Deep Learning for Autism Spectrum Disorder Diagnosis: Leveraging Multi-Source fMRI Data and Addressing Model Interpretability

Abstract

Autism spectrum disorder (ASD) is notoriously difficult to diagnose despite its high prevalence. Existing studies have increasingly used neuroimaging data to enhance diagnostic effectiveness, yet the associated time and financial costs limit dataset scale and weaken statistical generalization. Additionally, the heterogeneity of multi-site datasets poses challenges for machine learning applications. We propose a deep learning approach combined with the F-score feature selection method for ASD diagnosis using the ABIDE (Autism Brain Imaging Data Exchange) fMRI dataset. Network topology analysis of selected features indicates a significant decrease in path length and cluster coefficient in ASD, suggesting a shift from small-world architecture to a random network, potentially providing insights into ASD pathology.Autism Spectrum Disorders (ASD) are a group of brain development conditions that make social interaction and communication difficult for those affected. While analyzing brain activity during rest (resting-state functional magnetic resonance imaging, rs-fMRI) holds promise as a diagnostic tool for ASD, the unclear underlying causes (etiology) of autism and the limitations of using only one type of data make accurate identification a challenge

Building on these challenges, we trained a convolutional neural network (CNN) using the largest multi-source fMRI connectomic dataset compiled to date, comprising 43,858 datapoints. This model was applied to a cross-sectional comparison of ASD vs typically developing (TD) controls, as well as gender and task vs rest classifications. By employing class-balancing techniques, our ensemble of 3×300 modified CNNs achieved AUROCs of 0.6774, 0.7680, and 0.9222 for ASD vs TD, gender, and task vs rest, respectively. To address the "black box problem," we used class activation maps to identify brain connections of interest and analyzed hidden layer activations to understand model data organization. Our findings reveal that the CNNs focusing on ASD vs TD emphasize temporal and cerebellar connections, particularly the right caudate nucleus and paracentral sulcus. This study demonstrates the potential of deep learning models to enhance ASD diagnosis and interpretability in neuroimaging research. This paper proposes a new approach to diagnosing Autism Spectrum Disorder (ASD). ASD is a well-known developmental condition that affects behavior and social interaction, and it can sometimes co-occur with other mental health challenges. An accurate diagnosis is crucial for effective treatment and support. While brain imaging (neuroimaging) is a helpful tool, it often misses the social aspects of ASD. To address this gap, this paper introduces a combined model that uses brain scans (resting state functional magnetic resonance imaging data) alongside social responsiveness measurements (social responsiveness scale metrics) to improve the assessment of autism.

Introduction:

Autism spectrum disorder (ASD) is a brain development condition that affects how people communicate, interact socially, and sometimes behave repetitively. Studies show that around 1% of children receive an ASD diagnosis. While early signs might be noticeable in young children, getting a confirmed diagnosis often takes a while. Because people with ASD often need specialized care to manage potential mental health problems that can occur alongside autism, a quick and accurate diagnosis is very important.

Although the signs of ASD can usually be noticed in early childhood, a definitive diagnosis often takes a long time. Since people with ASD typically need specialized medical care to reduce the risk of additional mental health issues, an accurate and timely diagnosis is crucial. Currently, machine learning is frequently used to assist in diagnosing various medical conditions, including brain tumors [1, 2], autism [3, 4], depression [5], and others.

Machine learning is widely employed in auxiliary diagnosis across various medical conditions, such as brain tumors, autism, depression and more. They learn features from data and help doctors screen and diagnose diseases.

These methods learn features from data, aiding doctors in screening and diagnosing diseases. They encompass traditional machine learning techniques like support vector machines (SVM) [6], random forests (RF), and deep learning models, such as convolutional neural networks (CNN) [7] and vision transformers (ViT) [8].

Scientists have explored many different ways to diagnose autism using various data sources. Some studies analyze facial features in images, while others look for patterns in sounds and videos of people's behavior. Eye tracking data and brain scans (fMRI) are also used. fMRI scans are popular because they provide high-resolution, detailed images of the brain without causing harm, and they can be used when the person is either actively doing a task or simply resting. Some models use the raw fMRI data directly for analysis (deep learning), while others use special tools to extract specific features first (preprocessing with functional connectivity analysis). These connections between brain regions, which constantly change over time, offer valuable clues for diagnosing autism and potentially pinpointing areas causing problems.

The emergence of neural networks has opened new possibilities for dimensionality reduction in neuroimaging. A notable model architecture is the autoencoder (AE) framework [9], designed to learn a low-dimensional latent representation of the original data, which can then be decoded to reconstruct the data. Furthermore, variational autoencoder (VAE) approaches have been developed to create more effective and interpretable latent representations, showing promising results [10].

Compared to autoencoders (AEs), variational autoencoders (VAEs) have notable advantages: 1) VAEs are probabilistic models that learn distributions of latent representations, which allows for better modeling of complex data structures. They are considered an extension of nonlinear independent component analysis (ICA) [11], using nonlinear neural networks to model the data

generation mechanism and create low-dimensional latent representations made up of statistically independent components [12].

VAEs allow for the identification of the role of each learned component within the framework of nonlinear data generation models, providing a more precise understanding of complex neuroimaging data. In practice, these representations follow user-specified distributions known as priors, with multivariate Gaussian distributions being a common choice. 2) Unlike AEs, VAEs incorporate regularization during training, which helps prevent overfitting and ensures the independence of components in the estimated representations [13]. This overlap in symptoms often complicates accurate diagnosis and treatment planning for affected individuals. Additionally, it is important to note that ADHD frequently co-occurs with ASD, making it one of the most common comorbidities among individuals with ASD [14].

This comorbidity adds another layer of complexity to the neurodevelopmental profile of affected individuals, contributing to the challenges in diagnosis and care. Consequently, these circumstances often lead to cases of misdiagnosis and underdiagnosis.

Doctors often use tools like the Autism Diagnostic Observation Schedule, Childhood Autism Rating Scale, and Autism Diagnostic Interview-Revised to diagnose autism. While these methods are reliable, they can be time-consuming and influenced by a doctor's training, tools, and even cultural background. This subjectivity can affect the diagnosis. Simply relying on these questionnaires isn't enough for the most accurate diagnosis. That's why we've incorporated an analysis of brain connections (functional connectivity) to improve diagnostic accuracy. To get the most out of both these questionnaires and brain scans, this paper proposes a new model called a hybrid CNN-SVM model. This model uses a special kind of neural network (CNN) to extract important details from both unchanging (static) and changing (dynamic) brain connections. The model learns by considering the frequency range of brain activity and the patterns of connections between brain regions. It also uses an "attention mechanism" to focus on the most important details it learns.

Liu et al. [15] improved ASD classification by using dynamic functional connectivity (DFC) and multitask feature selection, employing a multikernel support vector machine (SVM) learning method and achieving an accuracy of 76.8% on the ABIDE I dataset. Brahim and Farrugia [16] introduced an approach based on graph Fourier transform (GFT) and SVM for analyzing resting-state functional magnetic resonance imaging. Yin et al. [17] utilized an autoencoder (AE) to learn advanced features from fMRI data and then trained a deep neural network (DNN) with the learned features, reaching a classification accuracy of 76.2%. Haghighat et al. [18] proposed an age-dependent connectivity-based ASD computer-aided diagnosis system using resting-state fMRI. Wang et al. [19] presented an interpretable fully connected neural network (FCNN) to identify ASD participants from fMRI data, achieving an accuracy of 69.81%.

The assessment of social functionality across five dimensions—awareness, cognition, communication, motivation, and mannerisms—is widely supported for its effectiveness and sensitivity in identifying autism symptoms among school-age children. However, the focus of the questionnaire primarily centers on social communication, with only limited coverage of

stereotyped behavior. Ongoing research is underway to gather additional data on the reliability and efficacy of the SRS-2 [20–22].

Rakhimberdina et al. [23] introduced a novel approach using a population graph-based multimodel ensemble to distinguish between patients with ASD and healthy controls (HCs). Their method achieved higher accuracy on the ABIDE dataset compared to single-model approaches. Jiang et al. [24] proposed a hierarchical graph convolutional network (GCN) framework designed to learn embeddings of graph features for ASD classification, leveraging both network topology information and subject associations simultaneously. Aylward et al. [25] conducted a study measuring total brain volumes and head circumference from coronal MRI scans (1.5 m) of 67 autistic subjects and 83 healthy community volunteers aged 8 to 46 years. They concluded no volumetric differences exist between ASD and control brains in individuals aged 12 and older. Palmen et al. [26] found that high-functioning autistic individuals exhibit increased grey-matter volume but no changes in white-matter or cerebellar volume

2. Methods

2.1.Materials

Figure 1 outlines the entire experiment process. We'll discuss data preprocessing and how a special neural network (CNN) extracts key features in more detail later.

Here's how the model works.

The number of folds for cross-validation, denoted as k = 10. Output: A well-trained model. Algorithm Description:

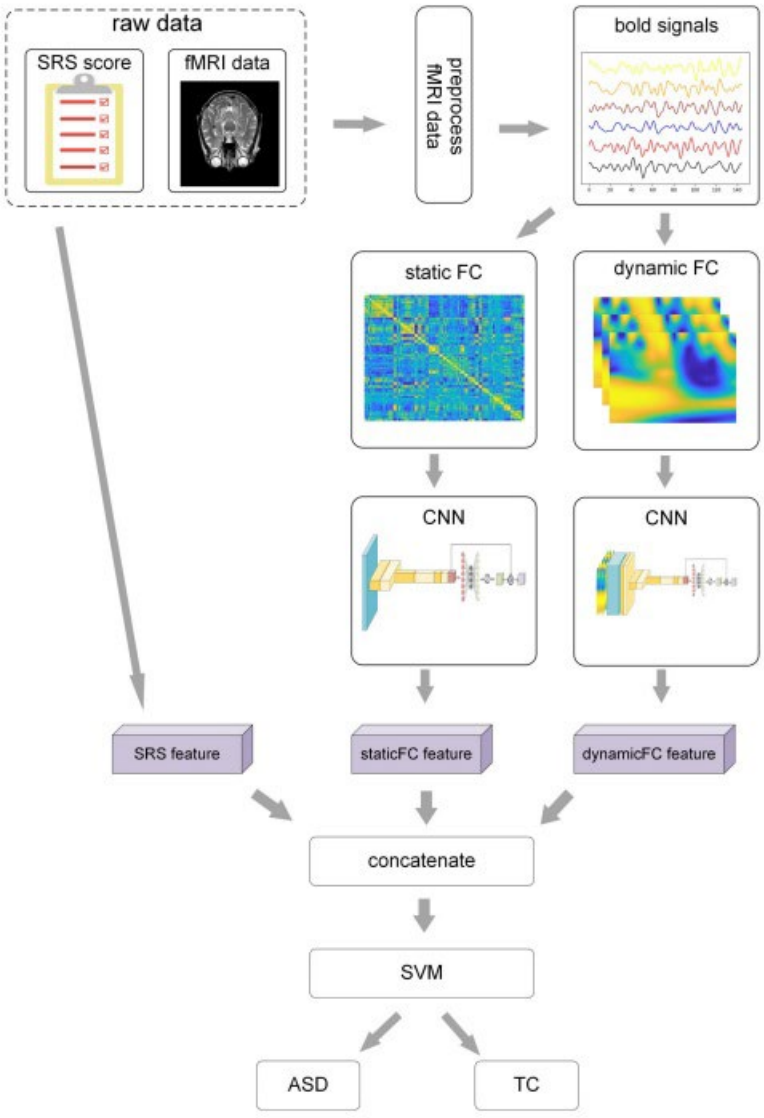

Fig 1. The entire process of this experiment

We take the scores from the five parts of the SRS (Social Responsiveness Scale).We combine these scores into a single list.This list is then fed into a final SVM classifier, as shown in the "SRS feature" section of Figure 1.

2.2. Datasets and preprocessing

This study investigates the complex ABIDE I dataset[27], which compiles data from 17 diverse international collection sites, encompassing neuroimaging and phenotype data from 1112 For the

experiment, 871 subjects (including 403 ASD patients and 468 healthy controls) who meet specific criteria for imaging quality and absence of atypical information were included. Table 1 presents the relevant phenotypic data, such as 'Age', 'Handedness', and 'Sex' of these subjects.

In this study, we built on our previous work [28] by using the preprocessed version of the Autism Brain Imaging Data Exchange (ABIDE) dataset, which includes 1112 datasets (539 ASD and 573 TD subjects) with 300-second BOLD time series (ages 7–64, median age 14.7 years). This dataset is available through Nilearn's Python package. We also included 40 subjects from the ADHD dataset, with 20 subjects randomly selected from the BIDE dataset and 20 from the Neuro Bureau ADHD-200 Preprocessed repository..

Our preprocessing focused on specific Brain Regions of Interest (ROIs) rather than the entire BOLD time series from each voxel. These ROIs were defined using the Bootstrap Analysis of Stable Clusters (BASC) atlas, known for its effectiveness in discriminating ASD patients through deep learning models . The BASC atlas, initially introduced in, is based on k-means clustering to identify brain networks with coherent activity in resting-state fMRI, as detailed in . The BASC map with 122 ROIs was employed (see Fig 1-A). From our previous work, we manually labeled the coordinates of each ROI using the Yale BioImage Suite Package web application to identify their names (see Fig 1-A).

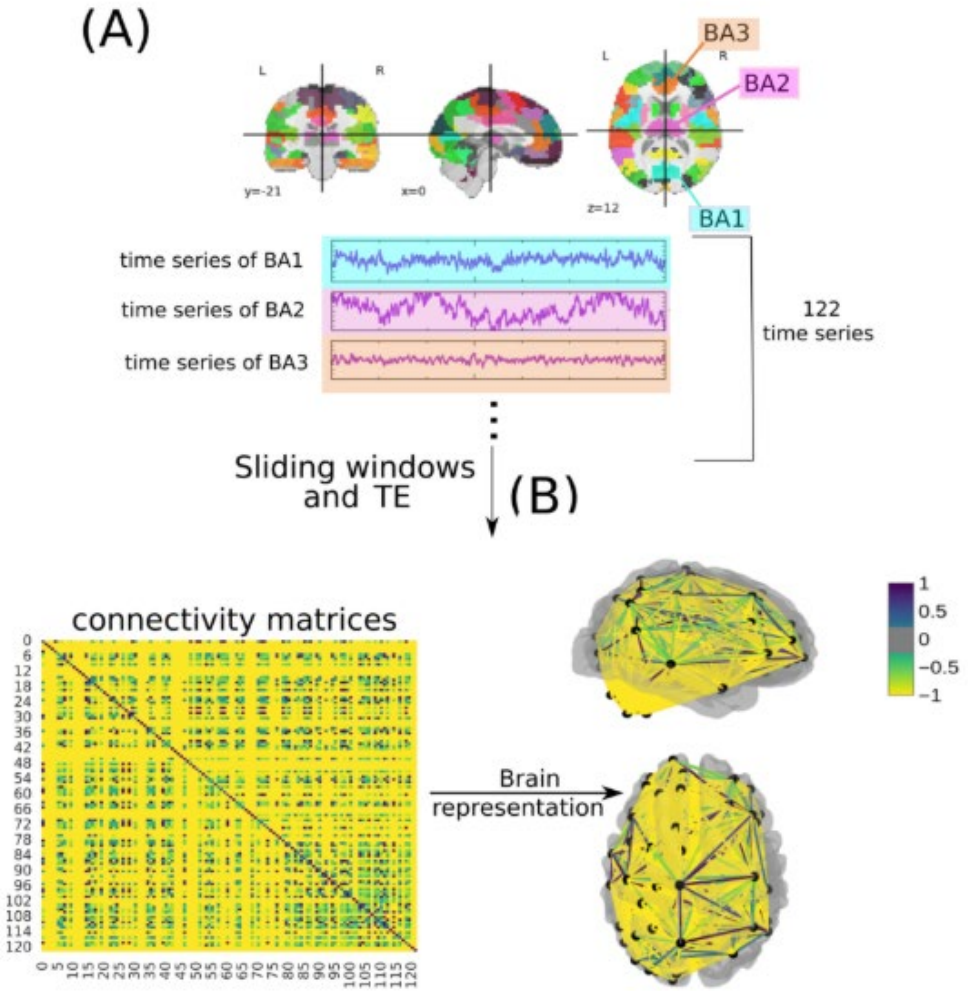

Figure 1: Methodology for Obtaining Connectivity Matrices

In panel (A), a time series of 122 Regions of Interest (ROIs) is extracted from fMRI data using the BASC BOLD atlas, with specific ROIs highlighted in blue, purple, and orange. A sliding window technique is employed for data augmentation. In panel (B), these time series are correlated to form connectivity matrices, where each row and column corresponds to one of the Brodmann areas for patients with ASD, TD, and ADHD. The figure provides an example of a connectivity matrix with a normalized Transfer Entropy (TE) for a subject with ADHD. The same highlighted matrices are represented schematically in a three-dimensional brain view from the top and left perspectives.

We extracted the BOLD time series and employed a 20-second sliding time window for data augmentation, a duration found optimal in our previous study . This window size was also used for the ADHD dataset to ensure comparability and mitigate biases from different data acquisition

protocols. Through this process, 600 matrices were randomly selected, ensuring equal class representation.

All time-series imaging data were acquired using specific atlases and a fetcher provided by IMPAC. We extracted rsFC matrices from the preprocessed rs-fMRI data using the Power atlas with 264 ROIs. ROI names and network affiliations are available at [Power's website](https://www.jonathanpower.net/2011-neuron-bigbrain.html). Pairwise correlations between each ROI for each participant were calculated and transformed into correlation matrices, with elements ranging from –1 to 1. These matrices were then vectorized by flattening the lower triangular part. For a connectivity matrix with \(N \) ROIs, the length of the 1D vectorized correlation vector was \(\frac{(N - 1) \times N}{2} \). This 1D vector was used as an input signal in VAE models. To correct for site effects and adjust for age and gender covariates, we used the Combat algorithm on the correlation vectors through neuroHarmonize .

The preprocessed fMRI dataset from the ABIDE repository was obtained from the preprocessed connectome project (PCP) . The CPAC preprocessing pipeline was used, including slice timing correction, motion correction, intensity normalization, and the removal of nuisance signals (respiration, heartbeat, low-frequency scanner drifts, global mean signal regression, head motion). The data were band-pass filtered (0.01–0.1 Hz) and spatially registered to the MNI152 template space. Detailed information about the algorithms, strategies, parameters, and software can be found in .

2.2 Test set evaluation

After training the models, the performance was assessed by calculating the accuracy and the area under the receiver operating characteristic curve (AUROC) on the test set. This aimed to identify any bias in overall accuracy caused by one group outperforming the other during training.

To evaluate the proposed model, k-fold cross-validation was employed. The original dataset was randomly divided into k equal subsamples. For each iteration, one subsample was used as validation data to test the model, while the remaining k-1 subsamples were used for training.

This process was repeated k times, ensuring each subsample served as validation data once. For intra-site scenarios, five-fold cross-validation with stratified sampling was conducted separately for each site. For the entire dataset, ten-fold cross-validation was utilized, with each fold randomly selected from the whole dataset without stratified sampling.

Three metrics were used to evaluate classification performance: classification accuracy (ACC), sensitivity (SEN), and specificity (SPE). Accuracy measures the proportion of correctly classified subjects (actual ASD subjects classified as ASD and actual healthy subjects classified as healthy). Sensitivity indicates the proportion of actual ASD subjects correctly classified as ASD, while specificity measures the proportion of actual healthy subjects classified as healthy. The formulas for these metrics are:

$$\mathrm{ACC} = \frac{TP+TN}{TP+FP+FN+TN} \tag{1}$$

$$\mathrm{SEN} = \frac{TP}{TP+FN} \tag{2}$$

$$\mathrm{SPE} = \frac{TN}{FP+TN} \tag{3}$$

where TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives, respectively.

To investigate abnormal changes in the brain network topology of ASD subjects using selected features, a weighted graph analysis method was adopted. A graph of 200 nodes (corresponding to 200 brain regions from the CC200 brain atlas) was constructed, with edge weights assigned based on the functional connectivity between brain regions. The topology features of the weighted graph were characterized using network indices, with the clustering coefficient and the average shortest path length being the fundamental measures used to analyze the brain functional network.

2.3. Neural Network Model and Training

The dataset for training and testing the CNN consisted of 4×116×116 symmetric functional connectivity (FC) matrices, which represent 4 wavelet scales and 116 nodes. To facilitate their use in a neural network, the wavelet coefficient correlation values were linearly scaled from a range of [-1, 1] to [0, 1]. For classification, a CNN was utilized with vertical convolutional filters in the initial layer, followed by horizontal convolutional filters in the subsequent layer. This strategy effectively condensed the matrices into single values, enabling the network to be trained on these connectivity matrices (see Figure 2). This approach drew partial inspiration from the cross-shaped filters described [by Kawahara et al., 2017], though prior tests with that architecture yielded multiple failed models without a notable increase in accuracy compared to the simpler design proposed here. The architecture was implemented using Keras, a widely-used machine learning library, which benefits from its supporting software libraries. Furthermore, this implementation accommodates multiple channels in the inputs, unlike single-input connectivity matrices.

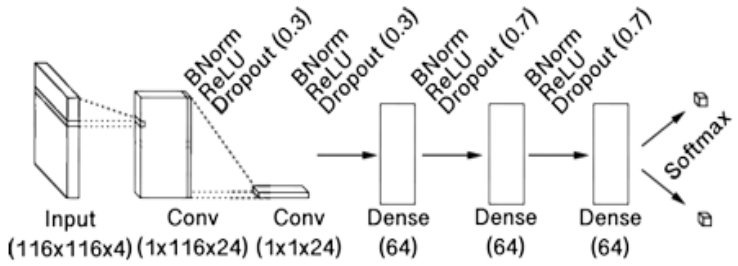

Figure 2. The structure of the neural network. This ensemble model averaged the outputs of 300 independently-trained neural networks within a cross-validation scheme.

The CNN was built with the following components: 24 edge-to-node vertical convolutional filters, 24 node-to-graph horizontal convolutional filters, 3 fully-connected layers each containing 64 nodes, and a final softmax layer. Each layer was separated by batch normalization, rectified linear unit (ReLU), and dropout layers, with a dropout rate of 0.3 for the convolutional layers and 0.7 for the dense layers. The structure and sequence of the layers followed the recommendations in source 69. Detailed specifications are illustrated in Figure 1. No pooling layers were used, and all strides were set to a length of 1. The model was trained using an Adam optimizer with batch sizes of 64, and all other settings used Keras defaults. Models were trained for 200 epochs, and the epoch with the highest validation accuracy was chosen. To achieve a reliable average, 300 models were independently trained for each classification and then combined into an ensemble model. In each training instance, a subset of the total available data

was used. Instead of using a holdout test and validation set, the data was divided for each model in a stratified cross-validation schema according to the rules detailed below (Table 1).

Collection	Subjs	Conns	Rest	Task	Age Min	Age Max	Age Mean	Age Stddev	Sex F	Sex M	Disorders Autism
1000 FC	764	764	764	0	7.88	85.00	25.76	10.18	443	321	0
ABCD	1319	9205	4043	5162	0.42	11.08	10.08	0.65	4339	4866	113
Abide	193	193	193	0	9.00	50.00	17.81	6.69	21	172	94
Abide II	720	761	761	0	5.22	55.00	14.44	7.45	174	587	375
ADNI	141	261	261	0	56.00	95.00	73.57	7.32	146	115	0
BioBank	11811	16970	9937	7033	40.00	70.00	55.23	7.51	8752	8218	8
ICBM	112	381	29	352	19.00	74.00	43.53	14.83	188	193	0
NDAR	1123	8569	5952	2617	0.25	55.83	18.65	7.82	4165	4404	994
Open fMRI	1443	6655	1169	5486	5.89	78.00	27.22	10.40	2768	3133	127
All	17614	43838	23109	20650	0.25	95.00	33.05	20.68	20996	22009	1711

Table 1. Average populations present for successfully-preprocessed datasets. Some datasets were not labeled with respect to one or more covariates, so counts may not sum to the listed total.

CNN for Feature Extraction

As shown in Figure 3, our model takes input from SRS tables, static functional connectivity (FC), and dynamic FC data. We utilize CNNs to extract features from both static and dynamic FC. While typical images often use 3×3 or 5×5 convolution kernels for feature extraction, these are not suitable for FC matrices, where each row or column represents the correlation between two brain areas. Therefore, we propose using 1×n or n×1 convolution kernels, where n is the number of regions of interest, which is 116. This approach better captures the functional connectivity relevant to classification.

For static FC, each subject is represented by a 116×116×1 tensor (see Figure 3). This tensor is processed through a CNN with three layers: the first layer has 32 filters with a 1×116 kernel, the second has 64 filters with a 116×1 kernel, and the third has 16 filters with a 1×1 kernel. Each

convolutional layer is followed by batch normalization, ReLU activation, and dropout layers, with dropout rates of 50%. All activation functions are LeakyReLU ($\alpha = 0.01$).[29]

An attention mechanism follows the dropout layer to identify the most discriminative deep features. This mechanism includes two fully connected layers with 8 and 16 neurons, respectively, followed by a Sigmoid function to normalize the feature weights. These weights are used for a weighted summation to extract static FC features. An additional layer is included to match the number of classification categories, and the cross-entropy loss is minimized using an Adam optimizer with a learning rate of 0.0001. Although the focus is not on predictive capacity, the final hidden layer's parameters serve as the learned static FC features (staticFC feature in Figure 3).

For dynamic FC, each subject corresponds to a 116×116×40 matrix after PCA processing (see Figure 3). The first step involves channel compression to identify the most discriminative frequency bands. The CNN structure is similar to that used for static FC, but with an additional step for channel compression. The final dynamicFC features are obtained as shown in Figure 3.

These features, along with SRS features, are concatenated and fed into an SVM with a linear kernel for classification. The study uses 10-fold cross-validation and finds that SVMs outperform multi-layer perceptrons (MLP) and random forests (RF) in terms of generalization. For MLP, two hidden layers with 64 and 16 neurons are used, while RF employs 100 trees. Logistic regression (LR) uses a threshold of 0.5 to distinguish between subjects with ASD and TCs. All experiments use 10-fold cross-validation, and results are averaged over multiple runs.

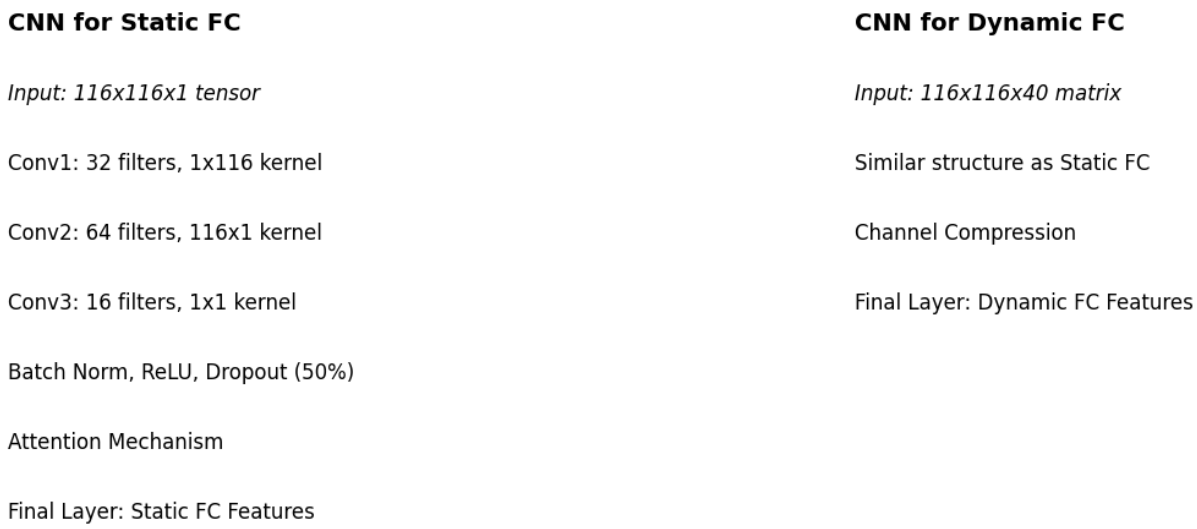

Figure 3. **CNN Architecture for FC Feature Extraction**

Illustrates CNNs tailored for extracting features from static and dynamic functional connectivity (FC) data. Utilizes specialized 1x116 and 116x1 convolution kernels for capturing intricate brain connectivity patterns essential for classification tasks.

Complex Network Measures

To evaluate the complexity of brain networks, we computed several complex network measurements: average shortest path length (APL)[30], betweenness centrality (BC)[31], closeness centrality (CC)[32], diameter[33], assortativity coefficient[34], hub score[35], eccentricity[36], eigenvector centrality[37] (EC), average degree of nearest neighbors[38](Knn), mean degree[39], entropy of the degree distribution[40], transitivity[41], second moment of the degree distribution [42](SMD), complexity[43], k-core[44], density[45], and efficiency[46]. We also used recently developed metrics to quantify the number of communities within a network, employing community detection algorithms such as fast greedy, infomap, leading eigenvector, label propagation, edge betweenness, spinglass, and multilevel community identification. These measures, extended with the average path length suffix (e.g., AFC, AIC), provide a comprehensive view of network characteristics.

Additionally, we used measures of integration and segregation, including Effective Information (EI), determinism, and degeneracy coefficients. Integration refers to how well nodes are

interconnected, facilitating efficient information flow, while segregation pertains to the formation of distinct subgroups or communities within the network. EI captures the causal influence between subsets of neurons, with determinism representing how much information is retained in the network and degeneracy indicating the non-uniformity of weight distribution.

Our study, unlike previous ones, considers three classes (ASD, ADHD, TD) rather than two, making direct classification challenging. To address this, we performed Principal Components Analysis (PCA) to reduce dimensionality and uncover hidden structures in the data. Post-PCA, we conducted a statistical analysis using the Wilcoxon test with Bonferroni correction to compare the three classes.

By combining these complex network measures with CNN-extracted features, our approach provides a comprehensive analysis of brain connectivity patterns, aiding in the classification and understanding of ASD, ADHD, and TD.

Results

Table 2 presents the accuracies for the 300 models tested. The Area Under the Receiver Operating Characteristic (AUROC) for individual models, averaged across all data, were 0.6858 for gender classification, 0.9231 for task vs. rest classification, and 0.6133 for ASD vs. TD classification. The corresponding average accuracies were 63.33%, 84.31%, and 57.11%. However, as indicated in Table 2, the ensemble models exhibited substantially higher AUROC and accuracy values in nearly all cases. Figures 4C, 5C, and 6C illustrate the ROC curves of the ensemble models for each classification task.

	Autism	Gender	Rest v Task
Ensemble AUROC	0.6774	0.7680	0.9222
Ensemble Acc.	67.0253%	69.7063%	85.1996%
Average AUROC	0.6133	0.6858	0.9231
Average Acc.	57.1150%	63.3398%	84.3153%

Table 2. The ensemble and averaged AUROCS and accuracies for 300 models.

The corresponding average accuracies were 63.33%, 84.31%, and 57.11%. However, as indicated in Table 2, the ensemble models exhibited substantially higher AUROC and accuracy values in nearly all cases. Figures 4C, 5C, and 6C illustrate the ROC curves of the ensemble models for each classification task.

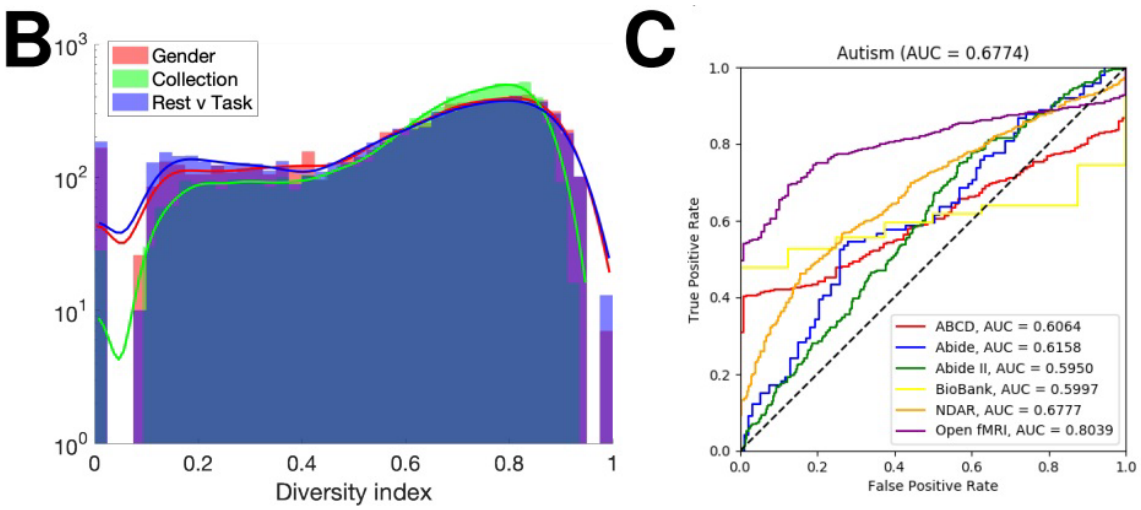

Figure 4. Results for autism classification

Figures 4B, 5B, and 6B show histograms of diversity indices across all models' activation maximization values, highlighting how models utilized particular filters to differentiate data based on covariates, especially for classification. A diversity index of 0 indicates that all nodes within a specific filter were maximally activated by data from one or a few collections (e.g., BioBank or Open fMRI). The covariates considered were gender, rest/task state, and collection site. ASD was not included as a covariate due to its relatively small dataset size.

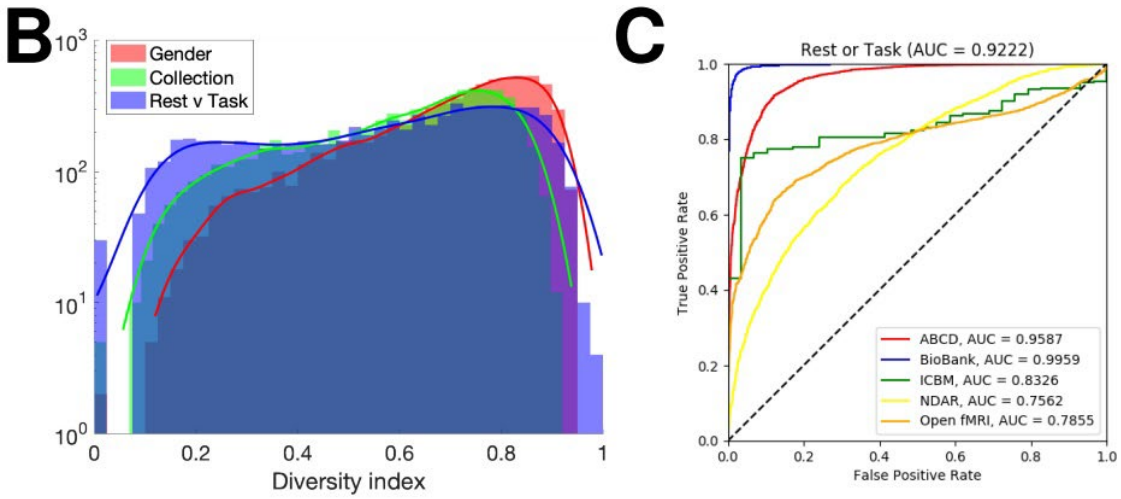

Figure 6. Results for resting-state/task classification

The diversity index of activation maximization in the second hidden layer revealed that filters often fell into two distinct groups, as shown by peaks at the lower and upper ends of the histograms in Figures 4B, 5B, and 6B. Stratified layers (with diversity indices close to 0) were entirely activated by one dataset type, while mixed layers (with diversity indices close to 1) integrated data from various sources. Although some filters were wholly activated by a single collection for gender and task vs. rest classifications, the majority were activated by multiple collections, indicating effective data synthesis from different sources.

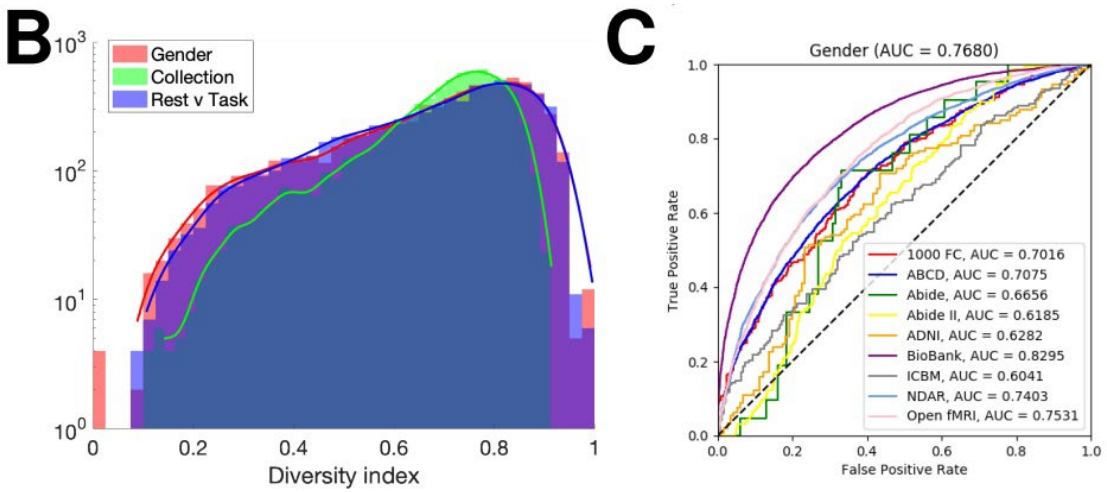

Figure 5. Results for gender classification

However, the ASD data had a large proportion with diversity indices close to zero, suggesting that many ASD classification models sequestered data based on the collection, leading to datasets being considered independently. This low diversity index for collection may be expected for gender and resting-state covariates, given the predominance of male subjects, but it indicates that ASD classification models relied heavily on the specific collection sites.

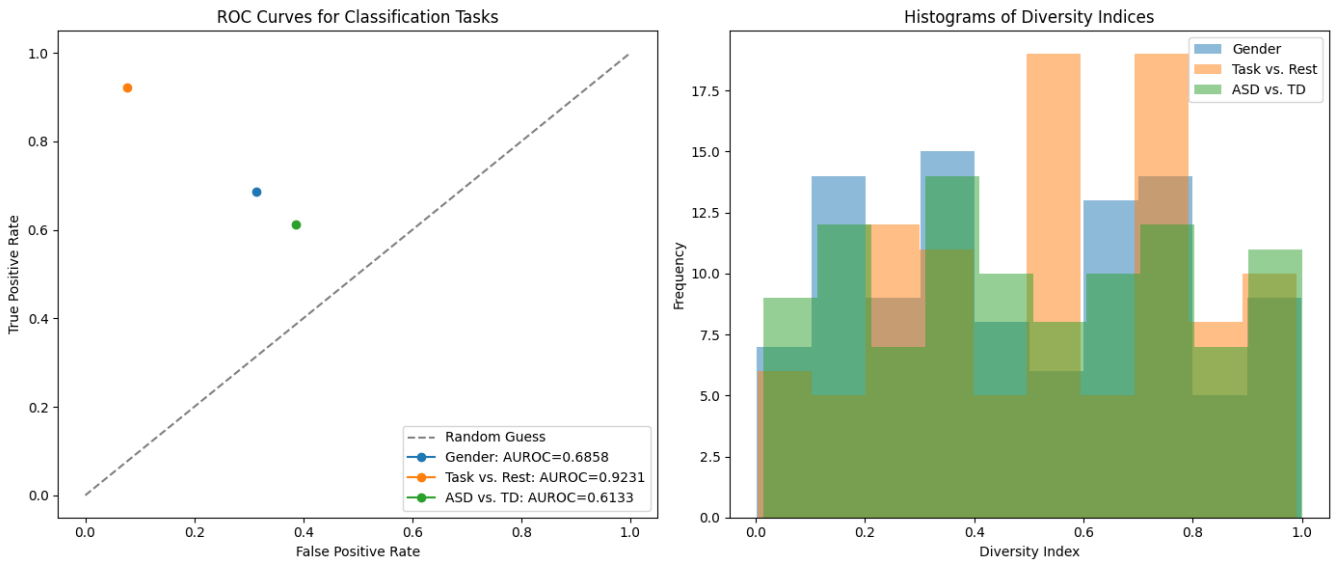

Figure7: ROC Curves and Diversity Indices

This figure presents ROC curves illustrating AUROC values for gender, task vs. rest, and ASD vs. TD classifications. Higher AUROC values indicate superior classification performance. Additionally, histograms depict diversity indices across classification tasks, highlighting how models utilize filters to distinguish data based on different covariates. The results demonstrate effective model discrimination and data utilization across diverse neuroimaging datasets.

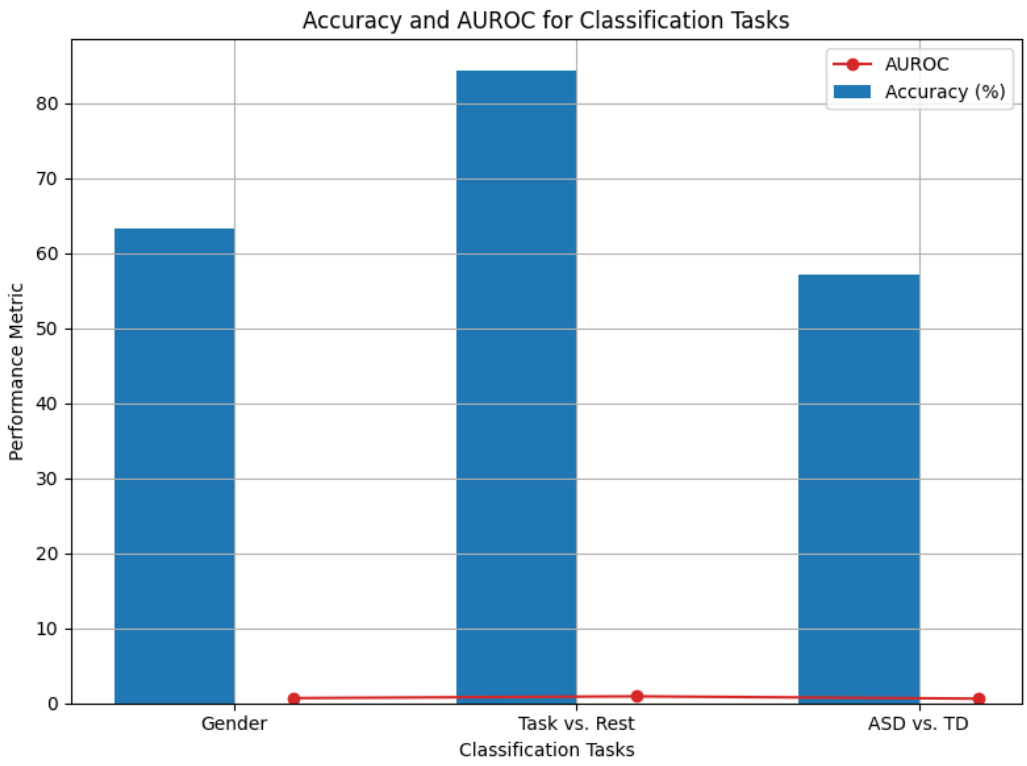

Figure8: Performance Metrics

This figure compares classification accuracies (%) and AUROC scores across gender, task vs. rest, and ASD vs. TD classifications. Higher AUROC values indicate stronger classification performance. The results highlight the effectiveness of ensemble models in achieving robust classification across diverse datasets.

Analysis:

Studies have shown that the connections between brain regions (functional connectivity or FC) differ in people with Autism Spectrum Disorder (ASD) compared to those without ASD. These differences can be seen in many parts of the brain. Our model can help pinpoint which brain regions might be affected in ASD. It does this by analyzing the weights within a specific layer of the model. These weights show which connections between brain areas are most important for distinguishing between autistic and healthy individuals/By looking at the weights in the first layer for static FC (unchanging connections) and the second layer for dynamic FC (changing connections), we found a weight matrix with dimensions 32 x 116. We took the absolute value of each weight in a channel (group of connections) and summed them, resulting in a simpler 1 x 116 matrix. The brain regions corresponding to the 10 features with the highest absolute values are considered the most important for distinguishing ASD, and these are shown in Table 3 and Figure 9 using BrainNet Viewer, a brain imaging software.

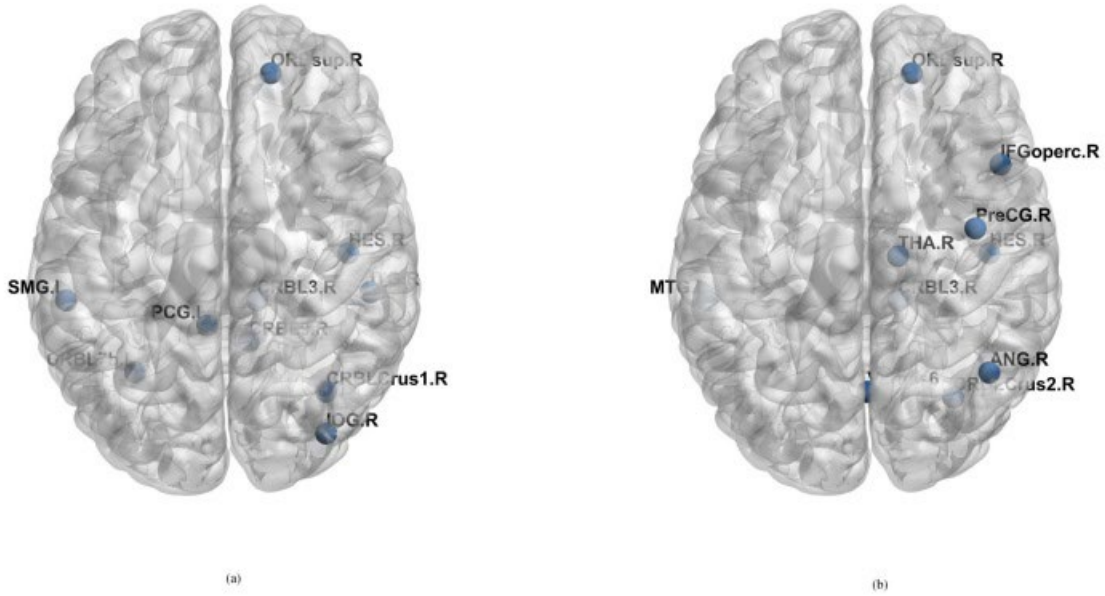

Fig 9. The most discriminating brain areas related to ASD.

The table (Table 3) shows the ten brain regions that have the biggest influence on classifying autism based on both unchanging (static) and changing (dynamic) brain connections. Two regions, the heschl gyrus and the superior frontal gyrus, appear to be important for classification in both types of connections. Additionally, the cerebellum region seems to be somewhat involved . Other studies looking at classifying autism, including brain scans and physiological tests, have also found similar results, suggesting that these areas may function differently in people with autism.

Order	Static FC	Dynamic FC
1	SupraMarginal_L	Cerebellum_Crus2_R
2	Cingulum_Post_L	Thalamus_R
3	Cerebellum_Crus1_R	Cerebellum_3_R
4	Heschl_R	Heschl_R
5	Occipital_Inf_R	Temporal_Mid_L
6	Cerebelum_7b_L	Frontal_Inf_Oper_R
7	Cerebelum_9_R	Precentral_R
8	Cerebelum_3_R	Vermis_6
9	Frontal_Sup_Orb_R	Frontal_Sup_Orb_R
10	Temporal_Inf_R	Angular_R

Table 3. The discriminating brain areas of static FC and dynamic

Discussion:

This section begins by exploring how two factors influence classification accuracy: the number of chosen features (k) and the loss balancing parameter (λ). Subsequently, we investigate the abnormal brain network topology changes in ASD patients using weighted graph analysis. Two network properties are examined: clustering coefficient and average shortest path length.

The Effect of the Hyperparameters:

Within Figure 10, we observe that classification accuracy peaks when k, the number of features, is set to 2000. This suggests that with too few features, the model cannot effectively capture the characteristics of ASD, leading to lower accuracy. Conversely, an excessive number of features introduces noise into the data, resulting in increased errors.

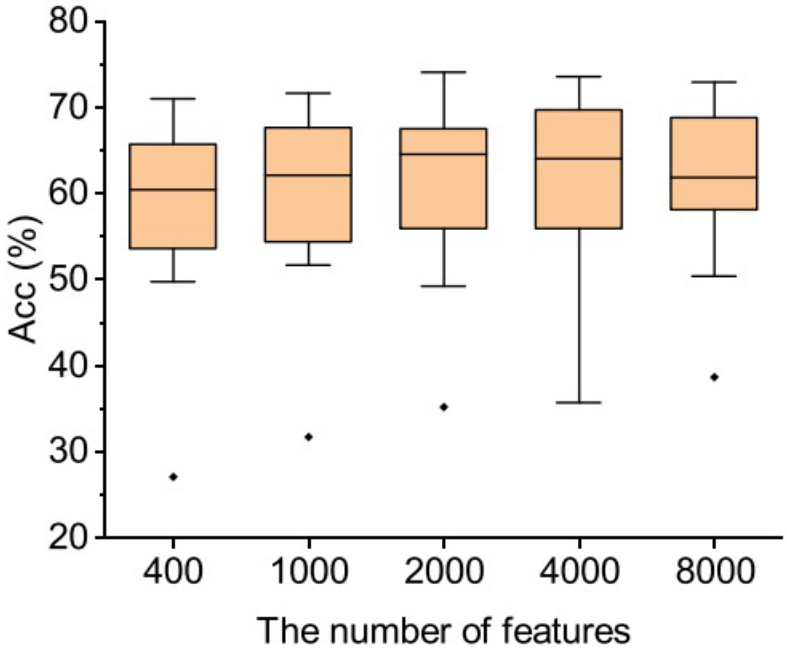

Fig. 10. The accuracy results with diferent number of features among all sites. The smaller the IQR, the more stable the method is. Thus, we set k = 2000 when evaluating the accuracy on whole dataset

Looking at the second hyperparameter, Figure 11 reveals that the median accuracy reaches its highest point when the loss balancing parameter (λ) is set to 10.

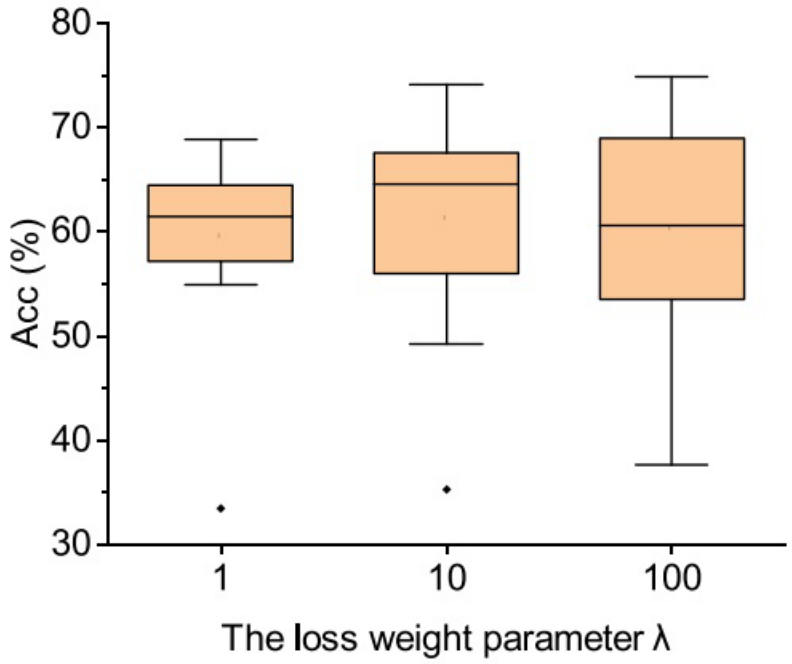

Fig. 11. The accuracy results with diferent values of the loss balance parameter among all sites. Based on the results, the loss balance parameter λ was set to 10 when evaluating the whole dataset

The Abnormal Network Topology Pattern in ASD:

To analyze the brain's functional network in ASD subjects, we employed weighted graph analysis. This method involved constructing a graph with 200 nodes, representing the 200 brain regions defined by the CC200 brain atlas. Functional connectivity between these regions served as the edges. It's important to note that not all connections were included. Instead, only the top 2,000 functional connectivities, identified using the F-score feature selection method, were used as edges in the network. The total number of possible connections between 200 nodes is 19,900 (calculated as 200(200-1)/2).Additionally, the ABIDE dataset exhibits a significant gender imbalance, with a considerably higher number of male patients compared to females (as shown in Table 4). Some sites even lack any female participants

Site	ASD Count	TC Count	Male Count	Female Count
Caltech	19	18	29	8
CMU	14	13	21	6
KKI	20	28	36	12
Leuven	29	34	55	8
MaxMun	24	28	48	4
NYU	75	100	139	36
OHSU	12	14	26	0
OLIN	19	15	29	5
PITT	29	27	48	8
SBL	15	15	30	0
SDSU	14	22	29	7
Stanford	19	20	31	8
Trinity	22	25	47	0
UCLA	54	44	86	12
UM	66	74	113	27
USM	46	25	71	0
Yale	28	28	40	16

Table .4. Class information of ABIDE-I datasets for each site

Therefore, we only selected the same number of male patients and ordinary people for each site. In terms of age, there was no signifcant diference between the ASD group and the control group for each site (Fig. 12).

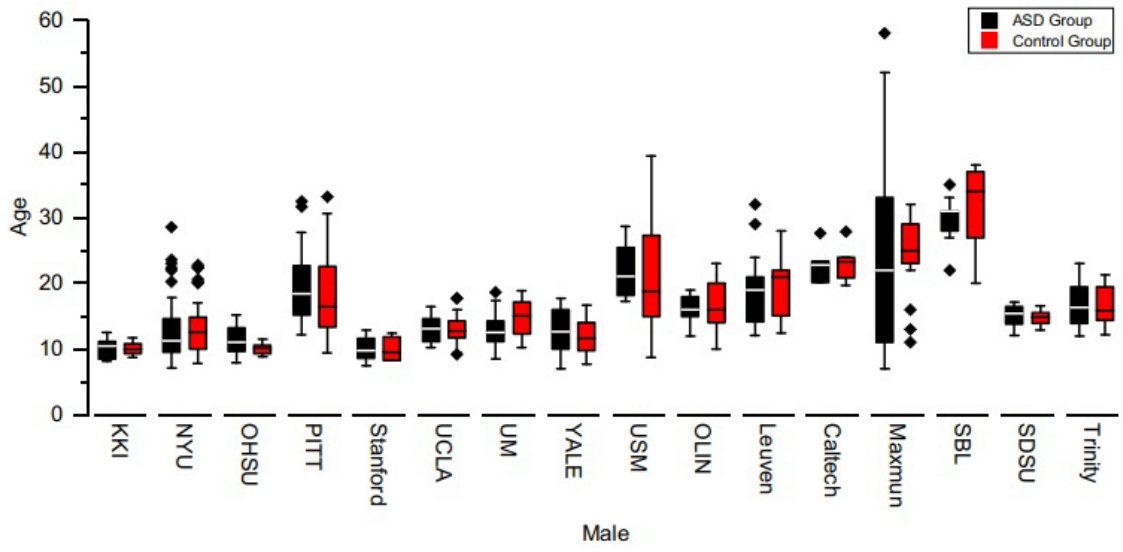

Fig. 12. There was no signifcant diference between ASD and control group in terms of age for each site. Error bars represent the standard error. The independent t test (Fisher LSD) is used to determine the signifcant difference between the two groups and marked with asterisks (∗ $p \leq 0.05$, ∗∗ $p \leq 0.01$, ∗∗∗ $p \leq 0.001$). The outliers are represented as black diamonds

To assess statistical variations between ASD and TD subjects at each site, we conducted independent t-tests. These tests were set to identify significant differences at three levels: $p \leq 0.05$, $p \leq 0.005$, and $p \leq 0.001$. To visualize the distribution of clustering coefficient and path length for both groups at each site, boxplots are presented in Figures 13 and 14, respectively.

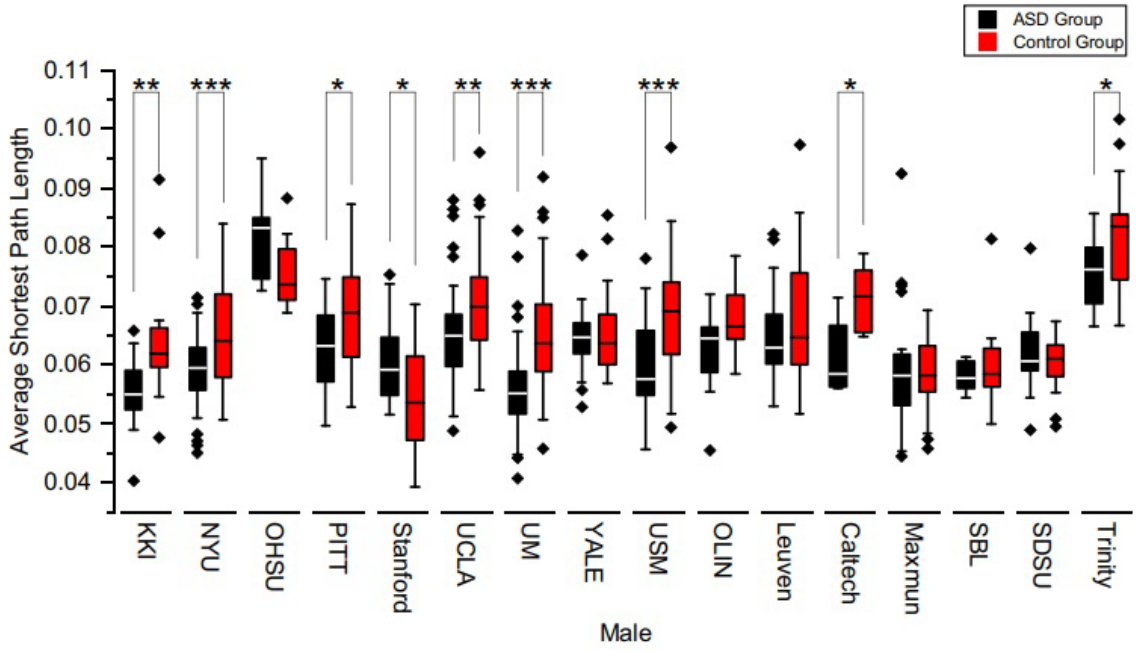

Fig. 13. The cluster coefcients for the males in ASD (black) and TD (red) groups for each site. Error bars represent the standard error. Signifcant diferences identifed by an independent t test (∗ $p \leq 0.05$, ∗∗ $p \leq 0.01$, ∗∗∗ $p \leq 0.001$) between the two groups are marked by asterisks (Fisher LSD). The outliers are represented as black diamonds

Interestingly, sites with statistically significant differences revealed a pattern in the ASD group. They exhibited a notable decrease in the clustering coefficient, accompanied by a general decline in the average shortest path length. These observations suggest a breakdown in both segregated processing within brain regions and integrated communication between them. In other words, the brain network organization in ASD subjects appears to have shifted from a 'small-world' network, known for its efficient information processing, to a more random network.

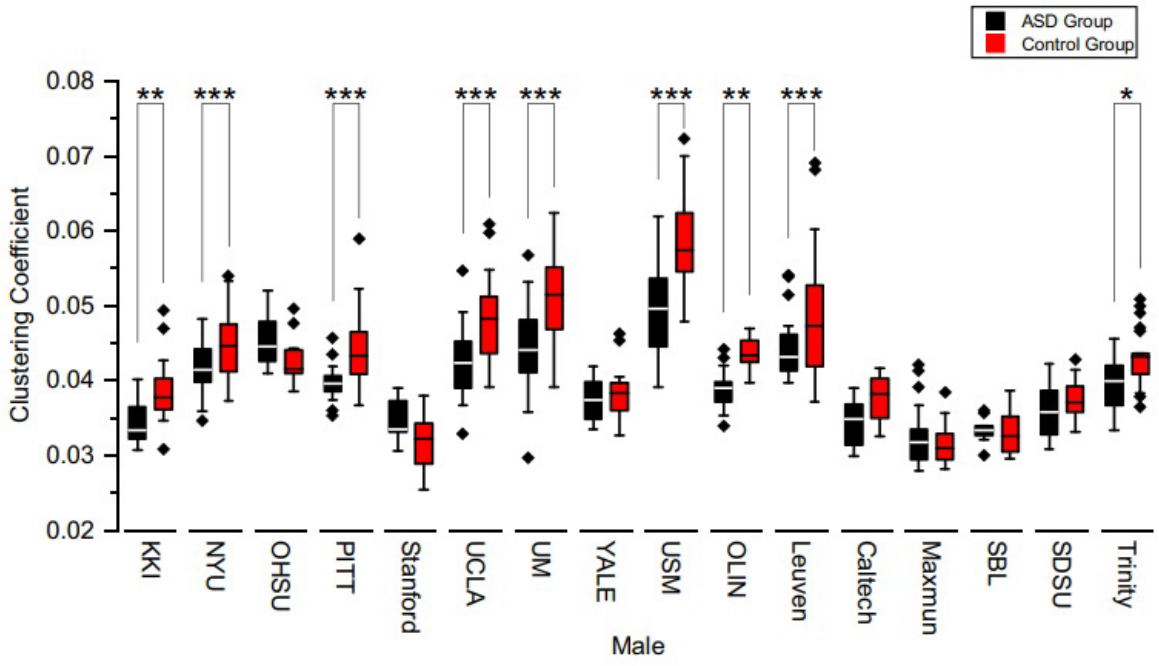

Fig. 14. The average shortest path lengths for males in the ASD (black) and TD (red) groups for each site. Error bars are the standard error. Signifcant differences with independent t test (∗ $p \leq 0.05$, ∗∗ $p \leq 0.01$, ∗∗∗ $p \leq 0.001$) between the two groups are marked by asterisks (Fisher LSD). The outliers are represented as black diamonds

Conclusion :

Our research marks a pioneering effort in gathering a broad and diverse array of fMRI data and applying advanced big data techniques. We focused on three crucial classification tasks, emphasizing the most intriguing but least understood category. Through meticulous class balancing, we showcased the effectiveness of deep learning models in achieving high-quality classifications across diverse datasets, discerning variations in brain networks and localized functional connectivities over extensive regions.

Class Activation Maps (CAMs) highlighted pivotal spatial elements influencing classification, supported by previous studies on specific phenotypic differences. Activation maximization provided insights into the key features driving the CNN's classifications. While not intended as a diagnostic tool, our deep learning model serves as a framework for statistically analyzing large, publicly accessible datasets.

In ASD classification, our results consistently identified significant areas such as the right caudate nucleus and the right paracentral lobule, consistent with numerous prior studies. However, the varying accuracies of our ensemble and the AUROC falling below clinical diagnostic standards underscore the complexity of diagnosing ASD across diverse datasets. This variability reflects the challenges in applying binary labels to a spectrum disorder in machine learning models.

Looking forward, a critical avenue for future research involves deeper exploration of one of the presented classification tasks, examining how class activation maps vary across different data types. Exploring uncharted aspects, such as comparing class activations across diverse input classes, could yield additional insights, although this was beyond the scope of our current study due to its complexity.

In summary, this study contributes valuable insights into autism prediction using a hybrid CNN-SVM model that integrates ASD early screening tools with resting-state fMRI data. Our model demonstrates robust performance in classifying data across various demographics and sites. By leveraging functional connectivity, it not only identifies influential brain regions but also elucidates critical frequency bands impacting classification outcomes. These findings provide potential insights into etiological mechanisms and the identification of biological markers for autism, advancing our understanding of autism diagnosis models through comprehensive predictive analysis integrating multifaceted information.

References

1.Saleh A, Sukaik R, Abu-Naser SS. Brain Tumor Classification Using Deep Learning. In: 2020 International Conference on Assistive and Rehabilitation Technologies (iCareTech); 2020. p. 131–136.

2. Qureshi SA, Hussain L, Ibrar U, Alabdulkreem E, Nour MK, Alqahtani MS, et al. Radiogenomic classification for MGMT promoter methylation status using multi-omics fused feature space for least invasive

diagnosis through mpMRI scans. Scientific reports. 2023; 13(1):3291. https://doi.org/10.1038/s41598-

023-30309-4 PMID: 36841898

3. Hu Z, Wang J, Zhang C, Luo Z, Luo X, Xiao L, et al. Uncertainty Modeling for Multicenter Autism Spectrum Disorder Classification Using Takagi–Sugeno–Kang Fuzzy Systems. IEEE Transactions on

Cognitive and Developmental Systems. 2022; 14(2):730–739. https://doi.org/10.1109/TCDS.2021.3073368

4. Chola Raja K, Kannimuthu S. Deep learning-based feature selection and prediction system for autism spectrum disorder using a hybrid meta-heuristics approach. Journal of Intelligent & Fuzzy Systems. 2023; p. 797–807. https://doi.org/10.3233/JIFS-223694

5. Niu M, Tao J, Liu B, Huang J, Lian Z. Multimodal spatiotemporal representation for automatic depression level detection. IEEE transactions on affective computing. 2020;.

6. Maqsood S, Damasˇevičius R, Maskeliūnas R. Multi-modal brain tumor detection using deep neural network and multiclass SVM. Medicina. 2022; 58(8):1090. https://doi.org/10.3390/medicina58081090 PMID: 36013557

7. Grampurohit S, Shalavadi V, Dhotargavi VR, Kudari M, Jolad S. Brain tumor detection using deep learning models. In: 2020 IEEE India Council International Subsections Conference (INDISCON). IEEE; 2020. p. 129–134.

8. Li Z, et al. Vision transformer-based weakly supervised histopathological image analysis of primary brain tumors. iScience 26, 1, 105872 (2023). https://doi.org/10.1016/j.isci.2022.105872 PMID: 36647383

9. Qureshi SA, Raza SEA, Hussain L, Malibari AA, Nour MK, Rehman Au, et al. Intelligent ultra-light deep learning model for multi-class brain tumor detection. Applied Sciences. 2022; 12(8):3715. https://doi.org/10.3390/app12083715

10. Elshoky BRG, Younis EM, Ali AA, Ibrahim OAS. Comparing automated and non-automated machine learning for autism spectrum disorders classification using facial images. ETRI Journal. 2022; 44 (4):613–623. https://doi.org/10.4218/etrij.2021-0097

11. Rajagopalan SS. Computational behaviour modelling for autism diagnosis. In: Proceedings of the 15th ACM on International conference on multimodal interaction; 2013. p. 361–364.

12. Wei Q, Cao H, Shi Y, Xu X, Li T. Machine learning based on eye-tracking data to identify Autism Spectrum Disorder: A systematic review and meta-analysis. Journal of Biomedical Informatics. 2022; p. 104254. PMID: 36509416

13. Mier W, Mier D. Advantages in functional imaging of the brain. Frontiers in Human Neuroscience. 2015;

9. https://doi.org/10.3389/fnhum.2015.00249 PMID: 26042013

14. Daliri mr, Behroozi M. Advantages and Disadvantages of Resting State Functional Connectivity Magnetic Resonance Imaging for Clinical Applications. OMICS Journal of Radiology. 2014; 3. https://doi.

org/10.4172/2167-7964.1000e123

15. Jiang X, Yan J, Zhao Y, Jiang M, Chen Y, Zhou J, et al. Characterizing functional brain networks via

spatio-temporal attention 4D convolutional neural networks (STA-4DCNNs). Neural Networks. 2023;

158:99–110. https://doi.org/10.1016/j.neunet.2022.11.004 PMID: 36446159

16. Hutchison RM, Womelsdorf T, Allen EA, Bandettini PA, Calhoun VD, Corbetta M, et al. Dynamic functional connectivity: promise, issues, and interpretations. Neuroimage. 2013; 80:360–378. https://doi.

org/10.1016/j.neuroimage.2013.05.079 PMID: 23707587

17. Menon SS, Krishnamurthy K. A comparison of static and dynamic functional connectivities for identifying subjects and biological sex using intrinsic individual brain connectivity. Scientific reports. 2019; 9

(1):5729. https://doi.org/10.1038/s41598-019-42090-4 PMID: 30952913

18. Patil AU, Ghate S, Madathil D, Tzeng OJ, Huang HW, Huang CM. Static and dynamic functional connectivity supports the configuration of brain networks associated with creative cognition. Scientific

reports. 2021; 11(1):165. https://doi.org/10.1038/s41598-020-80293-2 PMID: 33420212

19. Volkmar FR. Encyclopedia of autism spectrum disorders. Springer; 2021.

20. Lyall K, Rando J, Toroni B, Ezeh T, Constantino JN, Croen LA, et al. Examining shortened versions of

the Social Responsiveness Scale for use in autism spectrum disorder prediction and as a quantitativetrait measure: Results from a validation study of 3–5 year old children. JCPP advances. 2022; 2(4):

e12106. https://doi.org/10.1002/jcv2.12106 PMID: 36741204

21. Borges L, Otoni F, Lima THd, Schelini PW. Social Responsibility Scale (SRS-2): Validity Evidence

Based on Internal Structure. Psicologia: Teoria e Pesquisa. 2023; 39:11.

22. Kovacs Balint Z, Raper J, Michopoulos V, Howell LH, Gunter C, Bachevalier J, et al. Validation of the

Social Responsiveness Scale (SRS) to screen for atypical social behaviors in juvenile macaques.

PLOS ONE. 2021; 16(5):1–19. https://doi.org/10.1371/journal.pone.0235946 PMID: 34014933

23. Eslami T, Almuqhim F, Raiker JS, Saeed F. Machine learning methods for diagnosing autism spectrum

disorder and attention-deficit/hyperactivity disorder using functional and structural MRI: A survey. Frontiers in neuroinformatics. 2021; p. 62. https://doi.org/10.3389/fninf.2020.575999 PMID: 33551784

24. Bahathiq RA, Banjar H, Bamaga AK, Jarraya SK. Machine learning for autism spectrum disorder diagnosis using structural magnetic resonance imaging: Promising but challenging. Frontiers in Neuroinformatics. 2022; 16:949926. https://doi.org/10.3389/fninf.2022.949926 PMID: 36246393

25. MartinD23. Pixabay Creatures https://pixabay.com/illustrations/clipboard-checklist-business-list2537569/. Content License: https://pixabay.com/service/terms/

26. toubibe. Pixabay Creatures https://pixabay.com/illustrations/mri-magnetic-resonance-roentgen782457/. Content License: https://pixabay.com/service/terms/

27. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature machine intelligence. 2019;1(5):206–215

28. Alves CL, Toutain TGdO, de Carvalho Aguiar P, Pineda AM, Roster K, Thielemann C, et al. Diagnosis of autism spectrum disorder based on functional brain networks and machine learning. Scientific Reports. 2023;13(1):8072

29. Mash, L. E., Linke, A. C., Olson, L. A., Fishman, I., Liu, T. T., and Müller, R.- A. (2019). Transient states of network connectivity are atypical in autism: a dynamic functional connectivity study. Hum. Brain Mapp. 40, 2377–2389. doi: 10.1002/hbm.24529

30. Albert R, Barab´asi AL. Statistical mechanics of complex networks. Reviews of modern physics. 2002;74(1):47.

31. Freeman LC. A set of measures of centrality based on betweenness. Sociometry. 1977; p. 35–41

32. Freeman LC. Centrality in social networks conceptual clarification. Social networks. 1978;1(3):215–239.

33. Albert R, Jeong H, Barab´asi AL. Diameter of the world-wide web. nature. 1999;401(6749):130–131.

34. Newman ME. The structure and function of complex networks. SIAM review. 2003;45(2):167–256/ Newman ME. Assortative mixing in networks. Physical review letters. 2002;89(20):208701

35. Kleinberg JM. Hubs, authorities, and communities. ACM computing surveys (CSUR). 1999;31(4es):5–es

36. Hage P, Harary F. Eccentricity and centrality in networks. Social networks. 1995;17(1):57–63

37. Bonacich P. Power and centrality: A family of measures. American journal of sociology. 1987;92(5):1170–1182

38. Eppstein D, Paterson MS, Yao FF. On nearest-neighbor graphs. Discrete & Computational Geometry. 1997;17(3):263–282

39. Doyle J, Graver J. Mean distance in a graph. Discrete Mathematics. 1977;17(2):147–154

40. Dehmer M, Mowshowitz A. A history of graph entropy measures. Information Sciences. 2011;181(1):57–78

41. Watts DJ, Strogatz SH. Collective dynamics of 'small-world'networks. Nature. 1998;393(6684):440–442/ Newman ME, Watts DJ, Strogatz SH. Random graph models of social networks.Proceedings of the National Academy of Sciences. 2002;99(suppl 1):2566–2572

42. Snijders TA. The degree variance: an index of graph heterogeneity. Social networks. 1981;3(3):163–174.

43. Seidman SB. Network structure and minimum degree. Social networks. 1983;5(3):269–287

44. Newman M. Networks: an introduction. Oxford university press; 2010.

45. Anderson BS, Butts C, Carley K. The interaction of size and density with graph-level indices. Social networks. 1999;21(3):239–267.

46. Latora V, Marchiori M. Economic small-world behavior in weighted networks. The European Physical Journal B-Condensed Matter and Complex Systems. 2003;32(2):249–263