incorporating machine-learning frameworks to augment diagnostic precision [11–16]. Foundational research, for instance, scrutinized fMRI's capacity to discern aberrant neural connectivity in individuals afflicted with ASD, while supplementary inquiries concentrated on

algorithmic automation [9,17,18]. Subsequent scholarship delved into the integration of machine-learning algorithms within the analytical matrix of fMRI data, the elucidation of latent neural architectures that are pertinent to ASD, and the implementation of complex computational models for diagnostic purposes [17,19–21]. Despite these advancements, the scholarly focus remains disproportionately oriented toward adult populations, thereby creating a conspicuous void in pediatric-focused research [19,22,23]. A review of prior deep-learning studies reveals a focus on adult populations, the employment of various fMRI data, and the involvement of children with different characteristics.

Convolutional neural networks (CNNs) have garnered substantial acclaim for their proficiency in image-recognition tasks that have extended into the domain of medical imaging [24–26]. The intrinsic capability of CNNs to autonomously and adaptively decipher spatial feature hierarchies renders them uniquely apt for the intricate analysis of fMRI datasets [27–29]. Despite the availability of the Autism Brain Imaging Data Exchange (ABIDE I) dataset since 2014 and the NDAR dataset since 2006, there exists a significant lacuna in scholarly literature that integrates fMRI data with CNN methodologies for the targeted early diagnosis of ASD in pediatric populations.

The overarching objective of the current investigation was to address these existing research gaps. A CNN framework was deployed for the rigorous analysis of fMRI scans, targeting the incipient detection of ASD in pediatric subjects. In doing so, the study furnished a swift, unerring, and impartial diagnostic conduit, thereby enhancing the prospects for timely and effective therapeutic interventions. The seminal contributions of this study are delineated as follows:

- This study introduced a custom deep CNN for fMRI data, achieving high accuracy in distinguishing pediatric ASD from typical development.
- The study uncovered the model's feature learning through in-depth analysis and feature-map visualization, enabling robust ASD discrimination.
- The investigation yielded invaluable insights via feature-map analysis, corroborating the efficacy of the CNN architecture in hierarchical feature learning.

This study employed convolutional neural networks to scrutinize fMRI data for the nascent diagnosis of ASD in pediatric demographics. Subsequent sections of this article elucidate the data acquisition, preprocessing, and segmentation protocols that were used to compile an exhaustive dataset (Section 2); the proposed CNN architecture and its training regimen (Section 3); the empirical evaluation encompassing performance metrics, confusion matrix analytics, and feature-map visualizations (Section 4); and the concluding remarks and avenues for future research (Section 5).

2. Materials and Methodology

2.1. Data Collection and Preprocessing

MRI scans, obtained with a 3 Tesla Allegra, were performed within three months post-diagnosis for pediatric subjects. The data for this study were obtained from the Autism Brain Imaging Data Exchange (ABIDE I) repository, specifically originating from the New York University Langone Medical Center [17]. This subset encompassed 56 individuals who were clinically diagnosed with ASD and 70 age-matched controls who exhibited typical development (TD). Children who were on psychostimulants ceased medication 24 h before scanning, pending physician approval. During the resting-state fMRI, most participants relaxed with open eyes, focusing on a white cross-hair displayed against a black screen. The phenotypic attributes of the participants from the NYU dataset are systematically detailed in Table 1.

Table 1. The phenotypic characteristics of children who participated in this study.

Category	Value
Total	126
ASDs	56

Table 1. *Cont.*

Category	Value
TDs	70
ASD Males	51
ASD Females	5
TD Males	51
TD Females	19
Age Range of ASDs (Years)	7.13–18.00
Age Range of TDs (Years)	6.47–18.00
Average Age (SD) (Years)	12.02 (2.97)
Average ADOS scores of ASDs (SD)	11.20 (4.27)

In our rigorous investigation, we utilized pediatric fMRI data that had undergone a meticulous preprocessing regimen via the connectome computation system (CCS). This widely accessible computational framework offered an exhaustive preprocessing pipeline. Specific procedures within this pipeline encompassed a range of techniques, from the registration of anatomical brain masks to functional images using FMRIB's linear image registration tool (FLIRT) to slice-timing correction achieved with 3dTshift. Additional steps included skull stripping via Analysis of Functional NeuroImages' 3dAutomask, motion correction through 3dvolreg, voxel intensity normalization, nuisance signal removal, and band-pass filtering in the 0.01–0.1 Hz frequency range.

In the data-preprocessing phase, the four-dimensional fMRI datasets, stored in the Neuroimaging Informatics Technology Initiative (NIFTI) format, underwent a series of complex transformations. The initial operation entailed the decomposition of each four-dimensional dataset into a series of three-dimensional volumes, each corresponding to a specific temporal instance (Figure 1).
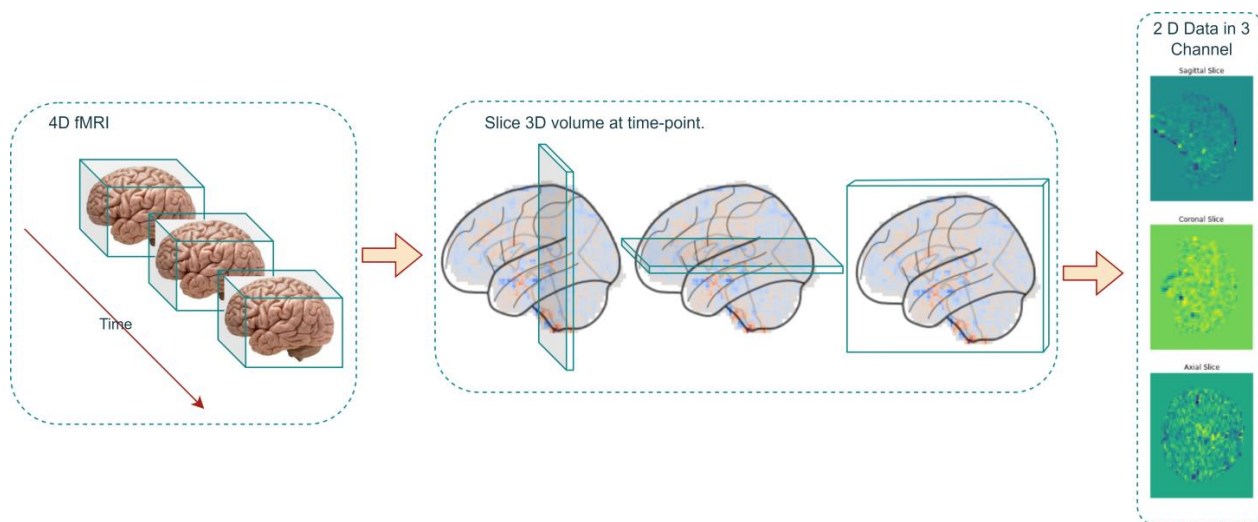

Figure 1. Process of converting 4D fMRI image to 2D echo planar data.

Mathematically, this decomposition is articulated as follows (Equation (1)):

$$V(x, y, z) = F(x, y, z, t_i) \tag{1}$$

where $V(x, y, z)$ represents the three-dimensional volume extracted at a particular time t_i. $F(x, y, z, t_i)$ denotes the original four-dimensional fMRI dataset, and t_i specifies the temporal instance under consideration.

Subsequent to this, the extracted three-dimensional volumes were further partitioned into two-dimensional slices. These slices were oriented along the anatomical planes—

sagittal, coronal, and axial—and were amalgamated to form a two-dimensional, three-channel data representation.

The mathematical formalism for this partitioning is encapsulated in the following equation (Equation (2)):

$$D = [S(x, z), C(y, z), A(x, y)] \tag{2}$$

where D constitutes the two-dimensional, three-channel data representation. Here, $S(x, z)$, $C(y, z)$, and $A(x, y)$ denote slices obtained along the sagittal, coronal, and axial planes, respectively. Specifically, $S(x, z) = V(x, y\_i, z)$, $C(y, z) = V(x\_i, y, z)$, and $A(x, y) = V(x, y, z\_i)$.

This intricate transformational sequence culminated in the conversion of each four-dimensional fMRI dataset into an array comprising 176 two-dimensional images. This array served as a rich substrate for subsequent analytical investigations.

This methodological approach generated a voluminous dataset of 2D echo planar data in three channels, amounting to 22,176 samples. Of this corpus, 17,740 samples were earmarked for the training set, while an additional 20% were set aside for validation purposes. The remaining 4436 samples constituted the test dataset. This meticulous procedure of data transformation and slicing effectively facilitated the assembly of a comprehensive echo planar dataset, optimally configured for subsequent deep-learning analyses.

2.2. Convolutional Neural Network Framework

In the present study, the CNN model was employed to analyze 2D echo planar data obtained from fMRI scans for the purpose of recognizing ASD in pediatric populations. The architectural framework of the model incorporated convolutional layers, max pooling layers, dense layers, batch normalization, the dropout technique, and a terminal softmax activation function [30–33]. To facilitate binary classification, the cross-entropy loss function was utilized, while model optimization was achieved through the Adam optimizer [34,35].

The convolutional layer is fundamental to any CNN. It applies filters to the input image to create feature maps, capturing local patterns [30,36]. The equation for this operation is as follows:

$$O_{ij} = \sum_m \sum_n D_{(i+m)(j+n)} \times K_{mn} \tag{3}$$

where O_{ij} is the output feature map, D is the input image, K is the kernel, and m, n are the dimensions of the kernel.

Max pooling layers serve to downsample the spatial dimensions of the input volume, thereby emphasizing salient features for more efficient computational processing. The equation for max pooling is as follows:

$$M = \max(R) \tag{4}$$

where M is the max-pooled output and R is a region in the input feature map.

Dense layers perform complex transformations on the flattened feature maps, contributing to the final classification. The equation for the dense layer is as follows:

$$y = \text{Activation}(W \times x + b) \tag{5}$$

where y denotes the output, W are the weights, x is the input, and b is the bias term.

Batch normalization normalizes the output from the activation function, aiding in faster and more stable training. The equation for batch normalization is as follows [32]:

$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} \tag{6}$$

where $\hat{x}$ is the normalized output, x is the input, μ is the mean, σ^2 is the variance, and ε is a small constant introduced to prevent division by zero.

The softmax activation function converts the network's output into a probability distribution over the target classes. The equation for softmax is as follows:

$$S(y_i) = \frac{e^{y_i}}{\sum\limits_{j} e^{y_j}} \tag{7}$$

where $S(y_i)$ is the softmax output for class i, y_i is the input for class i, and y_j is the input for all classes.

For the binary classification task, the cross-entropy loss function serves as the chosen loss metric. The equation for this loss function is as follows [34]:

$$L(y, \hat{y}) = -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})) \tag{8}$$

where L is the loss, y is the actual label, and $\hat{y}$ is the predicted label.

To optimize the model, the Adam optimizer was employed. The equation for the Adam optimizer is as follows [35]:

$$\theta_{t+1} = \theta_t - \frac{\alpha \cdot m_t}{\sqrt{v_t} + \epsilon} \tag{9}$$

where θ are the parameters, α is the learning rate, m_t and v_t are the first and second moment estimates, respectively, and ϵ is a small constant to avoid division by zero.

CNNs undergo parameter optimization via backpropagation algorithms that are aimed at minimizing a predefined loss function [37]. The optimizer updated the weights based on the gradients computed during backpropagation, refining the model for better performance.

2.3. Workflow for ASD Recognition Using CNNs

The establishment, training, and testing of a CNN model for ASD recognition is a rigorous, multi-step process that adheres to academic standards and mathematical formalism. The process commenced with the acquisition of 4D fMRI data from the ABIDE dataset (Algorithm 1). These data underwent comprehensive preprocessing, including motion correction and normalization, and were then decomposed into 3D volumes and 2D anatomical slices. These slices served as the CNN input, facilitating subsequent model initialization, training, and testing.

Algorithm 1: CNN Model for ASD Recognition

Data: Preprocessed 4D fMRI data $F(x,y,z,t)$
Result: ASD Recognition Performance
1   **Step 1: Data Preprocessing and Slicing**
2   **for** t_i in T *(Temporal Instances)* **do**
3       $V(x,y,z) = F(x,y,z,t_i)$;
4       $D = [S(x,z),C(y,z),A(x,y)]$;
5   Add D to Dataset
6   **Step 2: Dataset Partitioning**
7   Divide Dataset into Training, Validation, and Test Sets
8   **Step 3: Model Initialization**
9   Initialize CNN Model Parameters θ
10 **Step 4: Model Training**
    for $i = 1$ to N *(Iterations)* **do**
11   Train CNN Model on Training Set using θ and Adam Optimizer;
12   Validate Model on Validation Set **if** *Validation Loss $\leq$ Threshold*
13   **then**
14   Break
15 **Step 5: Model Testing**
16 Test Trained Model on Test Set
17 **Step 6: Performance Evaluation**
18 Compute ASD Recognition Performance Metrics

In accordance with the aforementioned pseudocode, the procedural framework for ASD recognition is delineated as follows:

Step 1: Data Preprocessing and Slicing

4D fMRI datasets are converted into 3D volumes, each representing a specific time t_i, as $V(x, y, z) = F(x, y, z, t\_i)$. These volumes are sliced into 2D images along anatomical planes to form a 2D, three-channel data $D = [S(x, z), C(y, z), A(x, y)]$.

Step 2: Dataset Partitioning

The dataset is methodically partitioned into training, validation, and testing subsets, serving the distinct functions of model training, hyperparameter optimization, and evaluative performance assessment, respectively.

Step 3: Model Initialization

The CNN model is instantiated with an initial set of stochastic parameters θ, thereby establishing the commencement point for subsequent optimization.

Step 4: Model Training

The model is subjected to a series of iterative training cycles, employing the Adam optimization algorithm to fine-tune parameters and minimize the objective loss function. The training phase culminates either upon achieving a pre-specified validation loss threshold or after a predetermined epoch count.

Step 5: Model Testing

The fully developed model undergoes stringent evaluation on the test subset to assess its capacity for generalization and its aptitude in accurately classifying ASD cases.

Step 6: Performance Evaluation

A comprehensive suite of evaluative metrics, encompassing accuracy, precision, recall, and F1 score, are computed to rigorously quantify the model's capabilities in ASD recognition.

3. Experimental Setup

3.1. Parameters for CNN

The given model is a CNN designed for image classification, comprising several key layers (Figure 2). The Conv2D layers are essential for feature extraction, applying filters to the input image to detect patterns like edges and textures. The MaxPooling2D layers reduce the spatial dimensions, summarizing the features in each region and thereby reducing overfitting. Batch normalization layers are used to normalize the activations, improving training speed and stability. The fully connected layers, commonly referred to as dense layers, execute classification tasks predicated on the features ascertained through prior computational layers, with the final dense layer using a softmax activation function to classify the input into two categories. Dropout layers are strategically integrated within the network architecture, serving to stochastically nullify a subset of input units during the training phase as a regularization technique to mitigate overfitting. Together, these layers form a robust architecture that is capable of learning complex patterns in the input data, making it suitable for various image-recognition tasks. The layer parameters of the proposed neural network are shown in Table 2.
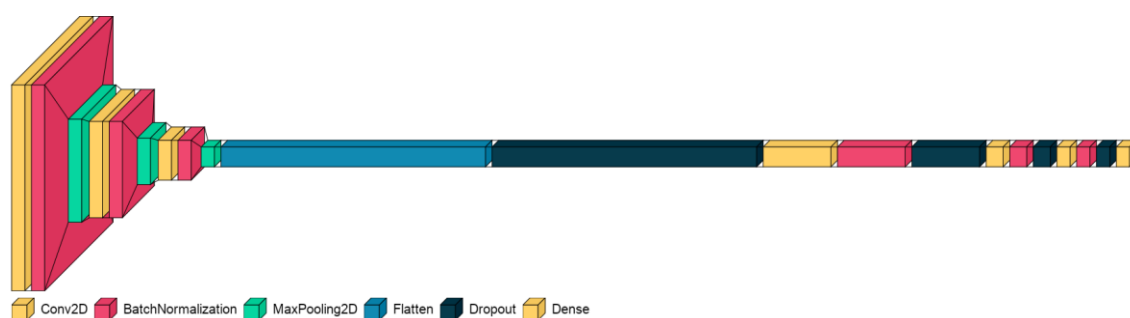

Figure 2. An advanced CNN framework for ASD classification.

Table 2. Configuration and parameterization of layers in the proposed network architecture.

Layer (Type)	Output Shape	Param
conv2d 0	(None, 62, 62, 32)	896
batch normalization	(None, 62, 62, 32)	128
max pooling2d	(None, 31, 31, 32)	0
conv2d 1	(None, 29, 29, 64)	18,496
batch normalization 1	(None, 29, 29, 64)	256
max pooling2d 1	(None, 14, 14, 64)	0
conv2d 2	(None, 12, 12, 128)	73,856
batch normalization 2	(None, 12, 12, 128)	512
max pooling2d 2	(None, 6, 6, 128)	0
flatten	(None, 4608)	0
dropout	(None, 4608)	0
dense	(None, 1024)	4,719,616
batch normalization 3	(None, 1024)	4096
dropout 1 (Dropout)	(None, 1024)	0
dense 1 (Dense)	(None, 256)	262,400
batch normalization 4	(None, 256)	1024
dropout 2 (Dropout)	(None, 256)	0
dense 2 (Dense)	(None, 64)	16,448
batch normalization 5	(None, 64)	256
dropout 3 (Dropout)	(None, 64)	0
dense 3 (Dense)	(None, 2)	130

The CNN model is tailored for binary classification, with a structure designed to balance complexity and computational efficiency. The "data shape" parameter, derived from the training data, ensures that the model aligns with the input dimensions, while the "kernel_size" of (3, 3) allows for the capture of local patterns without excessive computational cost. Initiating with 32 filters, the architecture incrementally doubles the quantity of convolutional kernels subsequent to each max-pooling layer, thereby augmenting its capacity for intricate pattern recognition. With "n_classes" set to 2, the output layer is configured for binary classification, using a softmax activation function to provide probability estimates. Collectively, these settings reflect a common and effective approach in deep learning, aimed at learning hierarchical features from image data for binary classification tasks.

3.2. Training Model

The algorithm designed for ASD recognition was instantiated on a Python 3 environment and architected atop the Keras framework. For enhanced computational efficiency and expedited training, the model was deployed on an NVIDIA A100 graphics processing unit (GPU). This amalgamation of hardware and software specifications engendered both robust performance and scalability, thereby facilitating the seamless execution of the algorithm's intricate mathematical computations and iterative training cycles.

For the model compilation, the loss function employed was "crossentropy", which is particularly apt for multi-class classification scenarios, where labels are represented as integers. The optimization of the model's weights was conducted via the "Adam" optimizer, and the performance was gauged using the "accuracy" metric. The training process was delineated across 50 epochs with a batch size of 64, implying that weight adjustments occurred every 64 samples. A fraction of 20% of the training data was earmarked for validation, serving as an evaluative gauge for the model's generalization capabilities. This orchestrated confluence of hyperparameters and methodologies constituted the training regimen, guiding the model to discern intricate patterns in the data for effective classification. Upon the culmination of this training regimen, the wall time—indicative of the actual elapsed duration for model training—was recorded at 1 min and 52 s. Figures 3 and 4 elucidate the trajectory of accuracy and loss metrics throughout the training phase.
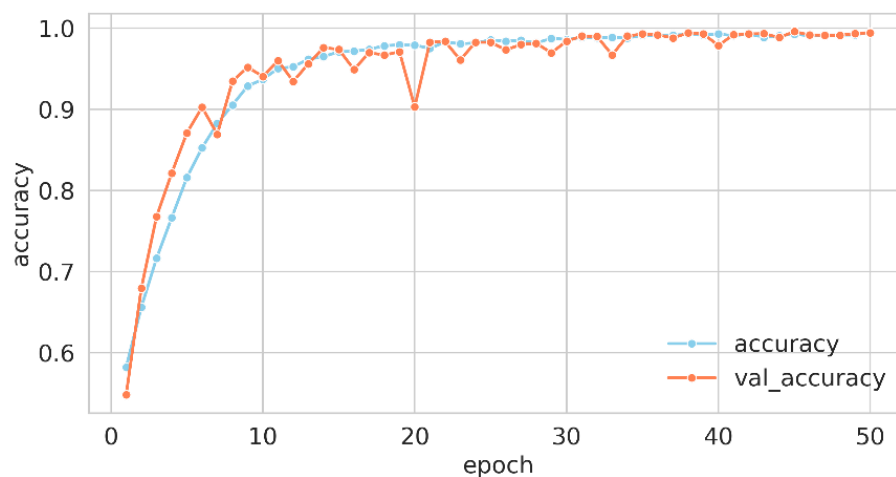
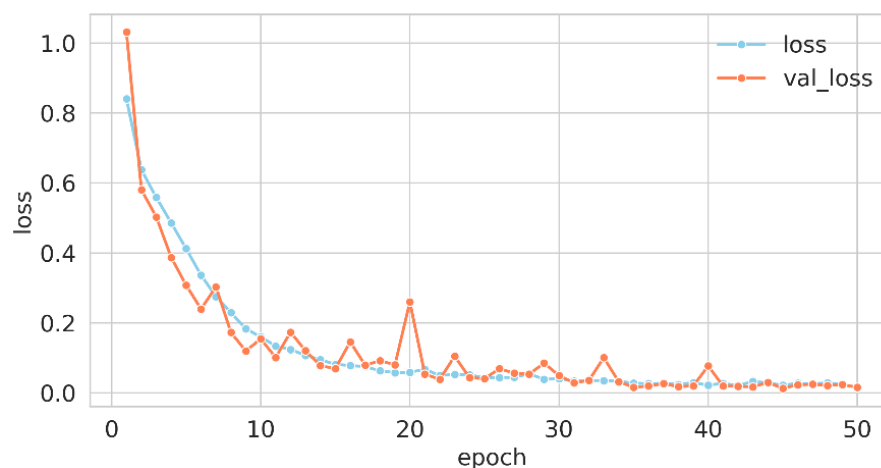**Figure 3.** Training accuracy curve for proposed net.

Figure 4. Training loss curve for proposed net.

The training history data—comprising accuracy, validation accuracy, loss, and validation loss—were retrieved from a file and plotted over 50 epochs. Two line plots were generated: one for accuracy and the other for loss. In each plot, training and validation metrics were distinguished by separate lines and colors, with markers emphasizing values at each epoch. Labels, legends, and suitable font sizes were incorporated for clarity.

The accuracy plot revealed the model's learning trajectory in precise classification. The sky-blue line for training accuracy and the coral line for validation accuracy presented the model's efficacy on both datasets across 50 epochs. An ascending trend in both lines signified effective learning, while a notable discrepancy may have hinted at overfitting or underfitting. The loss plot delineated the decrement in the model's error over time. Sky-blue for training loss and coral for validation loss represented errors on respective data. A declining trend in both was usually favorable, reflecting error minimization. Conversely, a rising validation loss, in contrast to a declining training loss, may have signified overfitting, a condition in which the model performed exceptionally well on the training data but exhibited poor generalization to new, unseen data.

3.3. Evaluation Metrics

In the assessment of binary classification models, four principal metrics—accuracy, precision, recall, and F1 score—are employed to gauge a model's performance. These

metrics collectively furnish a multifaceted evaluation of a model's capability to accurately categorize instances. The mathematical formulations for these key metrics are as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{11}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{12}$$

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{13}$$

where TPs is true positives, TNs are true negatives, FPs are false negatives and FNs are false negatives.

4. Results and Analysis

4.1. Model Performance

Utilizing these four indicators in tandem allows for a nuanced understanding of a model's strengths and weaknesses, guiding further refinement and optimization. The four indicators for the proposed model are shown in Table 3.

Table 3. Evaluation of the proposed model's performance on the test dataset.

Metric	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Value	99.39	99.85	98.80	99.32

The efficacy of the proposed model was quantified by multiple key metrics, as delineated in the accompanying table. An impressive accuracy rate of 99.39% underscored the model's robust capability for accurate instance classification. The precision rate of 99.85% reflected the model's accuracy in predicting positive instances, indicating that almost all predicted positives were true positives. The recall rate, at 98.80%, represented the model's sensitivity in identifying all actual positive instances, while the F1 score of 99.32% offered a balanced view of precision and recall, providing a holistic metric of the model's robustness. Collectively, these metrics highlighted the model's exceptional performance in both general classification and the nuanced handling of positive instances.

The confusion matrix (Figure 5) served as a graphical elucidation of the model's discriminative efficacy between the two diagnostic categories: "TD" and "ASD". Within this matrix, rows delineated the ground-truth classifications, while columns signified the model's predicted outcomes. The percentage values showed the proportion of correct and incorrect classifications: 99.88% of "TD" were correctly classified, and only 0.12% were misclassified as "ASD"; 98.80% of "ASD" were correctly identified, and 1.20% were misclassified as "TD". These percentages highlighted the model's high accuracy in distinguishing between the two classes, with very low rates of misclassification. This granular perspective facilitated a comprehensive evaluation of the model's merits and potential avenues for enhancement, thereby substantiating its robust performance in the specified classification task.

To further validate the CNN model, a five-fold cross-validation approach was employed, offering a more robust assessment of model performance compared to a singular train/test split. This method entailed partitioning the data into five equal subsets, followed by five iterations of training and evaluating the model—each iteration utilizing a distinct subset as the test set while amalgamating the remaining data for training. Through the examination of model performance across these multiple folds, a more reliable estimation of the model's generalization capacity on unseen data was attained.
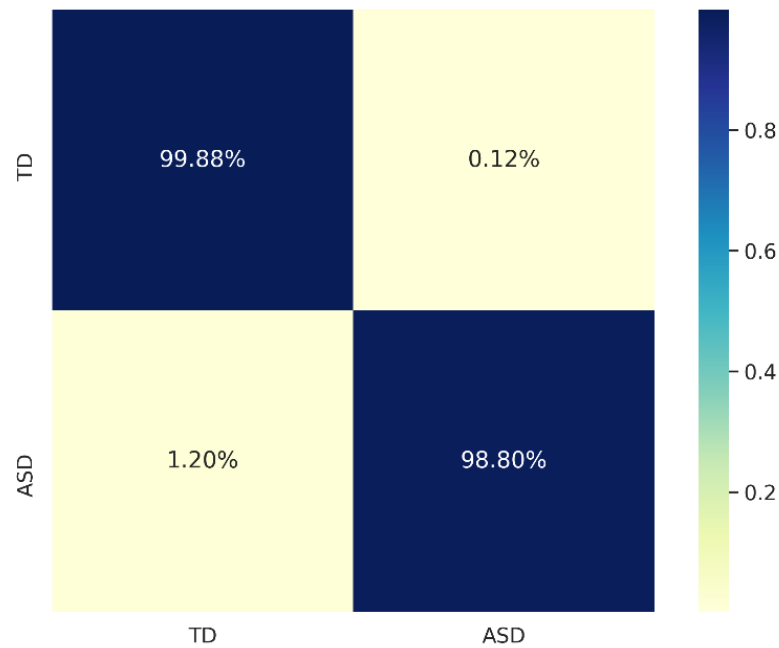
Figure 5. Confusion matrix for ASD and TD classification.

The fold accuracies, as delineated in Table 4, showcased a commendable level of consistency, signifying proficient performance of the model across varied data subsets. These values denoted the proportion of correct classifications within each fold, hence elucidating the model's robustness. Despite a high accuracy continuum across all folds, it accentuated the model's adeptness in differentiating between the "TD" and "ASD" classes.

Table 4. Accuracies of five-fold cross-validation.

Fold Number	1	2	3	4	5
Accuracy (%)	99.39	99.68	99.70	99.32	99.09

4.2. Feature Map Analysis

The given model architecture comprised a three-layer convolutional deep network that was designed to process fMRI graphics, and it provided a hierarchical approach to feature extraction (Figures 6–8).
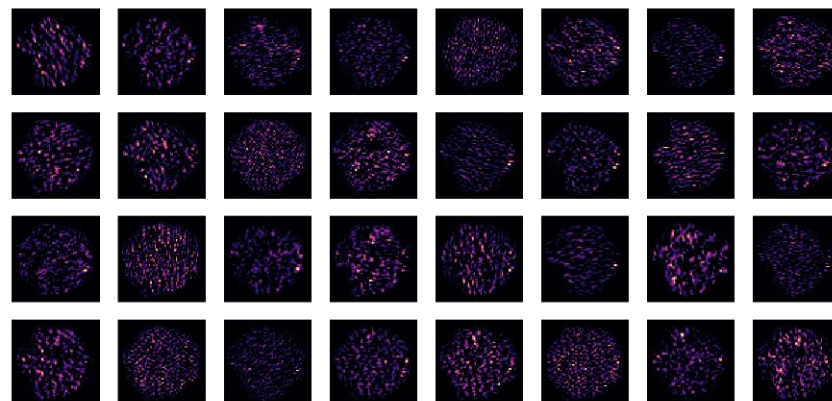

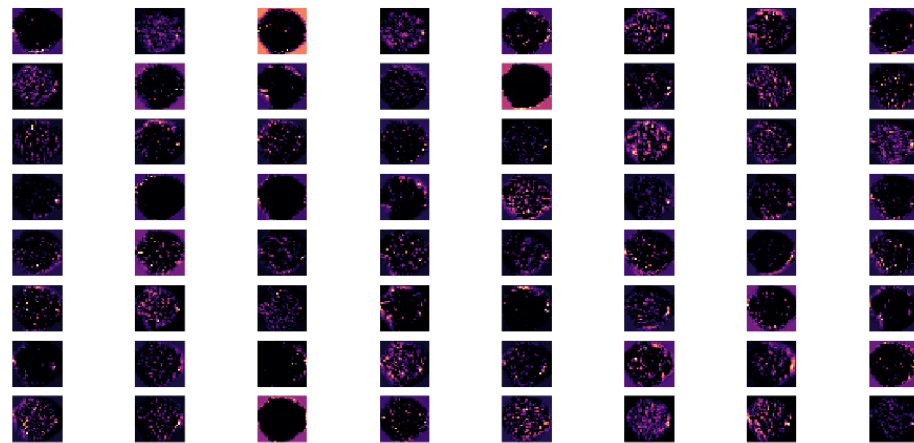**Figure 6.** The feature map of conv2d 0 layer.

Figure 7. The feature map of conv2d 1 layer.

Figure 8. The feature map of conv2d 2 layer.

The inaugural Conv2D layer ingested fMRI images and employed a predetermined quantity of filters with a designated kernel size, subsequently invoking a ReLU activation function (Figure 6). This layer was instrumental in discerning low-level features, such as edges and rudimentary shapes, within the fMRI data. The subsequent BatchNormalization normalized the activations, and MaxPooling2D reduced the spatial dimensions, focusing on the most important information. The quantity of filters was subsequently doubled for the ensuing layer.

The next Conv2D layer, with twice as many filters of the first, captured more complex and abstract features, such as textures and patterns (Figure 7). The ReLU activation introduced non-linearity, allowing the model to learn more intricate representations. Subsequent to the convolutional layer, BatchNormalization and MaxPooling2D were employed, the latter serving to further downsample the feature maps. Concurrently, the quantity of filters was doubled.

The third Conv2D layer continued this process, now with four times the original number of filters (Figure 8). At this stage, the model was likely detecting high-level characteristics, such as specific regions or structures within the fMRI images, that were germane to the specific diagnostic task. The ReLU activation, BatchNormalization, and MaxPooling2D continued the pattern of introducing non-linearity, normalizing activations, and reducing spatial dimensions.

Throughout these three layers, the model transitioned from detecting simple, spatial features to more abstract and complex characteristics. The progressive doubling of filters and the consistent use of ReLU, BatchNormalization, and MaxPooling2D created a hierarchy of feature maps that increased in abstraction and complexity. This hierarchical structure

was particularly suitable for interpreting fMRI graphics, where understanding both local details and global structures is essential. The final fully connected layers and dropout layers further refined the model's ability to classify based on these extracted features.

In this experiment, 500 samples were randomly extracted from the test set to investigate the role of feature maps in a deep neural network across four layers—three convolutional and one fully connected (Figure 9). After applying t-SNE for dimensionality reduction, the feature maps from these layers were visualized as scatter plots. The first convolutional layer showed moderate feature extraction capabilities, capturing basic elements like edges and textures, but with less distinct clustering. As we progressed to the second and third convolutional layers, the clusters became increasingly defined, indicating the improved feature extraction of more complex structures. The fully connected layer's feature map ideally displayed the most distinct clusters, indicating that it effectively combined high-level features for final classification.
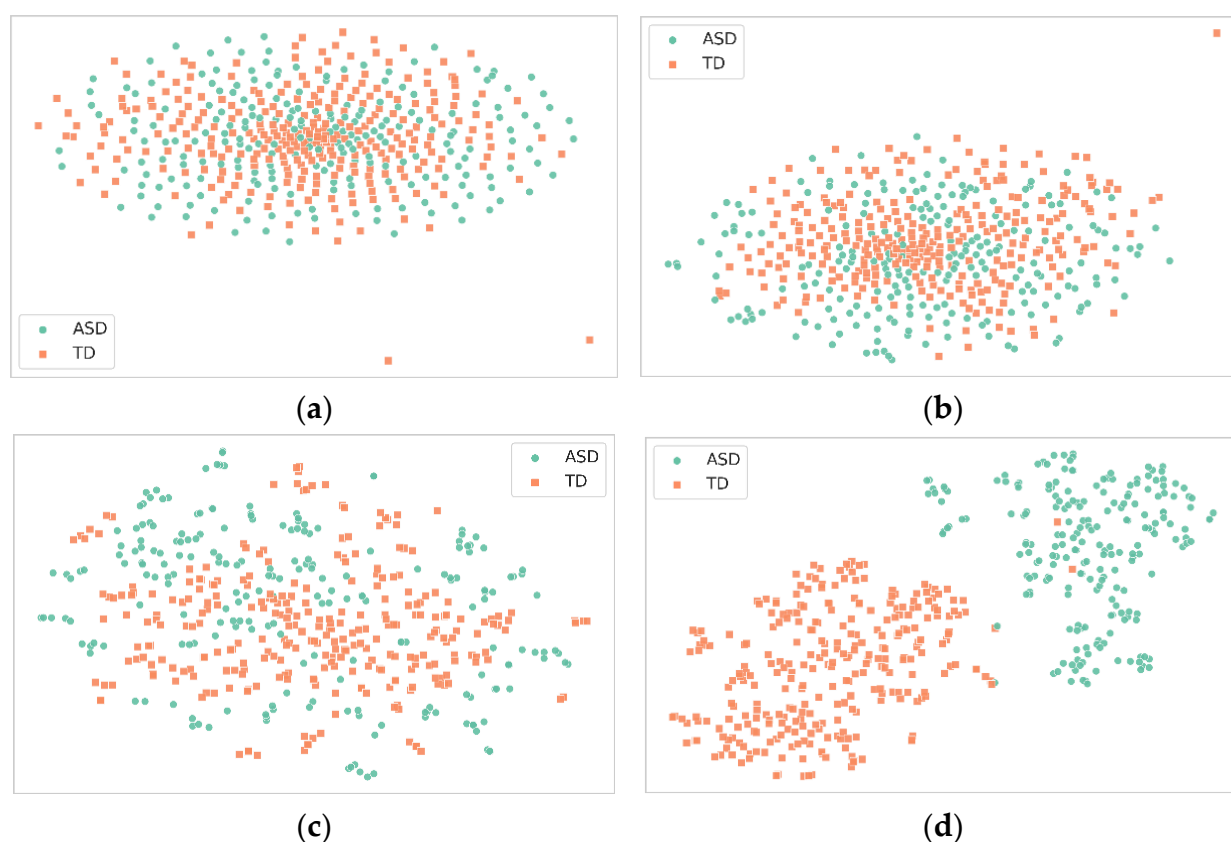

Figure 9. Visualizing t-SNE-reduced outputs of the proposed model's layers: (**a**) first convolutional layer, (**b**) second convolutional layer, (**c**) third convolutional layer, (**d**) initial fully connected layer.

The analysis of these feature maps leads to several conclusions. First, the network demonstrates a clear hierarchy of feature extraction, with each subsequent layer capturing increasingly complex features. This is evident from the progressively distinct clusters formed in the scatter plots. Second, the third convolutional stratum and the fully connected tier exhibit the most promising discriminatory power, suggesting that they play a critical role in effectively classifying between "TD" and "ASD". These observations affirm the importance of deeper layers in the network for complex classification tasks, thereby validating the architecture's efficacy.

4.3. Discussion

The proposed CNN model demonstrated a remarkable capability in distinguishing ASD subjects from TD subjects. This robust performance, substantiated by the model's

hierarchical feature extraction, suggested that the model captured meaningful and generalizable representations from the fMRI data. The high accuracy of the model has significant ramifications for the field of ASD diagnosis. Unlike traditional diagnostic methods, which often rely on subjective assessments and which are time-consuming, the proposed model offers a rapid, objective, and highly accurate diagnostic tool. This could be particularly transformative for preliminary ASD screening and could also contribute to the identification of novel biomarkers or neural signatures associated with ASD.

When the model's performance is juxtaposed with existing literature, its superiority becomes evident (Table 5). This study introduced a CNN model that not only attained an unparalleled accuracy rate of 99.39% in differentiating ASD from typical development, but also provided expeditious, objective, and highly precise ASD screening, with transformative potential for early diagnosis and intervention. This accuracy substantially outperformed the closest competing study by Yakolli et al. (2023), which achieved 88.0% accuracy using a CNN model. The performance delta is even more pronounced when it is compared to other methodologies, such as deep-broad learning and decision models, which have accuracies around the 70% mark. The model's exceptional performance was not confined to CNN-based methodologies, but extended to diverse computational approaches, including recurrent neural networks (RNNs) and capsule networks (CapsNets). This suggests that the unique architectural elements and feature-extraction strategies employed in the model rendered it exceptionally effective for the task at hand.

Table 5. ASD classification accuracy across various models.

References	Method	Pattern	Purpose	Accuracy (%)
Yakolli et al. [38]	CNN	FC and structural	Classification	88.0
Jiang et al. [39]	3D CNN	FC	Classification	72.46
Hao et al. [12]	Deep-broad learning	ROI	Classification	71.8
Husna et al. [40]	CNN	FC	Classification	87.0
Shi et al. [41]	Decision model	ROI	Classification	75.41
Niu et al. [42]	DANN	FC/ROI	Classification	73.2
Byeon et al. [16]	RNN	FC	Classification	74.54
Jiao et al. [15]	CapsNets	FC	Classification	71.0
Yin et al. [14]	DNN, AE	ROI	Classification	79.20
Anirudh et al. [13]	GCNN	ROI	Classification	70.86
Zhao et al. [11]	3D CNN	ROI	Classification	70.1
Guo et al. [43]	DNN	FC	Classification	86.36
Ktena et al. [44]	GCNN	ROI	Classification	62.9
The proposed model	CNN	FC	Classification	99.39

Expanding upon the model's notable performance metrics, the data preparation phase involves segmenting extracted three-dimensional fMRI volumes into two-dimensional slices across three anatomical planes, thereby preserving maximal information from the original 3D dataset. The model incorporates 3×3 convolution kernels, ReLU activation functions, and MaxPooling layers to effectively capture spatial features at various scales. This specialized feature-extraction methodology addresses shortcomings in extant models, thereby contributing to the model's enhanced performance. Furthermore, the model's consistently high accuracy across a pediatric age range of 6–18 years and diverse ASD severity levels substantiates its generalizability and diagnostic sensitivity. The employment of whole-brain resting-state fMRI data, without regional isolation, accentuates the model's capacity for identifying distributed neural patterns, thereby bolstering its robustness.

However, this study was not without limitations. The model's efficacy was principally validated within a school-aged demographic, thereby raising questions regarding its generalizability to younger pediatric populations, which is essential for early ASD diagnosis.

The acquisition of fMRI data from younger children with ASD presents challenges, due to sensory sensitivities and behavioral considerations. Consequently, the applications of alternative, child-friendly neuroimaging modalities, such as functional near-infrared spectroscopy (fNIRS) and electroencephalography (EEG), warrant further investigation. Additionally, the binary classification framework of the model fell short of capturing the nuanced heterogeneity of ASD symptoms and their varying degrees of severity. Moreover, eye-tracking methods, coupled with machine-learning techniques, offer another promising avenue for autism diagnosis, as exemplified by recent studies [45]. The binary classification nature of the model did not encapsulate the spectrum of ASD symptoms, which vary in severity and manifestation. The study also fell short in providing insights into the model's decision-making rationale, which is a crucial factor for clinical adoption. Furthermore, the computational resources requisite for training the model may not be ubiquitously available, thereby constituting a potential impediment to its broad-scale deployment. Another area of critical concern is that ethical considerations are integral to the deployment of AI methodologies, such as CNNs, in ASD diagnosis. Such concerns encompass data confidentiality, informed consent, and the risk of algorithmic bias, which may disproportionately impact specific demographics. Addressing these ethical dimensions is vital for the responsible application and clinical acceptance of AI-based diagnostic systems.

5. Conclusions

This investigation pioneered the development of a CNN model for the incipient diagnosis of ASD in pediatric cohorts, utilizing fMRI datasets. The study encompassed meticulous data preprocessing, bespoke CNN architecture training, performance evaluation manifesting unparalleled accuracy metrics exceeding 99%, and feature-map visualization that substantiated the model's adeptness in hierarchical feature learning. Collectively, the framework constituted a swift, objective, and remarkably precise computer-assisted diagnostic instrument for ASD screening, leveraging the synergies between fMRI data and advanced deep-learning algorithms.

Based on the investigation, the salient conclusions may be articulated as follows:

(1) The proffered CNN model exhibited extraordinary proficiency in differentiating ASD from typically developing (TD) subjects, attaining an accuracy metric of 99.39%, thereby markedly eclipsing prior scholarly endeavors. This underscored the model's efficacy in discerning discriminative features within the fMRI datasets.

(2) Scrutiny of the model's feature maps corroborated its capabilities in hierarchical feature extraction, with the more advanced layers serving a pivotal role in demarcating ASD from TD. This suggested that the model was adept at learning increasingly intricate and abstract data representations.

(3) The model furnished a rapid, unerring, and highly precise diagnostic apparatus for ASD screening and identification, with the potential to revolutionize conventional, subjective, and protracted diagnostic frameworks.

(4) The model's exceptional performance across a gamut of computational methodologies accentuated the efficacy of its unique architectural design and feature extraction paradigms for this specific classification task.

This investigation elucidated a CNN model with an extraordinary accuracy rate exceeding 99% in the classification of ASD, utilizing fMRI datasets, thereby significantly outperforming extant computational methodologies. While this study's findings are auspicious, the actualization of this technology's clinical utility mandates further scholarly pursuits in model validation across heterogeneous demographic subsets, the integration of multi-modal data, the interpretability of algorithmic decisions, resource-efficient deployment, classification granularity across autism subtypes and severity gradations, and prospective real-world evaluations. Rigorous endeavors to surmount these limitations and translational obstacles will be indispensable for metamorphosing this high-performing deep-learning framework from a technological novelty into a clinically deployable decision-

support mechanism, thereby enhancing early screening and personalized therapeutic interventions across the autism spectrum.

References

1. Kim, H.; Lim, C.-S.; Kaang, B.-K. Neuronal Mechanisms and Circuits Underlying Repetitive Behaviors in Mouse Models of Autism Spectrum Disorder. *Behav. Brain Funct.* **2016**, *12*, 3. [CrossRef] [PubMed]
2. Al Sagheer, T.; Haida, O.; Balbous, A.; Francheteau, M.; Matas, E.; Fernagut, P.-O.; Jaber, M. Motor Impairments Correlate with Social Deficits and Restricted Neuronal Loss in an Environmental Model of Autism. *Int. J. Neuropsychopharmacol.* **2018**, *21*, 871–882. [CrossRef] [PubMed]
3. Owens, D.G.C.; Johnstone, E.C. Precursors and Prodromata of Schizophrenia: Findings from the Edinburgh High Risk Study and Their Literature Context. *Psychol. Med.* **2006**, *36*, 1501–1514. [CrossRef] [PubMed]
4. Martinussen, M.; Friborg, O.; Schmierer, P.; Kaiser, S.; Øvergård, K.T.; Neunhoeffer, A.-L.; Martinsen, E.W.; Rosenvinge, J.H. The Comorbidity of Personality Disorders in Eating Disorders: A Meta-Analysis. *Eat. Weight. Disord.* **2017**, *22*, 201–209. [CrossRef]
5. Kaiser, M.D.; Hudac, C.M.; Shultz, S.; Lee, S.M.; Cheung, C.; Berken, A.M.; Deen, B.; Pitskel, N.B.; Sugrue, D.R.; Voos, A.C.; et al. Neural Signatures of Autism. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 21223–21228. [CrossRef]
6. Liu, S.; Liang, B.; Wang, S.; Li, B.; Pan, L.; Wang, S. NF-GAT: A Node Feature-Based Graph Attention Network for ASD Classification. *IEEE Open J. Eng. Med. Biol.* **2023**, 1–6. [CrossRef]
7. Singh, S.; Malhotra, D.; Mengi, M. TransLearning ASD: Detection of Autism Spectrum Disorder Using Domain Adaptation and Transfer Learning-Based Approach on RS-FMRI Data. *Artif. Intell. Commun. Technol.* **2023**, 863–871. [CrossRef]
8. El Nakieb, Y.; Ali, M.T.; Elnakib, A.; Shalaby, A.; Mahmoud, A.; Soliman, A.; Barnes, G.N.; El-Baz, A. Understanding the Role of Connectivity Dynamics of Resting-State Functional MRI in the Diagnosis of Autism Spectrum Disorder: A Comprehensive Study. *Bioengineering* **2023**, *10*, 56. [CrossRef]
9. Anderson, J.L.; Nielsen, J.A.; Froehlich, A.L.; DuBray, M.B.; Druzgal, T.J.; Cariello, A.N.; Cooperrider, J.R.; Zielinski, B.A.; Ravichandran, C.; Fletcher, P.T.; et al. Functional Connectivity Magnetic Resonance Imaging Classification of Autism. *Brain* **2011**, *134*, 3742–3754. [CrossRef]
10. Polyakova, S. The Pathophysiological Rationale for Personalized Metabolic Therapy of Asd. Promising Treatments. *Autism Dev. Disord.* **2019**, *17*, 55–70. [CrossRef]
11. Zhao, Y.; Dong, Q.; Zhang, S.; Zhang, W.; Chen, H.; Jiang, X.; Guo, L.; Hu, X.; Han, J.; Liu, T. Automatic Recognition of fMRI-Derived Functional Networks Using 3-D Convolutional Neural Networks. *IEEE Trans. Biomed. Eng.* **2018**, *65*, 1975–1984. [CrossRef]
12. Hao, X.; An, Q.; Li, J.; Min, H.; Guo, Y.; Yu, M.; Qin, J. Exploring High-Order Correlations with Deep-Broad Learning for Autism Spectrum Disorder Diagnosis. *Front. Neurosci.* **2022**, *16*, 1046268. [CrossRef]
13. Anirudh, R.; Thiagarajan, J.J. Bootstrapping Graph Convolutional Neural Networks for Autism Spectrum Disorder Classification. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; IEEE: New York, NY, USA, 2019; pp. 3197–3201.
14. Yin, W.; Mostafa, S.; Wu, F.-X. Diagnosis of Autism Spectrum Disorder Based on Functional Brain Networks with Deep Learning. *J. Comput. Biol.* **2021**, *28*, 146–165. [CrossRef] [PubMed]

15. Jiao, Z.; Li, H.; Fan, Y. Improving Diagnosis of Autism Spectrum Disorder and Disentangling Its Heterogeneous Functional Connectivity Patterns Using Capsule Networks. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; IEEE: New York, NY, USA, 2020; pp. 1331–1334.

16. Byeon, K.; Kwon, J.; Hong, J.; Park, H. Artificial Neural Network Inspired by Neuroimaging Connectivity: Application in Autism Spectrum Disorder. In Proceedings of the 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), Busan, Republic of Korea, 13–16 February 2020; IEEE: New York, NY, USA, 2020; pp. 575–578.

17. Di Martino, A.; Yan, C.-G.; Li, Q.; Denio, E.; Castellanos, F.X.; Alaerts, K.; Anderson, J.S.; Assaf, M.; Bookheimer, S.Y.; Dapretto, M.; et al. The Autism Brain Imaging Data Exchange: Towards a Large-Scale Evaluation of the Intrinsic Brain Architecture in Autism. *Mol. Psychiatry* **2013**, *19*, 659–667. [CrossRef] [PubMed]

18. Jeyapratap, A.S.; Parker, D.; Yerys, B.E.; Verma, R.; Osmanlioglu, Y. Analysis of Structural Network Topology Alterations in Autism and Its Relationship with Autism Severity. *bioRxiv* **2022**, 1–4. [CrossRef]

19. Dong, X.L.; Rekatsinas, T. Data Integration and Machine Learning. *Proc. Vldb Endow.* **2018**, *11*, 2094–2097. [CrossRef]

20. Rashid, M.; Singh, H.; Goyal, V. The Use of Machine Learning and Deep Learning Algorithms in Functional Magnetic Resonance Imaging—A Systematic Review. *Expert Syst.* **2020**, *37*, e12644. [CrossRef]

21. Chawla, P.; Rana, S.B.; Kaur, H.; Singh, K. Computer-Aided Diagnosis of Autism Spectrum Disorder From EEG Signals Using Deep Learning With FAWT and Multiscale Permutation Entropy Features. *Proc. Inst. Mech. Eng. Part H J. Eng. Med.* **2022**, *237*, 282–294. [CrossRef] [PubMed]

22. Bakkour, A.; Lewis-Peacock, J.A.; Poldrack, R.A.; Schonberg, T. Neural Mechanisms of Cue-Approach Training. *Neuroimage* **2017**, *151*, 92–104. [CrossRef]

23. Bryan, C.J.; Rudd, M.D. The Importance of Temporal Dynamics in the Transition From Suicidal Thought to Behavior. *Clinical Clin. Psychol. Sci. Pract.* **2016**, *23*, 21–25. [CrossRef]

24. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A Survey on Deep Learning in Medical Image Analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef] [PubMed]

25. Yang, Y.; Ge, Y.; Guo, L.; Wu, Q.; Peng, L.; Zhang, E.; Xie, J.; Li, Y.; Lin, T. Development and Validation of Two Artificial Intelligence Models for Diagnosing Benign, Pigmented Facial Skin Lesions. *Ski. Res. Technol.* **2020**, *27*, 74–79. [CrossRef] [PubMed]

26. Sarvamangala, D.; Kulkarni, R.V. Convolutional Neural Networks in Medical Image Understanding: A Survey. *Evol. Intell.* **2021**, *15*, 1–22. [CrossRef] [PubMed]

27. Yamins, D.L.K.; Hong, H.; Cadieu, C.F.; Solomon, E.A.; Seibert, D.; DiCarlo, J.J. Performance-Optimized Hierarchical Models Predict Neural Responses in Higher Visual Cortex. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 8619–8624. [CrossRef]

28. Xu, J.; Yu, X. Detection of Concrete Structural Defects Using Impact Echo Based on Deep Networks. *J. Test. Eval.* **2021**, *49*, 20190801. [CrossRef]

29. Xu, J.; Yu, X. Pavement Roughness Grade Recognition Based on One-Dimensional Residual Convolutional Neural Network. *Sensors* **2023**, *23*, 2271. [CrossRef]

30. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; IEEE: New York, NY, USA, 2015; pp. 1–9.

31. Huang, G.; Liu, Z.; Maaten, L.V.D.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [CrossRef]

32. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.

33. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.

34. Hu, K.; Zhang, Z.; Niu, X.; Zhang, Y.; Cao, C.; Xiao, F.; Gao, X. Retinal Vessel Segmentation of Color Fundus Images Using Multiscale Convolutional Neural Network with an Improved Cross-Entropy Loss Function. *Neurocomputing* **2018**, *309*, 179–191. [CrossRef]

35. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980. [CrossRef]

36. Simonyan, K. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556. [CrossRef]

37. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-Based Learning Applied to Document Recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]

38. Yakolli, N.; Anusha, V.; Khan, A.A.; Shubhashree, A.; Chatterjee, S. Enhancing the Diagnosis of Autism Spectrum Disorder Using Phenotypic, Structural, and Functional MRI Data. *Innov. Syst. Softw. Eng.* **2023**. in review.

39. Jiang, W.; Liu, S.; Zhang, H.; Sun, X.; Wang, S.-H.; Zhao, J.; Yan, J. CNNG: A Convolutional Neural Networks With Gated Recurrent Units for Autism Spectrum Disorder Classification. *Front. Aging Neurosci.* **2022**, *14*, 948704. [CrossRef] [PubMed]

40. Husna, R.N.S.; Syafeeza, A.R.; Hamid, N.A.; Wong, Y.C.; Raihan, R.A. Functional Magnetic Resonance Imaging for Autism Spectrum Disorder Detection Using Deep Learning. *J. Teknol.* **2021**, *83*, 45–52. [CrossRef]

41. Shi, C.; Xin, X.; Zhang, J. Domain Adaptation Using a Three-Way Decision Improves the Identification of Autism Patients from Multisite fMRI Data. *Brain Sci.* **2021**, *11*, 603. [CrossRef] [PubMed]

42. Niu, K.; Guo, J.; Pan, Y.; Gao, X.; Peng, X.; Li, N.; Li, H. Multichannel Deep Attention Neural Networks for the Classification of Autism Spectrum Disorder Using Neuroimaging and Personal Characteristic Data. *Complexity* **2020**, *2020*, 1357853. [CrossRef]

43. Guo, X.; Dominick, K.C.; Minai, A.A.; Li, H.; Erickson, C.A.; Lu, L.J. Diagnosing Autism Spectrum Disorder from Brain Resting-State Functional Connectivity Patterns Using a Deep Neural Network with a Novel Feature Selection Method. *Front. Neurosci.* **2017**, *11*, 460. [CrossRef]

44. Ktena, S.I.; Parisot, S.; Ferrante, E.; Rajchl, M.; Lee, M.; Glocker, B.; Rueckert, D. Distance Metric Learning Using Graph Convolutional Networks: Application to Functional Brain Networks. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2017*; Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2017; Volume 10433, pp. 469–477. ISBN 978-3-319-66181-0.

45. Cilia, F.; Carette, R.; Elbattah, M.; Dequen, G.; Guérin, J.-L.; Bosche, J.; Vandromme, L.; Le Driant, B. Computer-Aided Screening of Autism Spectrum Disorder: Eye-Tracking Study Using Data Visualization and Deep Learning. *JMIR Hum. Factors* **2021**, *8*, e27706. [CrossRef]