ASD-GResTM: Deep Learning Framework for ASD classification using Gramian Angular Field

Fahad Almuqhim and Fahad Saeed
Knight Foundation School of Computing and Information Sciences (KFSCIS)
Florida International University (FIU)
Miami, FL, USA
Email: {falmu027@fiu.edu, fsaeed@fiu.edu}

2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 979-8-3503-3748-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/BIBM58861.2023.10385743

***Abstract*—Autism Spectrum Disorder (ASD) is a heterogeneous disorder in children, and the current clinical diagnosis is accomplished using behavioral, cognitive, developmental, and language metrics. These clinical metrics can be imperfect measures as they are subject to high test-retest variability, and are influenced by assessment factors such as environment, social structure, or comorbid disorders. Advances in neuroimaging coupled with machine-learning provides an opportunity to develop methods that are more quantifiable, and reliable than existing clinical techniques. In this paper, we design and develop a deep-learning model that operates on functional magnetic resonance imaging (fMRI) data, and can classify between ASD and neurotypical brains. We introduce a novel strategy to transform time-series data extracted from fMRI signals into Gramian Angular Field (GAF) while locking in the temporal and spatial patterns in the data. Our motivation is to design and develop a novel framework that could encode the time-series, acquired from fMRI data, into images that can be used by deep-learning architectures that have been successful in computer vision. In our proposed framework called *ASD-GResTM*, we used a Convolutional Neural Network (CNN) to extract useful features from GAF images. We then used a Long Short-Term Memory (LSTM) layer to learn the activities between the regions. Finally, the output representations of the last LSTM layer are applied to a single-layer perceptron (SPL) to get the final classification. Our extensive experimentation demonstrates high accuracy across 4 centers, and outperforms state-of-the-art models on two centers with an increase in the accuracy of 17.58% and 6.7%, respectively as compared to the state of the art. Our model achieved the maximum accuracy of 81.78% with high degree of sensitivity and specificity. All training, validation, and testing was accomplished using openly available ABIDE-I benchmarking dataset.**

***Index Terms*—ASD, GAF, ResNet, LSTM, deep learning**

I. Introduction

Autism spectrum disorder (ASD) is a heterogeneous disorder, and the current diagnosis relies on behavioral observations in different settings (e.g., home and school), which might result in misdiagnosis [1] or delayed diagnosis [2]. Reliable and fast diagnosis of ASD remains challenging, and may lack specificity, and sensitivity [3] or the rigor required for other diseases such as diabetes. Currently, the ASD diagnosis is performed by a clinician with tools such as Autism Diagnostic Interview-Revised (ADI-R) [4] – a standardized care-giver interview, and other tools such as Autism Diagnostic Schedule (ADOS-2) [5] which is semi-structured observations for individuals (mostly kids) suspected of ASD. In addition to being a time-consuming and laborious process, the diagnosis is unreliable and inconsistent when performed in different settings or different health practitioners [1]. Recognizing the benefits of early diagnosis and intervention [6], there is a pressing demand for *biomarkers* that can facilitate timely identification of ASD [7].

Resting state functional magnetic resonance imaging (rs-fMRI) – is a noninvasive and fast technique to record functional activities in brain networks. In this data modality, the brain is represented by small cubic voxels where each voxel registers blood-oxygen-level-dependent (BOLD) volumes in the brain. Since fMRI can capture fine-grained data, the brain is parcellated into regions of interest (ROIs), and the BOLD volume of each region is the average brain activity (correlated with the brain oxygen level) of all voxels in that region. At the end of such an imaging process, the fMRI scan consists of a time series of BOLD volumes for each ROI. In recent years, there has been an increasing interest in using fMRI data for classifying mental disorders such as ASD, and Attention-deficit/hyperactivity disorder (ADHD) ([8] [9]) using biomarkers extracted using advance computational methods. While multiple machine-learning models have been proposed uptill date [10], [11], relatively little work has explored feature extraction and learning for time series data from fMRI scans, and current deep learning architectures. Our motivation is to design and develop a novel framework that could encode the time-series, acquired from fMRI data, into images that can be used by deep-learning architectures that have been successful in computer vision. Ideally such deep-learning structures will learn features and identify specific ASD and health biomarkers leading to effective diagnostic classification method. We assert that melfrequency cepstral coefficients (MFCCs) or perceptual linear predictive coefficient (PLPs) used for acoustic or speech data will not be useful in transforming and extracting the biomarkers from fMRI data.

In this paper, we designed and developed a comprehensive machine-learning pipeline that transforms the fMRI data into

This material is based upon work supported by the National Institutes of Health (NIH) grant number R01GM134384 and partly by National Science Foundation (NSF) grant NSF TI-2213951. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health (NIH) or the National Science Foundation (NSF).

Gramian Angular Field (GAF) to encode the time series of fMRI data to images which are then fed into our model. The GAF uses polar coordinates to encode each time series of each ROI as an image, and these transformed images are passed to a Residual Network (ResNet) model to extract features specific to ASD. The output from ResNet is then fed into the Bidirectional Long Short-Term Memory (BiLSTM) layer to extract the temporal and spatial patterns, and correlations between different ROIs. The learned vector space in BiLSTM is then used by a single perceptron layer (SPL) for final classification. We report our classification results using openly available ABIDE-I benchmark, and demonstrate that our classification performance is competitive with the state of the art ASD-DiagNet [12] model. Our experimentation also demonstrates that our method works better for shorter fMRI data sets (which leads to shorter time series data) as compared to traditional CNN based methods. In this paper, we provide an in-depth analysis in terms of the duality between time series from fMRI data and generated GAF images used for our machine-learning training and validation. Our extensive experimentation demonstrates high accuracy across 4 centers, and outperforms state-of-the-art models on two centers with an increase in the accuracy of 17.58% and 6.7%, respectively as compared to the state of the art. Our model achieved the maximum accuracy of 81.78% with high degree of sensitivity and specificity. The structure of the paper starts by showing related work in Section II. Section IV describes the proposed GAF algorithm. In Section V, we present the proposed framework. Section VI presents the results of the ASD-GResTM framework. Section VII presents conclusions and future work.

II. Related work

In recent years, numerous published papers have utilized fMRI data from ABIDE-I to identify biomarkers related to ASD brains vs. neurotypical (NT) brains ([13][14][15]). Most published research utilize Pearson's correlations to calculate the correlation of individual regions. As a result, the input vector for each subject is a representation of the brain's functional connectivity. Given that there are N regions within the brain, the resulting input vector will have a size of $n \times (n-1)/2$. In such scenarios, various methods employ deep learning techniques for the purpose of feature reduction. Examples of these techniques include the utilization of Autoencoders (AE) and Sparse Autoencoders (SAE). Guo et al. [13] developed a deep neural network (DNN) consisting of several stacked sparse autoencoders for feature extraction followed by a softmax regression (SR) layer for classification. They achieved 86.36% accuracy using only the data collected from the UM site from the ABIDE-I dataset. Zhang et al. [15] proposed an F-score feature selection technique for feature reduction, and a deep learning model for classification, and achieved an average of 64.53% on intra-site datasets. Heinsfeld et al.[16] implemented a method involving two stacked denoising autoencoders to extract features, which were then fed into a multi-layer classifier. This approach was employed separately on each site, yielding an average accuracy of 52%. ASD-DiagNet[12] adopted a tied autoencoder, where the same weights were utilized for both the encoding and decoding stages. A singular perceptron layer was employed for classification, resulting in an average accuracy of 63.8% across different sites. In the case of ASD-SAENet [17], a sparse autoencoder was employed for feature extraction, followed by a deep neural network for classification. The reported average accuracy across each site stood at 64.6%. In this paper, we transform the time series data extracted from fMRI data into spatio- and temporal-invariant images which can then be used for training and validation of computer vision models such as ResNet [18]. We assert that transforming the time series into images – such that such transformation can capture the biomarkers specific to the ASD – will lead to development of machine-learning models that have been successfully used in various image processing and computer vision solutions. However, the image conversion must conserve the spatial and temporal patterns specific to ASD which is not possible using trivial image conversions. While not used in brain imaging, such works include Barra et al [19] who applied the GAF algorithm to encode time series to images for financial forecasting. A study by Vidal et al. [20] used the GAF technique to transform the time series of gold prices into images, and used a combination of LSTM and VGG to predict the future volatility of gold. To the best of our knowledge, GAF i.e. time-series to image conversion has not been exploited for diagnostic classification of ASD to neurotypical using fMRI data.

III. Problem Statement

We are given a labeled dataset of N fMRI scans $X = \{(x_i, y_i)\,|i = 1, \ldots, N\}$ and each scan consists of a time series of BOLD volumes v for each ROI r. Each subject $x_i = \{v_r^i | r = 1, \ldots, M\}$, where M is the number of ROIs, and $v_r \in R^{t_i}$ where t_i is the length of the time series of subject i. Each sample is associated with a label $y_i \in \{ASD, NT\}$. Our strategy is to transform the time series of x_i into GAF images. Therefore, M GAF images of size $t_i \times t_i$ are generated for each subject i in the dataset. The goal is to use a combination of a pretrained ResNet50 model on ImageNet and an LSTM layer to extract useful representations of the data that will assist to perform well on classifying ASD subjects from neurotypical (NT) subjects.

IV. Encoding the time series

To date, various classification methods have been developed and introduced using only the Pearson's correlation between regions. In this study, we consider encoding the time series into images to allow the use of massive innovative computer vision models such as ResNet and VGG. We used the Gramian Angular Field (GAF) which was prepared according to the procedure of Wang et al. [21]. GAF encodes a time series into images using polar coordinates, preserving the temporal correlation. It does not modify the data of fMRI scans, but instead transforms the representations into images that show the temporal correlation of the time series. Let X be the time series $X = x_1, x_2, \ldots x_n$ where n is the number of time points

in the time series, which is first normalized to be between $[-1, 1]$. The time series must be scaled to allow calculating the arccos value of each time point. The normalization can be calculated using the following equation:

$$\hat{x} = \frac{(x_i - \max(X) + (x_i + \min(X))}{\max(X) - \min(X)} \tag{1}$$

Then, the normalized time series will be represented in polar coordinates, where the value is the angular cosine and the radius is the time stamp:

$$\begin{cases} \theta = \arccos(\hat{x}_i), & -1 \leq \hat{x}_i \leq 1, \hat{x}_i \in \hat{X} \\ & r = \frac{t_i}{N}, t_i \in N \end{cases} \tag{2}$$

The above equation represents the angular cosine to one and only one value in the polar coordinates, which makes it possible to retrieve the raw time series. A Gramian matrix is computed using the polar coordinate system, where each element in the matrix is a trigonometric sum of each angle. The GAF matrix is defined as follows:

$$G = \begin{bmatrix} \cos(\theta_1 + \theta_1) & \dots & \cos(\theta_1 + \theta_n) \\ \cos(\theta_2 + \theta_1) & \dots & \cos(\theta_2 + \theta_n) \\ \vdots & \ddots & \vdots \\ \cos(\theta_n + \theta_1) & \dots & \cos(\theta_n + \theta_n) \end{bmatrix} \tag{3}$$

$$= \hat{X}' \cdot \hat{X} - \sqrt{I - \hat{X}^2}' \cdot \sqrt{I - \hat{X}^2} \tag{4}$$

In equation 4, I represents the unit row vector $[1, 1, 1 \dots 1]$. The size of the GAF matrix is $n \times n$ where n is the length of the time series. In our dataset, each subject has 200 time series where each time series represents the BOLD volumes of each region. Using GAF, we generated 200 images for each subject where each image represents one region of the brain. These images will be used in our deep learning model, ASD-GResTM, to distinguish subjects with ASD from neurotypical subjects. In the dataset, time-series lengths ranged from 78 to 316 time points. Therefore, the dimensions of the GAF images are based on the length of the time series. In this case, the dimensions range from 78×78 to 316×316. Because of the requirement when using a pretrained ResNet50, the dimensions of the images must be at least 224 × 224. In this case, we used the entire time series to generate the GAFs, and then resized the images to 224 × 224. Fig. 1 illustrates the process for generating GAFs.

V. Deep Learning Model

Owing to the extensive use of deep learning models in computer vision, we used the GAF framework to encode the BOLD time series into images. The GAF preserves the temporal nature of a region's activities over time. ASD-GResTM framework combines three parts. First, we used a Convolutional Neural Network (CNN) to extract useful features from GAF images. We then used a Long Short-Term Memory (LSTM) layer to learn the activities between the regions. Finally, the output representations of the last LSTM layer are applied to a single-layer perceptron (SPL). In the CNN model, we

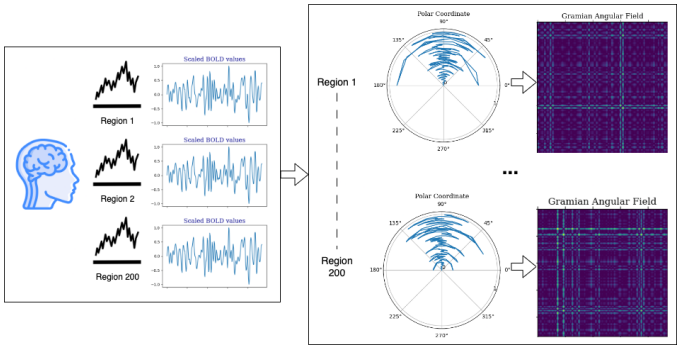

Fig. 1. Illustration of how to generate GAF images for each subject.

employed a pretrained ResNet50 which had been trained on the extensive ImageNet dataset. This choice was motivated by the fact that ImageNet comprises an extensive collection of images (with 1000 samples per 1000 classes) in contrast to our dataset. As a result, utilizing a pretrained model allowed us to initialize the model with the high-level performance achieved by ImageNet and then fine-tune it using our dataset. This approach aimed to capture valuable features without training from scratch on our relatively small dataset.

To facilitate this, we kept the ResNet50 layers frozen, except for the last linear layer. This strategy enables to extract relevant features based on the generated GAF images. To tailor the architecture to our needs, we adjusted the output vector size of the ResNet50 to consist of 512 neurons.

$$h_i = f(x_i) = ResNet(x_i) \tag{5}$$

In the LSTM block, we use a two-layer Bidirectional LSTM (BiLSTM) with an input size and hidden size of 512 and 256, respectively. Because it is a BiLSTM, the output of the last layer has a vector size of 512, which consists of 256 in each direction. For the classifier, once a ReLU nonlinearity is applied, a vector space with a dimensionality of 512 is fed into a single linear layer with two neurons, each corresponding to a distinct class. This computation is represented by the following equation:

$$z_i = g(h_i) = W^{(3)}\sigma(W^{(2)}h_i) \tag{6}$$

where σ denotes the ReLU function. The probability for each class is determined through the utilization of a Softmax function. Subsequently, the Cross-Entropy loss function is applied to compute the classification loss, thereby optimizing the performance of our model. These sequential steps are demonstrated by the following equations:

$$f(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}} \tag{7}$$

$$L_{CE} = -\sum_{i=1}^{N} y_i \log(f(z_i)) \tag{8}$$

where y_i represents the true label, and $f(z_i)$ represents the softmax probability of the i^{th} class.

The framework was developed using PyTorch-Lighting, a lightweight framework for flexibility and reproducibility built on top of PyTorch. Fig 2 shows the proposed framework.

VI. Experiments and results

A. Datasets

In this study, we use the Autism Brain Imaging Data Exchange I (ABIDE-I) dataset which is publicly available at (http://preprocessed-connectomes-project.org/abide/). The dataset consists of 1035 resting-state functional magnetic resonance imaging (rs-fMRI) scans derived from a diverse cohort of 505 subjects with Autism Spectrum Disorder (ASD) and 530 subjects with neurotypical (NT). These scans were collected from 17 distict sites across the United States. To extract the BOLD volumes, the scans must undergo meticulous preprocessing steps that require intensive execution time and effort. This dataset was pre-processed using several pipelines [22]. In this study, we used preprocessed data from a Configurable Pipeline for the Analysis of Connectomes (C-PAC). This pipeline applied several techniques to extract BOLD volumes, including skull stripping, surface reconstruction, slice-timing correction, head motion correction, and noise signal removal. More details of the preprocessing steps can be found at (http://preprocessed-connectomes-project.org/abide/cpac.html). Because the brain is divided into small cubics (4 mm × 4 mm × 4 mm), we used Craddock 200 (CC200), which parcellates the brain into 200 regions [23]. Most ASD classification frameworks, such as ASD-DiagNet [12] and ASD-SAENet [17] use Pearson's correlation between the regions, which transforms the input of each subject into a vector of $n \times (n-1)/2$, where n is the number of regions in the brain. In this study, we encoded the data into 200 GAF images for each participant. The scarcity of sufficient training samples per center adds to the complexity of the study. To conduct our evaluation, we selected four sites from the ABIDE-I dataset: Stanford, San Diego State University (SDSU), Social Brain Lab (SBL), and California Institute of Technology (Caltech). The distribution of classes for each individual site is outlined in Table I.

TABLE I
THE CLASS MEMBERSHIP OF EACH SITE.

Center	ASD	Neurotypical (NT)
Stanford	19	20
SDSU	14	22
SBL	15	15
Caltech	19	18

B. Computational Pipeline

We formulated a deep learning architecture capable of receiving sequential image data as input, and subsequently predicting the class of images. The ResNet50 model pretrained on ImageNet was used to extract useful representations from the GAF images. The ResNet's layers were frozen except for the last linear layer, which was fine-tuned for the GAF images. The output representations of each image were then passed to the BiLSTM for further sequence-related extraction. The features of the last layer of the BiLSTM were passed to a single preceptron to predict the final class. In this study, we used the dataset of each center to train, validate, and test the model. For each center run, the data were split into 5-fold croos validation. For each fold, the data was divided into 80% for the training set and 20% for the testing set. We extracted 10% of the training data for the validation set. The validation set was used to fine-tune the hyper-parameters of the model and determine the best model during the training phase to evaluate the final performance of the model using the untouched testing set. For training, the Adam optimizer was used to optimize the weight of the model. The learning rate was set to 0.0001 and the weight decay was 0.01. The model was trained for 25 epochs and the best model performance on the validation set was used for testing. The reported performance in this study is the average of the 5-fold cross validation.

C. Classification performance

To evaluate our framework, we use 5-fold cross validation. We computed the average of all folds to obtain the final accuracy, sensitivity, and specificity for each center. Table II shows the average accuracy, sensitivity, and specificity of each center. We compared our framework with state-of-the-art methods ([17] [12] [16]), where all frameworks used functional connectivity of the regions and a deep learning model for classification. Table III presents a performance comparison of different methods. We evaluated the performance of the model's diagnostic test by analyzing its Receiver Operating Characteristic (ROC) curves for each center. The Area Under the ROC Curve (AUC) measures the model's ability to distinguish between individuals with autism spectrum disorder (ASD) and neurotypical based on different thresholds. A higher AUC value indicates that the model is more effective in distinguishing between ASD and neurotypical subjects. Fig. 3 shows the ROC curves for each center. Our framework performed better on two of the sites compared to state-of-the-art methods. As the results show, encoding data from time series to images will open a wide space for applying many more deep learning models for feature extraction and classification. In this study, we twisted predesigned components to illustrate the benefits of the GAF method and confirm the exploration of the topic by designing a deep learning model that deals specifically with GAF images generated from fMRI data.

TABLE II
THE ACCURACY, SENSITIVITY, AND SPECIFICITY OF OUR MODEL ASD-GRESTM FOR EACH CENTER.

Center	Accuracy	Sensitivity	Specificity
Stanford	81.78	83.33	80
SDSU	61.42	36.66	77
SBL	63.3	66.66	60
Caltech	51.42	46.66	55

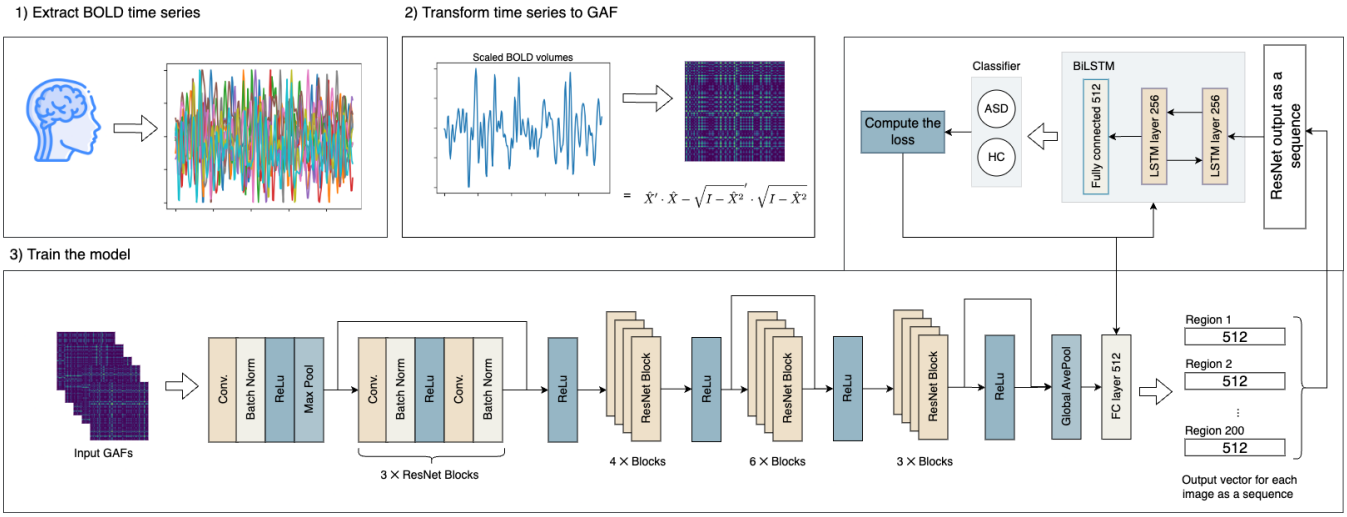

Fig. 2. Overall classification framework for ASD-GResTM. 1) The time series of BOLD volumes are extracted using extensive preprocessing steps which output (in our case) 200 time series of 200 regions. 2) Each region's time series is scaled to be between [-1, 1], and then converted to GAF images, as shown on the right. 3) Our deep learning architecture starts with a pretrained ResNet50 on ImageNet and finetunes the last linear layer for GAF-specific learning. The outputs of all the GAF images are stacked together to form a sequence vector of size 200 × 512. This sequence is then passed to a BiLSTM to capture useful representations, and the last output representations are passed to a single-layer perceptron to classify the probability of each class.

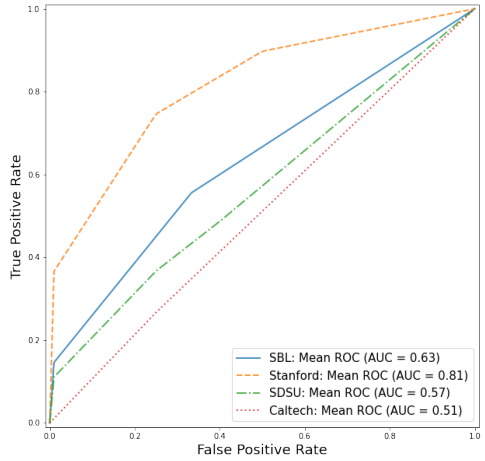

Fig. 3. The ROC curves for each center show how our model, ASD-GResTM, can distinguish between ASD and NT individuals using different threshold values, particularly for the Stanford center.

TABLE III
THE ACCURACY COMPARISON OF OUR MODEL ASD-GRESTM AND THE STATE-OF-THE-ART. THE HIGHEST ACCURACY IS INDICATED IN BOLD.

Center	Method	Accuracy
Stanford	ASD-SAENet [17]	53.2
	ASD-DiagNet [12]	64.2
	Heinsfeld et al. [16]	48.5
	ASD-GResTM	**81.78**
SDSU	ASD-SAENet [17]	**64.2**
	ASD-DiagNet [12]	63
	Heinsfeld et al. [16]	63.6
	ASD-GResTM	61.42
SBL	ASD-SAENet [17]	56.6
	ASD-DiagNet [12]	51.6
	Heinsfeld et al. [16]	46.6
	ASD-GResTM	**63.3**
Caltech	ASD-SAENet [17]	**56.7**
	ASD-DiagNet [12]	52.8
	Heinsfeld et al. [16]	52.3
	ASD-GResTM	51.42

TABLE IV
THE ACCURACY, SENSITIVITY, AND SPECIFICITY FOR EACH CENTER USING AAL ATLAS USING OUR MODEL ASD-GRESTM.

Center	Accuracy	Sensitivity	Specificity
Stanford	76.42	65	85
SDSU	55.35	20	76
SBL	56.66	46.66	66.66
Caltech	51.42	40	60

To demonstrate the effectiveness of the selection of brain parcellations, we tested ASD-GResTM on two other brain atlases, Automated Anatomical Labeling (AAL) [24] and Dosenbach 160 (DOS160) [25] which divide the brain into 116 and 160 regions of interest (ROIs) respectively. Table IV and Table V present the average accuracy, sensitivity, and specificity for each brain parcellation.

D. Interpretability and explainability of the model

To illustrate the findings of our proposed framework and the unique features of the GAF images used for the classification, we used Saliency Maps, which have been utilized for CNN models by Simonyan et al.. al. [26]. The purpose of the saliency method is to show the most noticeable locations of the images, which are used to differentiate between classes in the downstream task. In this section, we show different use cases in which the model can correctly classify ASD and neurotypical subjects in the testing phase. Two different regions are illustrated to visualize the GAF of different ROIs and the manner in which the model is capable of detecting discrete features of independent ROIs. Fig. 4 shows the GAF image of

TABLE V
THE ACCURACY, SENSITIVITY, AND SPECIFICITY FOR EACH CENTER USING DOS160 ATLAS USING OUR MODEL ASD-GRESTM.

Center	Accuracy	Sensitivity	Specificity
Stanford	74.64	73.33	75
SDSU	55.35	26	72
SBL	43.33	33.33	53.33
Caltech	48.57	46.66	55

a specific region of interest (ROI) for two correctly classified subjects (top is an ASD subject and bottom is a neurotypical subject). The image was accompanied by a corresponding saliency map generated after passing the region of interest (ROI) image through the model. To further demonstrate our approach, we provide another ROI for the same use case, as shown in Fig. 5. In ASD GAF images, useful representations are typically centralized in one area in the image, as annotated by a red rectangle, whereas in neurotypical GAF images, useful representations are scattered throughout the entire image as highlighted with a red rectangle. We present a neurotypical subject who was misclassified, and it can be seen that the representations resemble ASD patterns, but the subject is a neurotypical (NT), as depicted in Fig. 6. This situation highlights the need for more interpretable techniques and the examination of additional subjects in both groups to determine whether there are any common patterns within the classes that need to be disregarded, or if they indicate that the subjects are more likely to be found in between the two classes. As ASD is a spectrum disorder, it is important to consider the potential for greater variability in symptoms and behaviors.

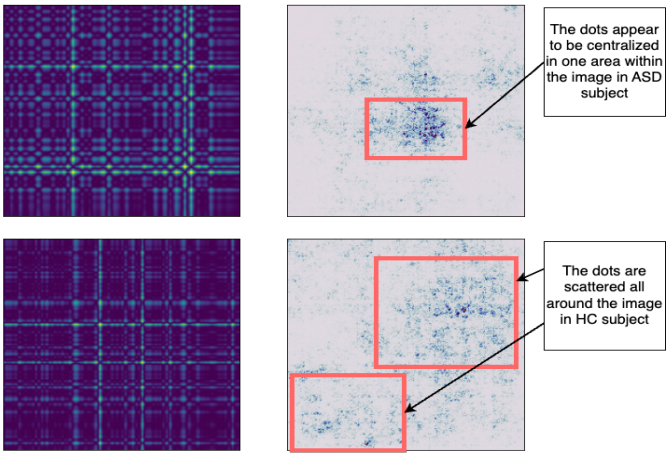

Fig. 4. On the left, we present a GAF image of a specific region of interest (ROI), and we present the output of the saliency method (the top is an ASD subject and the bottom is a neurotypical subject). These two subjects were classified correctly, and it can be seen that the formation of the represenations of the two subjects are varied with an open eye. This variation shows how the GAF representations can carry the spatial and temporal patterns of the time series, and that our downstream task can be learned to capture class-specific representations.

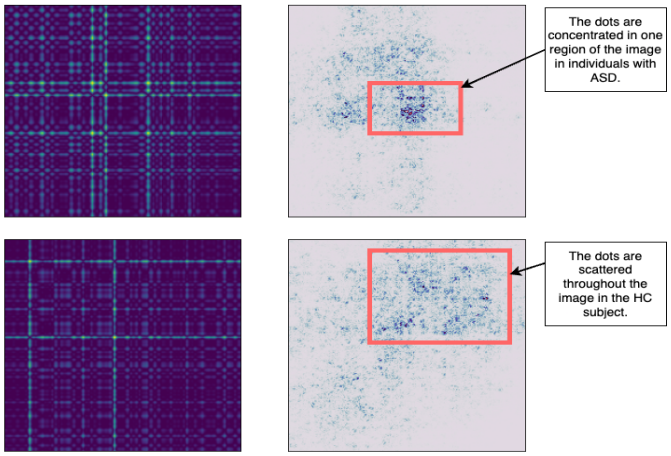

Fig. 5. GAF images' representations of another region of interest (ROI) for two correctly classified subjects (top is an ASD subject and bottom is a neurotypical subject). On the left, we show the GAF images of the ROI and their corresponding saliency method outputs on the right side. ASD representations on the top right are more centerized compared to the bottom right NT representations, which are scattered throughout the image.

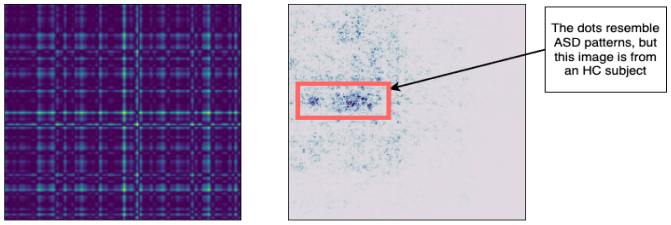

Fig. 6. The GAF image on the right is of a neurotypical subject, where it was misclassified as ASD. The saliency method output on the right shows that the representations of the input image are very similar to the ASD patterns, which are more centralized in one region of the image.

E. Scalability of the model

Our model ASD-GResTM is executable and scalable in terms of time and hardware resources. To illustrate how the execution time varies with sample size, we varied the sample size from 100 to 1000 subjects. It is worth noting that we set the batch size to 4 while training our model because of the size of the inputs in our case. The running time of our model ranges from 12 mins (100 subjects) up to 1.38 hours (1000 subjects). These times are comparable with the state-of-the-art methods, as the running time needed for each method is as follows: 41 mins by ASD-DaigNet [12], 52 mins by ASD-SAENet [17], and 6 hours by Heinsfeld et al. [16] Our proposed model can be trained and tested on 1K subjects within a reasonable time frame (1.38 hours) compared to most deep learning models, especially in the field of computer vision. Training was performed for only one data split: 80% for training and 20% for testing. The time reported was the time after the models were trained and tested. To execute our model, we used a Linux cluster node running Ubuntu 20.04 LTS contains two Intel Xeon Gold 5215, each having 10 cores with a total of 128 GBs of RAM. The node has an NVIDIA

RTX A6000 with 40 GBs of memory.

VII. Conclusion and Discussion

In this paper, we designed and developed a machine-learning framework called ASD-GResTM, which can be used as a classification method for autism spectrum disorder (ASD) using non-invasive fMRI data. We introduce a novel strategy to transform time-series data extracted from fMRI signals into Gramian Angular Field (GAF) which conserves the temporal and spatial patterns in the data. Our motivation was to encode the time-series, acquired from fMRI data, into images that can be used by deep-learning architectures that have been successful in computer vision. For ASD-GResTM framework, we used a Convolutional Neural Network (CNN) to extract useful features from GAF images. We then used a Long Short-Term Memory (LSTM) layer to learn the activities between the regions. Finally, the output representations of the last LSTM layer are applied to a single-layer perceptron (SPL) to get the final classification. Our extensive experimentation demonstrated that ASD-GResTM outperforms the state-of-art in two sites with a high accuracy of 81.78% in one of the centers while also exhibiting high sensitivity and specificity. Our preliminary results demonstrated in this paper opens new directions of encoding fMRI based time-series data into unique representation (in our case images) which can capture spatial and temporal patterns specific to ASD. Further research is needed to investigate data wrangling and transformation techniques that can be used to help ML models learn neural signatures that are specific to ASD and NT. The initial results being reported in this paper show exceptional potential for future research in this direction.

References

[1] R. E. Nickel and L. Huang-Storms, "Early identification of young children with autism spectrum disorder," *The Indian Journal of Pediatrics*, vol. 84, pp. 53–60, 2017.

[2] D. L. Christensen, "Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2012," *MMWR. Surveillance summaries*, vol. 65, 2016.

[3] L. Zwaigenbaum, A. Thurm, W. Stone, G. Baranek, S. Bryson, J. Iverson, A. Kau, A. Klin, C. Lord, R. Landa *et al.*, "Studying the emergence of autism spectrum disorders in high-risk infants: Methodological and practical issues," *Journal of autism and developmental disorders*, vol. 37, pp. 466–480, 2007.

[4] C. Lord, M. Rutter, and A. Le Couteur, "Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders," *Journal of autism and developmental disorders*, vol. 24, no. 5, pp. 659–685, 1994.

[5] C. Lord, M. Rutter, P. DiLavore, S. Risi, K. Gotham, S. Bishop *et al.*, "Autism diagnostic observation schedule–2nd edition (ados-2)," *Los Angeles, CA: Western Psychological Corporation*, vol. 284, 2012.

[6] J. Baio, L. Wiggins, D. L. Christensen, M. J. Maenner, J. Daniels, Z. Warren, M. Kurzius-Spencer, W. Zahorodny, C. R. Rosenberg, T. White *et al.*, "Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2014," *MMWR Surveillance Summaries*, vol. 67, no. 6, p. 1, 2018.

[7] S. Rogers and M. Talbott, "Early identification and early treatment of autism spectrum disorder," in *International review of research in developmental disabilities*. Elsevier, 2016, vol. 50, pp. 233–275.

[8] T. Yang, M. A. Al-Duailij, S. Bozdag, and F. Saeed, "Classification of autism spectrum disorder using rs-fmri data and graph convolutional networks," in *2022 IEEE International Conference on Big Data (Big Data)*. IEEE, 2022, pp. 3131–3138.

[9] Y. Qin, Y. Lou, Y. Huang, R. Chen, and W. Yue, "An ensemble deep learning approach combining phenotypic data and fmri for adhd diagnosis," *Journal of Signal Processing Systems*, vol. 94, no. 11, pp. 1269–1281, 2022.

[10] P. K. C. Prasad, Y. Khare, K. Dadi, P. Vinod, and B. R. Surampudi, "Deep learning approach for classification and interpretation of autism spectrum disorder," in *2022 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2022, pp. 1–8.

[11] C. J. Mellema, K. P. Nguyen, A. Treacher, and A. Montillo, "Reproducible neuroimaging features for diagnosis of autism spectrum disorder with machine learning," *Scientific reports*, vol. 12, no. 1, p. 3057, 2022.

[12] T. Eslami, V. Mirjalili, A. Fong, A. R. Laird, and F. Saeed, "Asd-diagnet: a hybrid learning approach for detection of autism spectrum disorder using fmri data," *Frontiers in neuroinformatics*, vol. 13, p. 70, 2019.

[13] X. Guo, K. C. Dominick, A. A. Minai, H. Li, C. A. Erickson, and L. J. Lu, "Diagnosing autism spectrum disorder from brain resting-state functional connectivity patterns using a deep neural network with a novel feature selection method," *Frontiers in neuroscience*, vol. 11, p. 460, 2017.

[14] A. El Gazzar, L. Cerliani, G. van Wingen, and R. M. Thomas, "Simple 1-d convolutional networks for resting-state fmri based classification in autism," in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–6.

[15] J. Zhang, F. Feng, T. Han, X. Gong, and F. Duan, "Detection of autism spectrum disorder using fmri functional connectivity with feature selection and deep learning," *Cognitive Computation*, pp. 1–12, 2022.

[16] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identification of autism spectrum disorder using deep learning and the abide dataset," *NeuroImage: Clinical*, vol. 17, pp. 16–23, 2018.

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

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[19] S. Barra, S. M. Carta, A. Corriga, A. S. Podda, and D. R. Recupero, "Deep learning and time series-to-image encoding for financial forecasting," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 3, pp. 683–692, 2020.

[20] A. Vidal and W. Kristjanpoller, "Gold volatility prediction using a cnn-lstm approach," *Expert Systems with Applications*, vol. 157, p. 113481, 2020.

[21] Z. Wang, T. Oates *et al.*, "Encoding time series as images for visual inspection and classification using tiled convolutional neural networks," in *Workshops at the twenty-ninth AAAI conference on artificial intelligence*, vol. 1. AAAI Menlo Park, CA, USA, 2015.

[22] C. Craddock, Y. Benhajali, C. Chu, F. Chouinard, A. Evans, A. Jakab, B. S. Khundrakpam, J. D. Lewis, Q. Li, M. Milham *et al.*, "The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives," *Frontiers in Neuroinformatics*, vol. 7, p. 27, 2013.

[23] R. C. Craddock, G. A. James, P. E. Holtzheimer III, X. P. Hu, and H. S. Mayberg, "A whole brain fmri atlas generated via spatially constrained spectral clustering," *Human brain mapping*, vol. 33, no. 8, pp. 1914–1928, 2012.

[24] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain," *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002.

[25] N. U. Dosenbach, B. Nardos, A. L. Cohen, D. A. Fair, J. D. Power, J. A. Church, S. M. Nelson, G. S. Wig, A. C. Vogel, C. N. Lessov-Schlaggar *et al.*, "Prediction of individual brain maturity using fmri," *Science*, vol. 329, no. 5997, pp. 1358–1361, 2010.

[26] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," *arXiv preprint arXiv:1312.6034*, 2013.