# A Deep-Generative Hybrid Model to Integrate Multimodal and Dynamic Connectivity for Predicting Spectrum-Level Deficits in Autism

Niharika Shimona D'Souza[1(✉)], Mary Beth Nebel[2,3], Deana Crocetti[2], Nicholas Wymbs[2,3], Joshua Robinson[2], Stewart Mostofsky[2,3,4], and Archana Venkataraman[1]

[1] Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, USA
Shimona.Niharika.Dsouza@jhu.edu
[2] Center for Neurodevelopmental and Imaging Research, Kennedy Krieger Institute, Baltimore, USA
[3] Department of Neurology, Johns Hopkins School of Medicine, Baltimore, USA
[4] Department of Psychiatry and Behavioral Science, Johns Hopkins School of Medicine, Baltimore, USA

**Abstract.** We propose an integrated deep-generative framework, that jointly models complementary information from resting-state functional MRI (rs-fMRI) connectivity and diffusion tensor imaging (DTI) tractography to extract predictive biomarkers of a disease. The generative part of our framework is a structurally-regularized Dynamic Dictionary Learning (sr-DDL) model that decomposes the dynamic rs-fMRI correlation matrices into a collection of shared basis networks and time varying patient-specific loadings. This matrix factorization is guided by the DTI tractography matrices to learn anatomically informed connectivity profiles. The deep part of our framework is an LSTM-ANN block, which models the temporal evolution of the patient sr-DDL loadings to predict multidimensional clinical severity. Our coupled optimization procedure collectively estimates the basis networks, the patient-specific dynamic loadings, and the neural network weights. We validate our framework on a multi-score prediction task in 57 patients diagnosed with Autism Spectrum Disorder (ASD). Our hybrid model outperforms state-of-the-art baselines in a five-fold cross validated setting and extracts interpretable multimodal neural signatures of brain dysfunction in ASD.

## 1 Introduction

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental disorder characterized by impaired social communicative skills and awareness, coupled

**Electronic supplementary material** The online version of this chapter (https://doi.org/10.1007/978-3-030-59728-3_43) contains supplementary material, which is available to authorized users.

with restricted/repetitive behaviors. These symptoms and levels of disability vary widely across the ASD spectrum. Neuroimaging techniques such as rs-fMRI and DTI are gaining popularity for studying aberrant brain connectivity in ASD [7]. Rs-fMRI allows us to assess the functional organization of the brain by tracking changes in steady-state co-activation [19], while DTI measures structural connectivity via the diffusion of water molecules in the brain [3]. However, the high data dimensionality, coupled with noise and patient variability, have limited our ability to integrate these modalities to understand behavioral deficits.

Techniques integrating structural and functional connectivity focus heavily on groupwise discrimination from the static connectomes. Methods include statistical tests on the node or edge biomarkers [29], data-driven representations [31], and neural networks [2] for classification. However, none of these methods tackle continuous-valued prediction, e.g., quantifying level of deficit. Deep learning is becoming increasingly popular for continuous prediction. The work of [17] proposes a specialized end-to-end convolutional network that predicts clinical outcomes from DTI connectomes. The authors of [14] combine a dictionary learning on the rs-fMRI correlations with an ANN to predict clinical severity in ASD patients. However, these methods focus on a single neuroimaging modality and do not leverage complementary information between structure and function.

There is now growing evidence that functional connectivity between regions is a dynamically evolving process [8], and that modeling this evolution is crucial to understanding disorders like ASD [25,27]. Hence, recent methods have been proposed that use either a sparse decomposition of the rs-fMRI connectomes [9], or a temporal clustering for ASD/control discrimination [26]. While promising, these approaches focus exclusively on rs-fMRI and ignore structural information.

We propose a hybrid deep-generative model that integrates structural and dynamic functional connectivity with behavior into a unified optimization framework. Our generative component is a structurally-regularized Dynamic Dictionary Learning (sr-DDL) model, which uses anatomical priors from DTI to regularize a time-varying decomposition of the rs-fMRI correlation matrices. Here, the connectivity profiles are explained by shared basis networks and time-varying patient-specific loadings. Simultaneously, these loadings are input to a deep network which uses an LSTM (Long Short Term Memory Network) to model temporal trends and an ANN (Artificial Neural Network) to predict clinical severity. Our optimization procedure learns the bases, loadings, and neural network weights most predictive of behavioral deficits in ASD. We obtain a representation which is both interpretable and generalizes to unseen patients, thus providing a comprehensive characterization of the disorder.

## 2    A Deep-Generative Hybrid Model for Connectomics

Figure 1 illustrates the generative (sr-DDL) and deep (LSTM-ANN) components of our framework. Let $P$ be the number of ROIs in our brain parcellation and $N$ be the number of patients. The rs-fMRI dynamic correlation matrices for patient $n$ are denoted by $\{\mathbf{\Gamma}_n^t\}_{t=1}^{T_n} \in \mathcal{R}^{P \times P}$, with $T_n$ being the number of time

steps. $\mathbf{L}_n \in \mathcal{R}^{P \times P}$ is the corresponding DTI connectivity information, and $\mathbf{y}_n \in \mathcal{R}^{M \times 1}$ is a vector of $M$ concatenated severity measures. Given the inputs $\mathcal{P} = \{\{\mathbf{\Gamma}_n^t\}, \mathbf{L}_n, \mathbf{y}_n\}_{n=1}^N$, our framework optimizes the following joint objective:

$$\mathcal{J}(\mathbf{B}, \{\mathbf{c}_n^t\}, \mathbf{\Theta}; \mathcal{P}) = \underbrace{\sum_n \mathcal{D}(\mathbf{B}, \{\mathbf{c}_n^t\}; \{\mathbf{\Gamma}_n^t\}, \mathbf{L}_n)}_{\text{sr-DDL loss}} + \lambda \underbrace{\sum_n \mathcal{L}(\mathbf{\Theta}, \{\mathbf{c}_n^t\}; \mathbf{y}_n)}_{\text{deep network loss}} \quad (1)$$
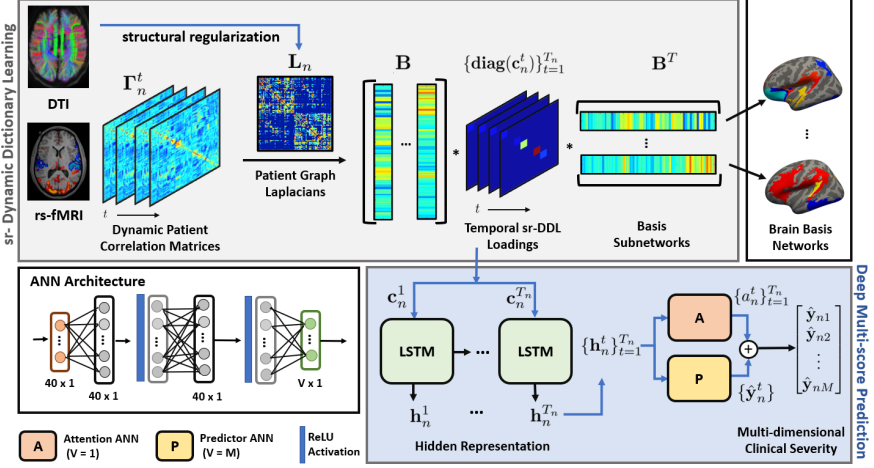


**Fig. 1.** Framework to integrate structural & dynamic functional connectivity for multi-task prediction **Gray Box:** sr-DDL module for rs-fMRI dynamic correlation matrices and DTI connectivity matrices. **Blue Box:** LSTM-ANN for multi-score prediction. (Color figure online)

**<u>sr-DDL Factorization:</u>** We represent the correlation matrices $\mathbf{\Gamma}_n^t$ by a shared basis $\mathbf{B} \in \mathcal{R}^{P \times K}$ that captures template patterns of co-activity and temporal loadings $\mathbf{c}_n^t \in \mathcal{R}^{K \times 1}$ that indicate their time-varying strength:

$$\mathcal{D}(\mathbf{B}, \{\mathbf{c}_n^t\}; \{\mathbf{\Gamma}_n^t\}, \mathbf{L}_n) = \sum_t \frac{1}{T_n} ||\mathbf{\Gamma}_n^t - \mathbf{B}\mathbf{diag}(\mathbf{c}_n^t)\mathbf{B}^T||_{\mathbf{L}_n} \quad s.t. \ \mathbf{B}^T\mathbf{B} = \mathcal{I}_K \quad (2)$$

Here, $K$ is the size of our basis, and $\mathbf{diag}(\mathbf{c}_n^t)$ is a diagonal matrix based on the elements of $\mathbf{c}_n^t$, and $\mathcal{I}_K$ is the identity matrix of size $K$. The positive semi-definiteness of $\{\mathbf{\Gamma}_n^t\}$ further implies that $\mathbf{c}_n^t$ is non-negative. The orthonormality constraint on $\mathbf{B}$ helps us learn uncorrelated sub-networks that explain the rs-fMRI data well and implicitly regularize the optimization.

Notice that Eq. (2) uses a weighted Frobenius norm, rather than the standard $\ell_2$ penalty. Mathematically, this norm is computed as $||\mathbf{X}||_{\mathbf{L}_n} = \text{Tr}(\mathbf{X}^T\mathbf{L}_n\mathbf{X})$ [20,28], with $\mathbf{X} = \mathbf{\Gamma}_n^t - \mathbf{B}\mathbf{diag}(\mathbf{c}_n^t)\mathbf{B}^T$. The matrix $\mathbf{L}_n \in \mathcal{R}^{P \times P}$ in our case
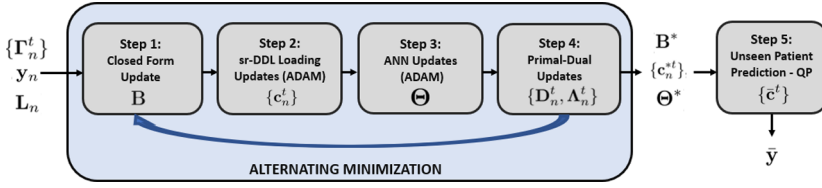
**Fig. 2.** Alternating minimization strategy for joint optimization of Eq. (4)

is the normalized graph Laplacian [4] derived from the DTI adjacency matrix for patient $n$. The DTI adjacency is 1 if there is at least one tract between the corresponding regions, and 0 otherwise. Essentially, this structural-regularization encourages the functional decomposition to focus on explaining the functional connectivity between regions with an a-priori anatomical connection.

**Deep Network:** The patient coefficients $\mathbf{c}_n^t$ are input to an LSTM-ANN network to predict the scores $\mathbf{y}_n$. The LSTM generates a hidden representation $\mathbf{h}_n^t$ over time. From here, the Predictor ANN (P-ANN) outputs a time varying estimate of the scores $\{\hat{\mathbf{y}}_n^t\}_{t=1}^{T_n}$. The Attention ANN (A-ANN) generates $T_n$ scalars, which we softmax across time to obtain the attention weights: $\{a_n^t\}_{t=1}^{T_n}$. These weights determine which time points for each patient are most relevant for behavioral prediction. The final prediction is an attention-weighted average across the estimates $\hat{\mathbf{y}}_n^t$. We use an MSE loss in Eq. (1) to obtain:

$$\mathcal{L}(\{\mathbf{c}_n^t\}, \mathbf{y}_n; \boldsymbol{\Theta}) = ||\hat{\mathbf{y}}_n - \mathbf{y}_n||_F^2 = \left|\left|\sum_t^{T_n} \hat{\mathbf{y}}_n^t a_n^t - \mathbf{y}_n\right|\right|_F^2 \tag{3}$$

We employ a two layered LSTM with hidden layer width 40. As seen in Fig. 1, both the P-ANN and the A-ANN have two hidden layers with width 40 with ReLU activations, with output size $(V)$ as $M$ and 1 respectively. We observed that these modeling choices are robust to issues with saturation and vanishing gradients that can hinder the training of deep neural networks. Finally, we fix the trade-off between the losses in Eq. (1) at $\lambda = 3$, and the number of networks to $K = 15$ based on a grid search.

## 2.1  Coupled Optimization Strategy

We use alternating minimization to optimize Eq. (1) with respect to $\{\mathbf{B}, \{\mathbf{c}_n^t\}, \boldsymbol{\Theta}\}$. Here, we iteratively cycle through the updates for the dictionary $\mathbf{B}$, loadings $\{\mathbf{c}_n^t\}$, and the LSTM-ANN weights $\boldsymbol{\Theta}$ to obtain a *joint solution*.

We note that there is a closed-form Procrustes solution for quadratic objectives [15]. However, Eq. (1) is bi-quadratic in $\mathbf{B}$, so it cannot be directly applied. Therefore, we adopt the strategy in [12–14], by which we introduce the constraints of the form $\mathbf{D}_n^t = \mathbf{B}\mathrm{diag}(\mathbf{c}_n^t)$, with corresponding augmented

Lagrangian variables $\{\mathbf{\Lambda}_n^t\}$. Thus, our objective from Eq. (1) now becomes:

$$\mathcal{J}_c = \sum_{n,t} \frac{1}{T_n} ||\mathbf{\Gamma}_n^t - \mathbf{D}_n^t \mathbf{B}^T||_{\mathbf{L}_n} + \lambda \sum_n \mathcal{L}(\mathbf{\Theta}, \{\mathbf{c}_n^t\}; \mathbf{y}_n) \ \ s.t. \ \ \mathbf{B}^T\mathbf{B} = \mathcal{I}_K$$

$$+ \sum_{n,t} \frac{\gamma}{T_n} \left[ \mathrm{Tr}\left[ (\mathbf{\Lambda}_n^t)^T (\mathbf{D}_n^t - \mathbf{B}\mathbf{diag}(\mathbf{c}_n^t)) \right] + \frac{1}{2} ||\mathbf{D}_n^t - \mathbf{B}\mathbf{diag}(\mathbf{c}_n^t)||_F^2 \right] \quad (4)$$

Figure 2 outlines our coupled optimization strategy, with steps detailed as follows:

**Step 1: Closed form Solution for B.** Notice that Eq. (4) reduces to a Procrustes objective $\mathbf{B}^* = \arg\min_{\mathbf{B}: \ \mathbf{B}^T\mathbf{B}=\mathcal{I}_K} ||\mathbf{M} - \mathbf{B}||_F^2$ where:

$$\mathbf{M} = \sum_n \frac{1}{T_n} \sum_t (\mathbf{\Gamma}_n^t \mathbf{L}_n + \mathbf{L}_n \mathbf{\Gamma}_n^t) \mathbf{D}_n^t + \gamma \mathbf{D}_n^t \mathbf{diag}(\mathbf{c}_n^t) + \gamma \mathbf{\Lambda}_n^t \mathbf{diag}(\mathbf{c}_n^t)$$

Given the singular value decomposition $\mathbf{M} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, then $\mathbf{B}^* = \mathbf{U}\mathbf{V}^T$. Thus, $\mathbf{B}$ spans the anatomically weighted space of patient correlation matrices.

**Step 2: Updating the sr-DDL Loadings $\{\mathbf{c}_n^t\}$.** The objective $\mathcal{J}_c$ in Eq. (4) decouples across patients. We can also incorporate the non-negativity constraint $\mathbf{c}_{nk}^t \geq 0$ by passing an intermediate vector $\hat{\mathbf{c}}_n^t$ through a ReLU. The ReLU pre-filtering allows us to optimize an unconstrained version of Eq. (4), which can be done via the stochastic ADAM algorithm [18]. In essence, this optimization couples the parametric gradient from the augmented Lagrangians with the backpropagated gradient from the deep network (defined by fixed $\mathbf{\Theta}$). After convergence, the thresholded loadings $\mathbf{c}_n^t = ReLU(\hat{\mathbf{c}}_n^t)$ are used in subsequent steps.

**Step 3: Updating the Deep Network Weights $\mathbf{\Theta}$.** We use backpropagation on the loss $\mathcal{L}(\cdot)$ to solve for $\mathbf{\Theta}$. Notice that we can handle missing clinical data by dropping the contributions of the unknown value of $\mathbf{y}_{nm}$ to the network during backpropagation. We use the ADAM [18] optimizer with random initialization, a learning rate of $10^{-4}$, scaled by 0.95 every 5 epochs, and batch-size 1.

**Step 4: Updating the Constraint Variables $\{\mathbf{D}_n^t, \mathbf{\Lambda}_n^t\}$.** We perform parallel primal-dual updates for the constraint pairs $\{\mathbf{D}_n^t, \mathbf{\Lambda}_n^t\}$ [1]. Here, we cycle through the closed form update for $\mathbf{D}_n^t$ and gradient ascent for $\mathbf{\Lambda}_n^t$ until convergence.

**Step 5: Prediction on Unseen Data.** In our cross-validated setting, we need to compute the sr-DDL loadings $\{\bar{\mathbf{c}}^t\}_{t=1}^{\bar{T}}$ for a new patient based on the training $\mathbf{B}^*$. Since we do not know the score $\bar{\mathbf{y}}$ for this patient, we remove the contribution $\mathcal{L}(\cdot)$ from Eq. (4) and assume the constraints $\bar{\mathbf{D}}^t = \mathbf{B}^*\mathbf{diag}(\bar{\mathbf{c}}^t)$ hold with equality, thus removing the Lagrangian terms. Essentially, the optimization for $\{\bar{\mathbf{c}}^t\}$ reduces to decoupled quadratic programming (QP) objectives $\mathcal{Q}_t$ across time:

$$\bar{\mathbf{c}}^{*t} = \arg\min_{\bar{\mathbf{c}}^t} \frac{1}{2} (\bar{\mathbf{c}}^t)^T \bar{\mathbf{H}}\bar{\mathbf{c}}^t + \bar{\mathbf{f}}^T \bar{\mathbf{c}}^t \ \ s.t. \ \ \bar{\mathbf{A}}\bar{\mathbf{c}}^t \leq \bar{\mathbf{b}}$$

$$\bar{\mathbf{H}} = 2(\mathbf{B}^{*T}\bar{\mathbf{L}}\mathbf{B}^*); \quad \bar{\mathbf{f}} = -[(\mathbf{B}^{*T}(\bar{\mathbf{\Gamma}}\bar{\mathbf{L}} + \bar{\mathbf{L}}\bar{\mathbf{\Gamma}})\mathbf{B}^*) \circ \mathcal{I}_K]\mathbf{1}; \quad \bar{\mathbf{A}} = -\mathcal{I}_K \ \bar{\mathbf{b}} = \mathbf{0}$$

where, ∘ denotes the Hadamard product. Finally, we estimate $\bar{\mathbf{y}}$ via a forward pass through the LSTM-ANN.

## 2.2    Baseline Comparisons

We compare the predictive performance of our framework against three baselines:

1. Two channel BrainNet CNN [17] on static rs-fMRI and DTI connectomes
2. PCA on DTI weighted dynamic rs-fMRI correlation features + LSTM-ANN
3. Decoupled sr-DDL factorization followed by the LSTM-ANN

The first baseline integrates multi-modal DTI connectivity with static rs-fMRI connectivity via the BrainNet CNN introduced in [17]. The original architecture is designed to predict cognitive scores from DTI. Here, we modify the BrainNet CNN to have two branches, one for rs-fMRI patient correlation matrices $\mathbf{\Gamma}_n \in \mathcal{R}^{P \times P}$ and the other for the DTI Laplacians $\mathbf{L}_n \in \mathcal{R}^{P \times P}$. We also modify the ANN in [17] to pool the learned representations and predict $M$ clinical severity measures. The hyperparameters are fixed according to [17].

For the second baseline, we weight the dynamic $P \times (P-1)/2$ rs-fMRI correlation features by the respective DTI Laplacian features. We then use Principal Component Analysis (PCA) to reduce the data dimensionality to $K = 15$, followed by a similar LSTM-ANN framework to map onto behavior.

Finally, we examine the score prediction upon excluding the DTI regularization from our deep-generative hybrid. This helps us evaluate the advantage of our multi-modal data integration, as opposed to analyzing rs-fMRI data alone.

## 3    Experimental Evaluation and Results

**Data and Preprocessing.** We validate our framework on a cohort of 57 children with high-functioning ASD. Rs-fMRI and DTI scans are acquired on a Philips $3T$ Achieva scanner (**rs-fMRI:** EPI, TR/TE = 2500/30 ms, flip angle = 70, res = $3.05 \times 3.15 \times 3$ mm, duration = 128 or 156 time samples; **DTI**: EPI, SENSE factor = 2.5, TR/TE = 6356/75 ms, res = $0.8 \times 0.8 \times 2.2$ mm, b-value = $700\,\mathrm{s/mm}^2$, 32 gradient directions). Rs-fMRI data was preprocessed through a standard pipeline that included motion correction, normalization to the MNI template, spatial and temporal filtering, and nuisance regression with CompCorr [6]. DTI data was preprocessed using the FDT pipeline in FSL [16]. We perform tractography using the BEDPOSTx and PROBTRACKx functions in FSL [5].

We use the Automatic Anatomical Labeling (AAL) atlas [32] to define 116 brain ROIs. A sliding window protocol (length = 45, stride = 5) was used to extract dynamic rs-fMRI correlations matrices. We subtract the first eigenvector, which is a roughly constant bias, and use the residual matrices as the inputs $\{\mathbf{\Gamma}_n^t\}$ for all methods. The DTI connectivity matrix is binary, where 1 corresponds to at least one tract between the two regions. We impute the DTI connectivity for the 11 patients, who do not have DTI based on the training data.

We rely on three clinical measures to characterize various impairments associated with ASD. The Autism Diagnostic Observation Schedule (ADOS) [23] captures socio-communicative deficits and restricted/repetitive behaviors via clinician evaluation (dynamic range: $0-30$). The Social Responsiveness Scale (SRS) [23] quantifies impaired social functioning via a parent/teacher questionnaire (dynamic range: $70-200$). Finally, Praxis [11,21] measures the ability to perform skilled motor gestures on command. A videotaped performance of the child is scored by two research-reliable raters (dynamic range: $0-100$).

**Multi-dimensional Severity Prediction.** Table 1 reports the *multi-score regression performance* of all methods. Figure 3 contrasts the performance of the deep-generative hybrid against the best performing baseline. We have included performance comparisons against the remaining baselines in Fig. 1 of the Supplementary Document. Here, we plot the severity score as given by the clinician on the **x**-axis, and the score predicted by the algorithm on the **y**-axis. The training and testing performance is illustrated by the red and blue points, respectively. The bold **x** = **y** diagonal line indicates ideal performance. Notice that all methods have a good training fit for all the scores. However, in case of testing performance, our method outperforms the baselines in almost all cases. Empirically, we are able to tune the baseline hyperparameters to obtain good testing performance on a single score (e.g. ADOS for Baseline 2), but the prediction of the remaining scores suffer. In contrast, the testing predictions from our framework follow the diagonal line more closely for all the scores. We believe that the representational flexibility of our deep network along with the joint optimization helps us generalize well.

**Table 1.** Performance based on **Median Absolute Error (MAE)** and **Mutual Information (MI)**. Lower MAE and higher MI indicate better performance.

| Score | Method | MAE train | MAE test | MI train | MI test |
|-------|--------|-----------|----------|----------|---------|
| ADOS | BrainNetCNN | 1.90 | 3.50 | 0.96 | 0.25 |
| | PCA & LSTM-ANN | 0.34 | **2.47** | 0.96 | **0.35** |
| | Without DTI reg | 0.13 | 3.27 | 0.99 | 0.26 |
| | **Our framework** | **0.08** | <u>2.84</u> | **0.99** | <u>0.34</u> |
| SRS | BrainNetCNN | 5.25 | 18.96 | 0.83 | 0.75 |
| | PCA & LSTM-ANN | 4.73 | 19.05 | 0.95 | 0.68 |
| | Without DTI reg | **0.49** | 18.70 | 0.97 | 0.77 |
| | **Our framework** | <u>0.51</u> | **17.81** | **0.98** | **0.88** |
| Praxis | BrainNetCNN | 3.78 | 15.15 | 0.95 | 0.19 |
| | PCA & LSTM-ANN | 2.21 | 20.71 | 0.90 | 0.47 |
| | Without DTI reg | 1.09 | 17.34 | 0.99 | 0.49 |
| | **Our framework** | **0.13** | **13.50** | **0.99** | **0.85** |
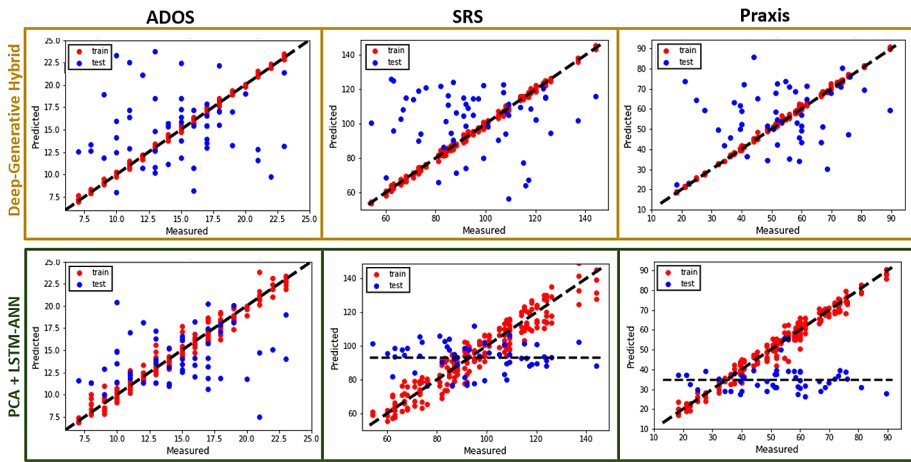
**Fig. 3.** Multi-score prediction by **Left:** ADOS **Middle:** SRS **Right:** Praxis by **Yellow Box:** deep-generative hybrid. **Green Box:** PCA+LSTM-ANN (Color online figure)
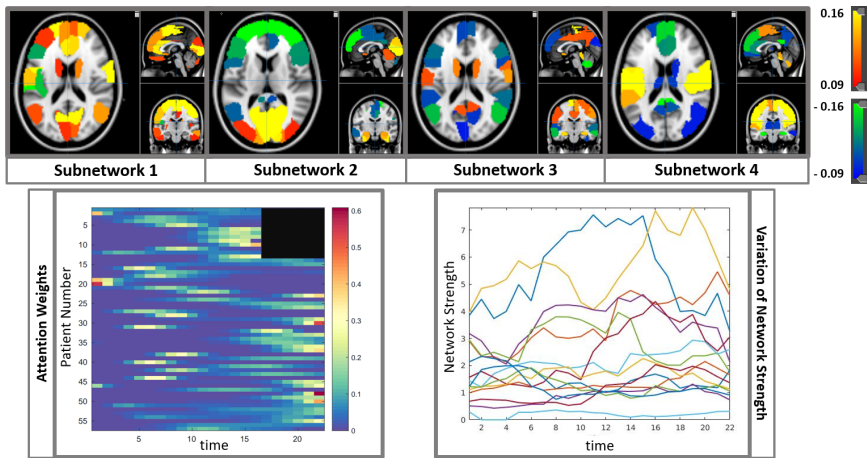


**Fig. 4. Top:** Subnetworks identified by the deep-generative hybrid. The red and orange regions are anti-correlated with the blue and green regions **Bottom: (Left)** Learned attention weights **(Right)** Variation of network strength over time (Color figure online)

**Clinical Interpretability.** Figure 4 (top) illustrates four representative subnetworks learned in **B**. (We have included the complete set of sub-network characterizations in Fig. 2 in the Supplementary Document.) Regions storing positive values are anticorrelated with negative regions. Subnetwork 1 includes regions from the default mode network (DMN), which has been widely reported in ASD [22]. Subnetwork 2 exhibits contributions from higher order visual processing and sensorimotor areas, concurring with behavioral reports of reduced visual-motor

integration in ASD [22]. Subnetworks 3 exhibits contributions from the central executive control network and insula, believed to be essential for switching between goal-directed and self-referential behavior [30]. Subnetwork 4 includes prefrontal and DMN regions, along with subcortical areas: associated with social-emotional regulation [24].

Figure 4 (bottom left) illustrates the learned attentions output by the A-ANN for all 57 patients during testing. We group patients with shorter scans in the first few rows of the plot. We have blackened the rest of the time points for these patients. The colorbar indicates the strength of the attention weights. The flagged non-zero weights denote intervals of the scan considered especially relevant for prediction. We observe that the network highlights the start of the scan for several patients, while it prefers focusing on the end of the scan for some others. This is indicative of the underlying patient heterogeneity.

Lastly, we illustrate the variation of the network strength for a patient in the cohort over the scan duration in Fig. 4 (bottom right). Each solid colored line corresponds to one of the 15 sub-networks. Over the scan duration, each network cycles through phases of activity and relative inactivity. Thus, only a few networks at each time step contribute to the patient's dynamic connectivity profile. This parallels the transient brain-states hypothesis in dynamic rs-fMRI connectivity [10], with active states as corresponding sub-networks in **B**.

## 4   Conclusion

We have introduced a novel deep-generative framework to integrate complementary information from the functional and structural neuroimaging domains, and simultaneously explain behavioral deficits in ASD. Our unique structural regularization elegantly injects anatomical information into the rs-fMRI functional decomposition, thus providing us with an interpretable brain basis. Our LSTM-ANN term not only models the temporal variation, but also helps isolate key dynamic resting-state signatures, indicative of clinical impairments. Our coupled optimization procedure ensures that we learn effectively from limited training data and generalize well to unseen patients. Finally, our framework makes very few assumptions and can potentially be applied to study other neuro-psychiatric disorders (e.g. ADHD, Schizophrenia) as an effective diagnostic tool.

## References

1. Afonso, M.V., Bioucas-Dias, J.M., Figueiredo, M.A.: An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. IEEE Trans. Image Process **20**(3), 681–695 (2010)
2. Aghdam, M.A., Sharifi, A., Pedram, M.M.: Combination of rs-fMRI and sMRI data to discriminate autism spectrum disorders in young children using deep belief network. J. Digit. Imaging **31**(6), 895–903 (2018)
3. Assaf, Y., Pasternak, O.: Diffusion tensor imaging (DTI)-based white matter mapping in brain research: a review. J. Mol. Neurosci. **34**(1), 51–61 (2008). https://doi.org/10.1007/s12031-007-0029-0

4. Banerjee, A., Jost, J.: On the spectrum of the normalized graph Laplacian. Linear Algebra Appl. **428**(11–12), 3015–3022 (2008)
5. Behrens, T.E., Berg, H.J., Jbabdi, S., Rushworth, M.F., Woolrich, M.W.: Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? Neuroimage **34**(1), 144–155 (2007)
6. Behzadi, Y., et al.: A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage **37**(1), 90–101 (2007)
7. Bennett, I.J., Rypma, B.: Advances in functional neuroanatomy: a review of combined DTI and fMRI studies in healthy younger and older adults. Neurosci. Biobehav. Rev. **37**(7), 1201–1210 (2013)
8. Cabral, J., Kringelbach, M.L., Deco, G.: Functional connectivity dynamically evolves on multiple time-scales over a static structural connectome: models and mechanisms. NeuroImage **160**, 84–96 (2017)
9. Cai, B., Zille, P., Stephen, J.M., Wilson, T.W., Calhoun, V.D., Wang, Y.P.: Estimation of dynamic sparse connectivity patterns from resting state fMRI. IEEE Trans. Med. Imaging **37**(5), 1224–1234 (2017)
10. Damaraju, E., et al.: Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. NeuroImage: Clin. **5**, 298–308 (2014)
11. Dziuk, M., Larson, J.G., Apostu, A., Mahone, E.M., Denckla, M.B., Mostofsky, S.H.: Dyspraxia in autism: association with motor, social, and communicative deficits. Dev. Med. Child Neurol. **49**(10), 734–739 (2007)
12. D'Souza, N.S., Nebel, M.B., Wymbs, N., Mostofsky, S., Venkataraman, A.: A generative-discriminative basis learning framework to predict clinical severity from resting state functional MRI data. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11072, pp. 163–171. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00931-1_19
13. D'Souza, N.S., Nebel, M.B., Wymbs, N., Mostofsky, S., Venkataraman, A.: A coupled manifold optimization framework to jointly model the functional connectomics and behavioral data spaces. In: Chung, A.C.S., Gee, J.C., Yushkevich, P.A., Bao, S. (eds.) IPMI 2019. LNCS, vol. 11492, pp. 605–616. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20351-1_47
14. D'Souza, N.S., Nebel, M.B., Wymbs, N., Mostofsky, S., Venkataraman, A.: Integrating neural networks and dictionary learning for multidimensional clinical characterizations from functional connectomics data. In: Shen, D. (ed.) MICCAI 2019. LNCS, vol. 11766, pp. 709–717. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32248-9_79
15. Everson, R.: Orthogonal, but not orthonormal, procrustes problems. Adv. Comput. Math. **3**(4), 782–790 (1998)
16. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M.: FSL. Neuroimage **62**(2), 782–790 (2012)
17. Kawahara, J., et al.: BrainNetCNN: convolutional neural networks for brain networks; towards predicting neurodevelopment. NeuroImage **146**, 1038–1049 (2017)
18. Kingma, D.P., Ba, J.L.: Adam: a method for stochastic optimization (2015)
19. Lee, M.H., Smyser, C.D., Shimony, J.S.: Resting-state fMRI: a review of methods and clinical applications. Am. J. Neuroradiol. **34**(10), 1866–1872 (2013)
20. Manton, J.H., Mahony, R., Hua, Y.: The geometry of weighted low-rank approximations. IEEE Trans. Sig. Process **51**(2), 500–514 (2003)
21. Mostofsky, S.H., Dubey, P., Jerath, V.K., Jansiewicz, E.M., Goldberg, M.C., Denckla, M.B.: Developmental dyspraxia is not limited to imitation in children with autism spectrum disorders. J. Int. Neuropsychol. Soc. **12**(3), 314–326 (2006)

22. Nebel, M.B., et al.: Intrinsic visual-motor synchrony correlates with social deficits in autism. Biol. Psychiatry **79**(8), 633–641 (2016)
23. Payakachat, N., et al.: Autism spectrum disorders: a review of measures for clinical, health services and cost-effectiveness applications. Expert Rev. Pharmacoeconomics Outcomes Res. **12**(4), 485–503 (2012)
24. Pouw, L.B., Rieffe, C., Stockmann, L., Gadow, K.D.: The link between emotion regulation, social functioning, and depression in boys with ASD. Res. Autism Spectrum Disord. **7**(4), 549–556 (2013)
25. Price, T., Wee, C.-Y., Gao, W., Shen, D.: Multiple-network classification of childhood autism using functional connectivity dynamics. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8675, pp. 177–184. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10443-0_23
26. Rabany, L., et al.: Dynamic functional connectivity in schizophrenia and autism spectrum disorder: convergence, divergence and classification. NeuroImage Clin. **24**, 101966 (2019)
27. Rashid, B., Damaraju, E., Pearlson, G.D., Calhoun, V.D.: Dynamic connectivity states estimated from resting fMRI identify differences among schizophrenia, bipolar disorder, and healthy control subjects. Front. Hum. Neurosci. **8**, 897 (2014)
28. Schnabel, R.B., Toint, P.L.: Forcing sparsity by projecting with respect to a non-diagonally weighted Frobenius norm. Math. Program. **25**(1), 125–129 (1983). https://doi.org/10.1007/BF02591723
29. Skudlarski, P., Jagannathan, K., Calhoun, V.D., Hampson, M., Skudlarska, B.A., Pearlson, G.: Measuring brain connectivity: diffusion tensor imaging validates resting state temporal correlations. Neuroimage **43**(3), 554–561 (2008)
30. Sridharan, D., et al.: A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. Proc. Nat. Acad. Sci. **105**(34), 12569–12574 (2008)
31. Sui, J., et al.: Combination of resting state fMRI, DTI, and sMRI data to discriminate schizophrenia by n-way MCCA+ JICA. Front. Hum. Neurosci. **7**, 235 (2013)
32. Tzourio-Mazoyer, N., et al.: Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage **15**(1), 273–289 (2002)