Demographic-Guided Attention in Recurrent Neural Networks for Modeling Neuropathophysiological Heterogeneity

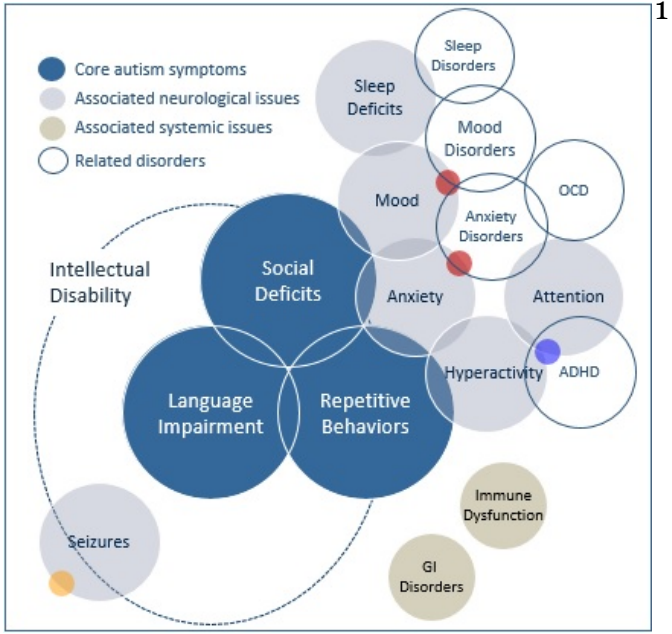
Nicha C. Dvornek, Xiaoxiao Li, Juntang Zhuang,
Pamela Ventola, and James S. Duncan

MLMI 2020
October 4, 2020

Yale SCHOOL OF MEDICINE

The Challenge of Learning from fMRI of Heterogeneous Psychiatric Disorders

- fMRI used to characterize pathophysiology of psychiatric disorders, e.g. autism spectrum disorder (ASD)
- ASD is extremely heterogeneous
- Early studies impose homogeneity
  - Restrict gender, age, etc.
  - → Smaller datasets
  - → Poor generalization of results
- Recent large open datasets (ABIDE)
  - Highly heterogeneous
  - → Poor classification accuracy of ASD/Control

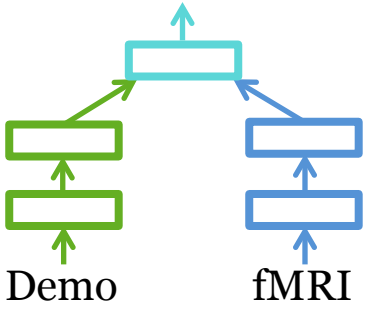


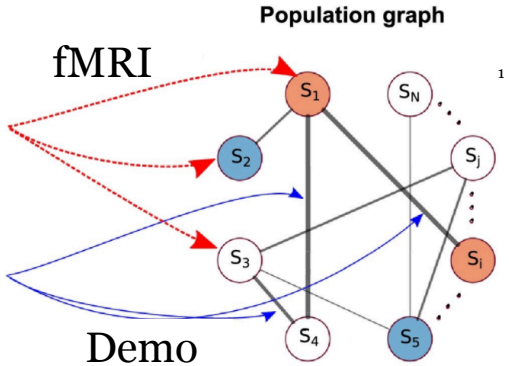
[1]

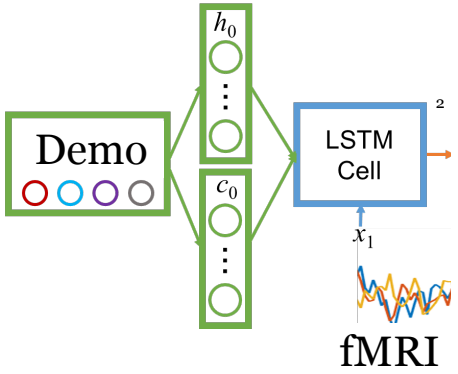
[1]www.autismspeaks.org

Include Demographic Information to Mitigate Heterogeneity Problem

- Non-imaging, scalar variables easy to obtain: Age, sex, IQ, …
- Many ways to incorporate demographic variables



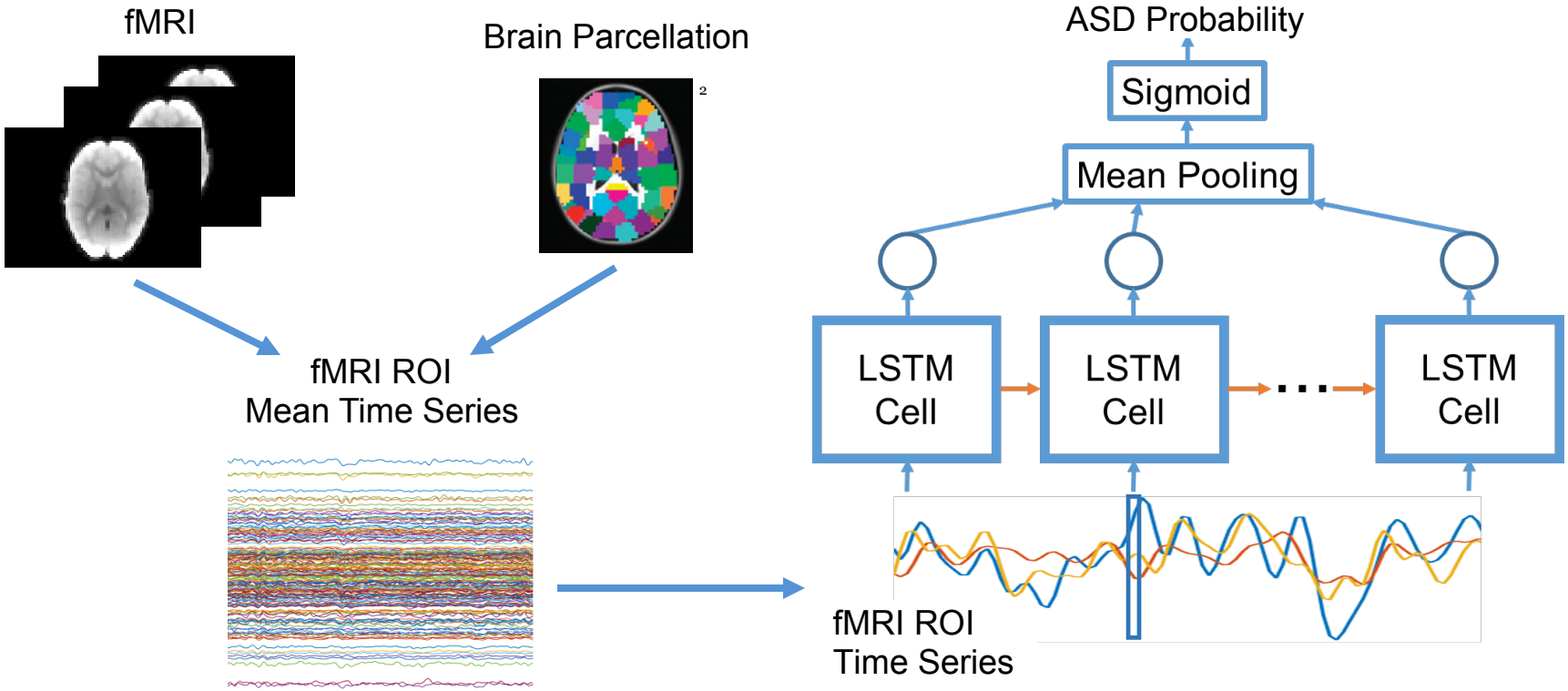




- No approach aims to modulate differences in neurological mechanisms
- **We model heterogeneous functional network patterns using a demographic guided attention + RNN model for fMRI**

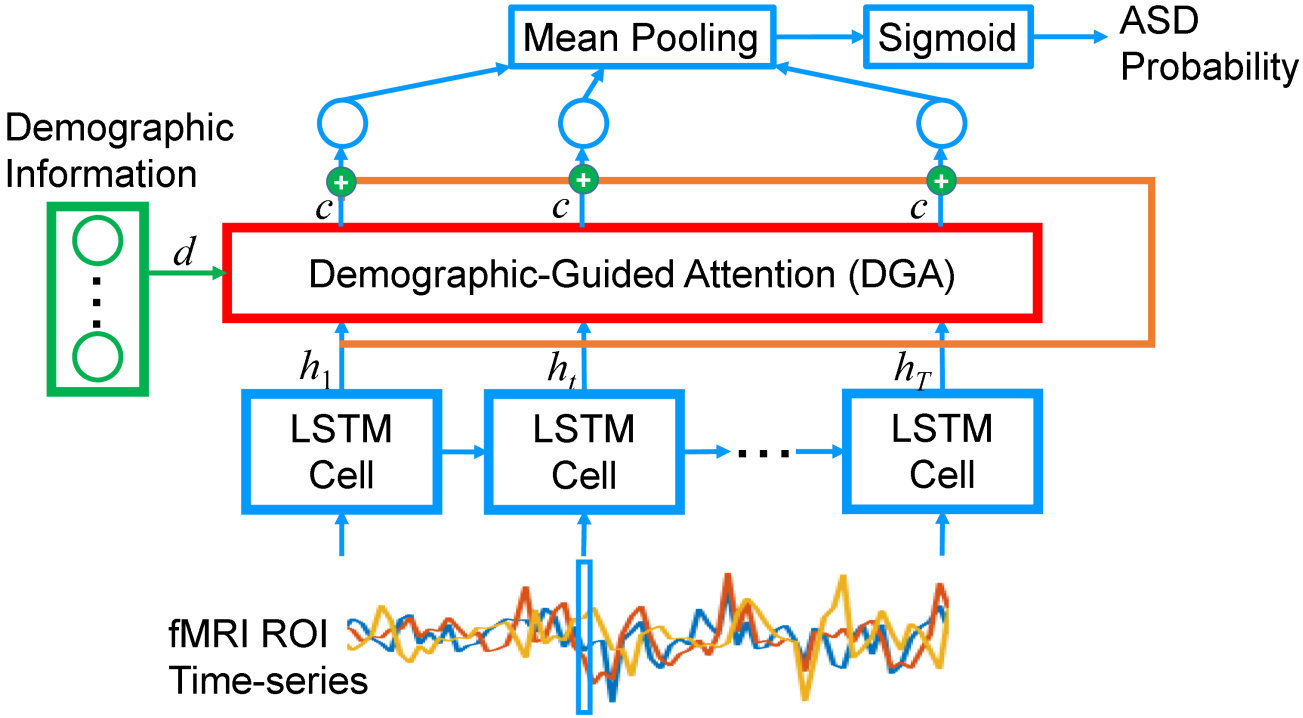
[1]Parisot et al., MedIA 2018; [2]Dvornek et al., MICCAI 2018

Baseline LSTM Network for fMRI Time-series Data[1]



[1]Dvornek et al., MLMI 2017; [2]Craddock et al., Nature Methods 2013

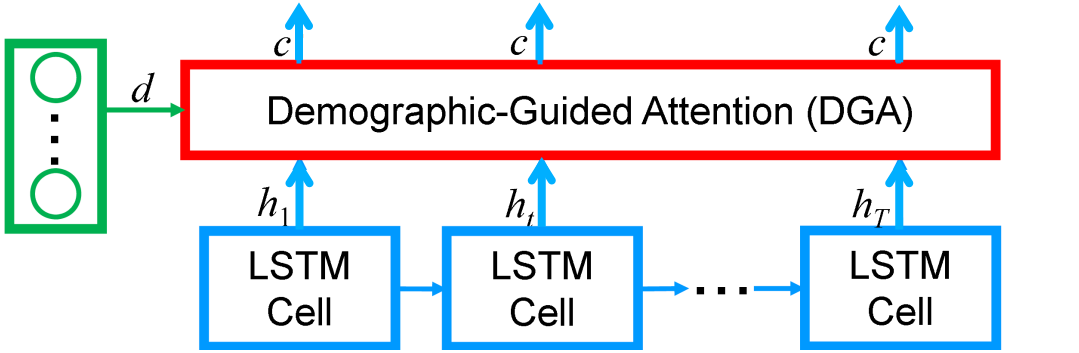
Proposed Demographic-Guided Attention Network



Generalized Attention Mechanism Based on Demographic and fMRI Information

- Query: Demographic information d
- Key and value: LSTM output h_t
- Scaled dot product attention computes context c:

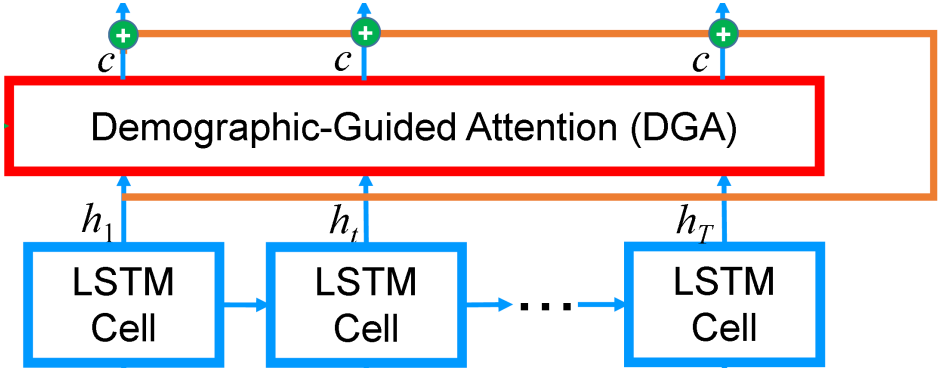
$$c = att\left(d, \{h_t\}\right) = \sum_{t=1}^{T} softmax\left[\underbrace{(W_q d)^T}_{\text{Query}} \underbrace{(W_k h_t)}_{\text{Key}} / \sqrt{m}\right] \underbrace{W_v h_t}_{\text{Value}}$$



Model Neurological Heterogeneity with Residual Connection between LSTM and Attention Outputs

- Use context to bias LSTM output

→ Change focus on LSTM nodes based on demographic information



- For multiple attention heads:
  - Process each head k output $c_k + h_t$ with separate FC layer
  - Take maximum score

Model Greater Neurological Heterogeneity with Multiple Attention Heads and Query Diversity Loss

- Single head: same demographics → same neuropathophysiology
- Multiple heads to model greater heterogeneity
- *Query Diversity Loss*: encourage *K* different attention heads to capture different underlying neuropathological modes:

$$L_{QD} = \sum_{i=1}^{N} \sum_{j=1}^{K-1} \sum_{k=j+1}^{K} \left| \frac{q_{ij}^T q_{ik}}{\|q_{ij}\| \, \|q_{ik}\|} \right|$$

Cosine proximity

$$q_{ij} = W_{q_j} d_i$$

= Query vector for subject *i*, attention mode *j*

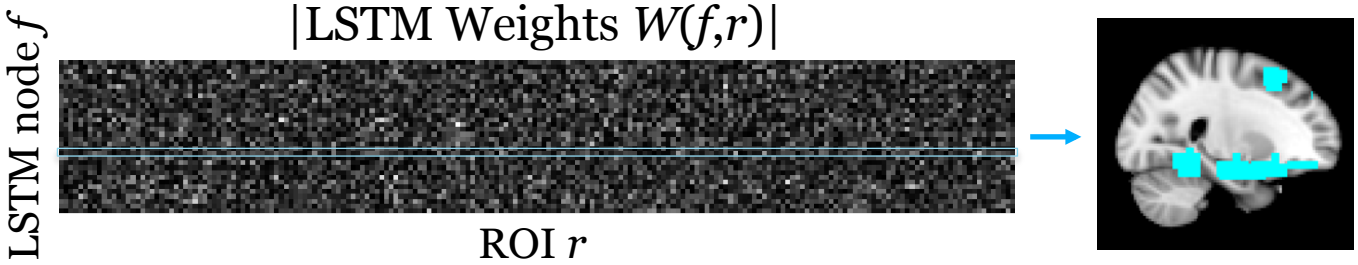
- Total loss: $L = L_C + \lambda L_{QD}$

Binary cross-entropy

0.5 in experiments

Interpretation of Demographic-Guided Attention as Neuropathological Heterogeneity

- LSTM node f : represents functional network
  - Assign membership by large LSTM weights of ROI inputs[1]



  - LSTM output $h(f)$: signal for functional network f
- Demographic information provides context for deciding which functional networks are important for ASD classification
  - c (f): demographic-guided attention to functional network f
  - Observe correlation between $d(i)$ and $c(f)$ across subjects

[1]Dvornek et al., MLMI 2017;

Datasets and Preprocessing

- Resting-state fMRI from multisite ABIDE I Dataset
- 3 Datasets from 3 prior publications
  - DS1[1]: N = 1100, CCS Pipeline, CC200 atlas
  - DS2[2]: N = 1035, CPAC Pipeline, CC200 atlas
  - DS3[3]: N = 860, CPAC Pipeline, HO atlas
- Standardize ROI mean time-series, resample at 2s interval
- Training: augment x10 by randomly cropping 3 min windows
- Inference: predict using all 3 min windows
- Demographic data: gender, age, handedness, full IQ, verbal IQ, performance IQ, eye status
  - Standardized to [-1,1]

[1]Dvornek et al., MLMI 2017; [2]Heinsfeld et al., Neuroimage clin., 2018; [3]Abraham et al., Neuroimage 2017

Classification of ASD vs. Healthy Control: Methods Compared

Model
Orig† [9]
LSTM [5]
DFuse [7]
DInit [6]
DGA1-C
DGA2-C
DGA1
DGA2
DGA2-QDL

Published results

DS1

Identifying Autism from Resting-State fMRI Using Long Short-Term Memory Networks

Nicha C. Dvornek[1(✉)], Pamela Ventola[2], Kevin A. Pelphrey[3], and James S. Duncan[1,4,5]

DS2

Identification of autism spectrum disorder using deep learning and the ABIDE dataset

Anibal Sólon Heinsfeld [a], Alexandre Rosa Franco [b, c, d], R. Cameron Craddock [f, g], Augusto Buchweitz [b, d, e], Felipe Meneguzzi [a, b]

DS3

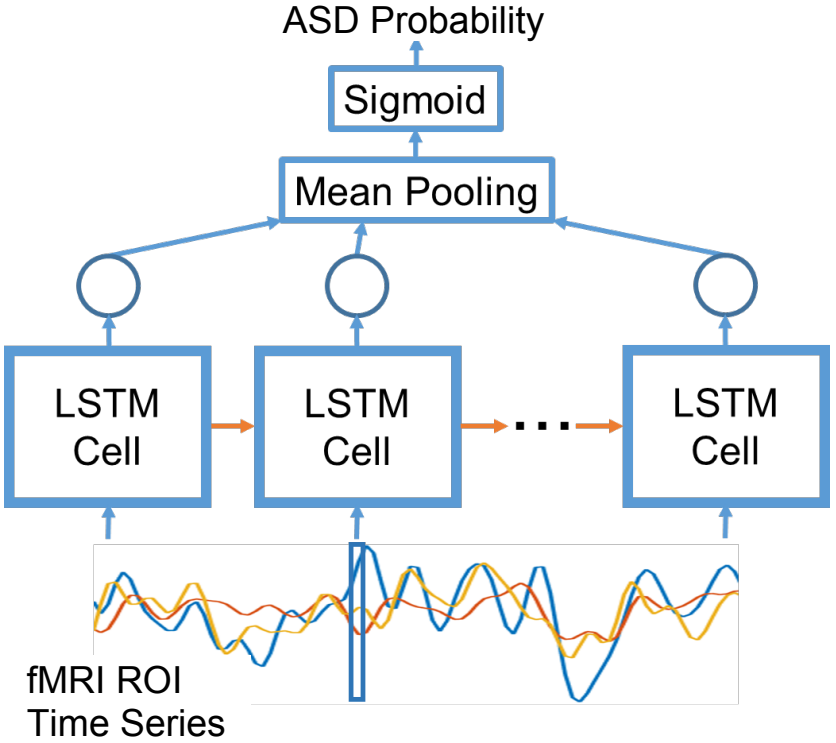
Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example

Alexandre Abraham [a, b], Michael P. Milham [e, f], Adriana Di Martino [g], R. Cameron Craddock [e, f], Dimitris Samaras [c, d], Bertrand Thirion [a, b], Gael Varoquaux [a, b]

Classification of ASD vs. Healthy Control: Methods Compared

Model
Orig† [9]
LSTM [5]
DFuse [7]
DInit [6]
DGA1-C
DGA2-C
DGA1
DGA2
DGA2-QDL

Baseline LSTM – no demographic information[1]

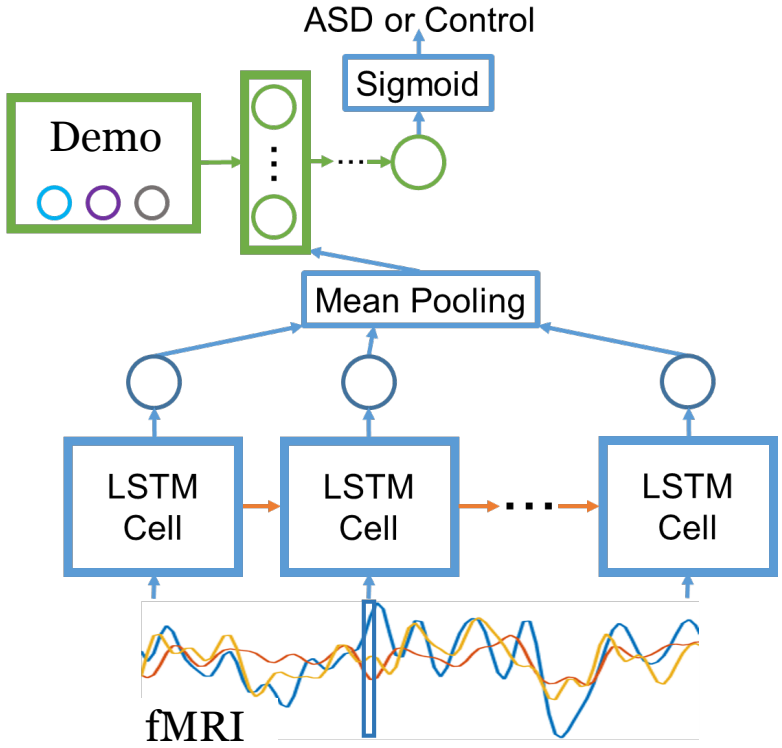


[1]Dvornek et al., MLMI 2017

Classification of ASD vs. Healthy Control: Methods Compared

Model
Orig† [9]
LSTM [5]
DFuse [7]
DInit [6]
DGA1-C
DGA2-C
DGA1
DGA2
DGA2-QDL

LSTM + late fusion of demographic information[1]

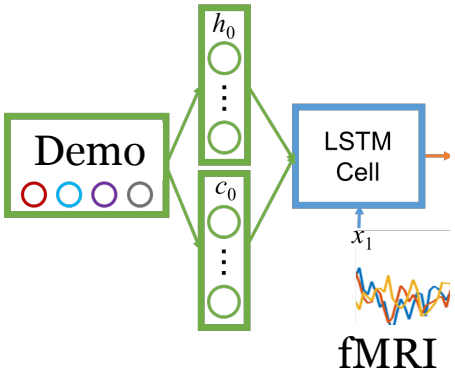


[1]Dvornek et al., ISBI 2018

Classification of ASD vs. Healthy Control: Methods Compared

Model
Orig† [9]
LSTM [5]
DFuse [7]
DInit [6]
DGA1-C
DGA2-C
DGA1
DGA2
DGA2-QDL

LSTM + state initialization via demographic information[1]

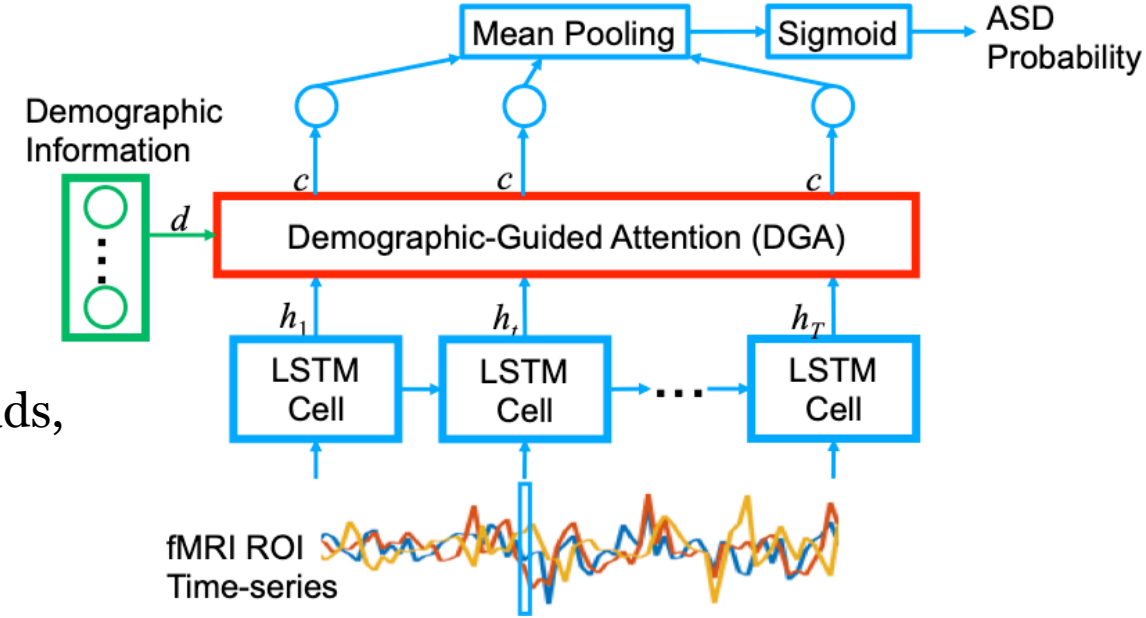


[1]Dvornek et al., MICCAI 2018

Classification of ASD vs. Healthy Control: Methods Compared

Model
Orig† [9]
LSTM [5]
DFuse [7]
DInit [6]
DGA1-C
DGA2-C
DGA1
DGA2
DGA2-QDL

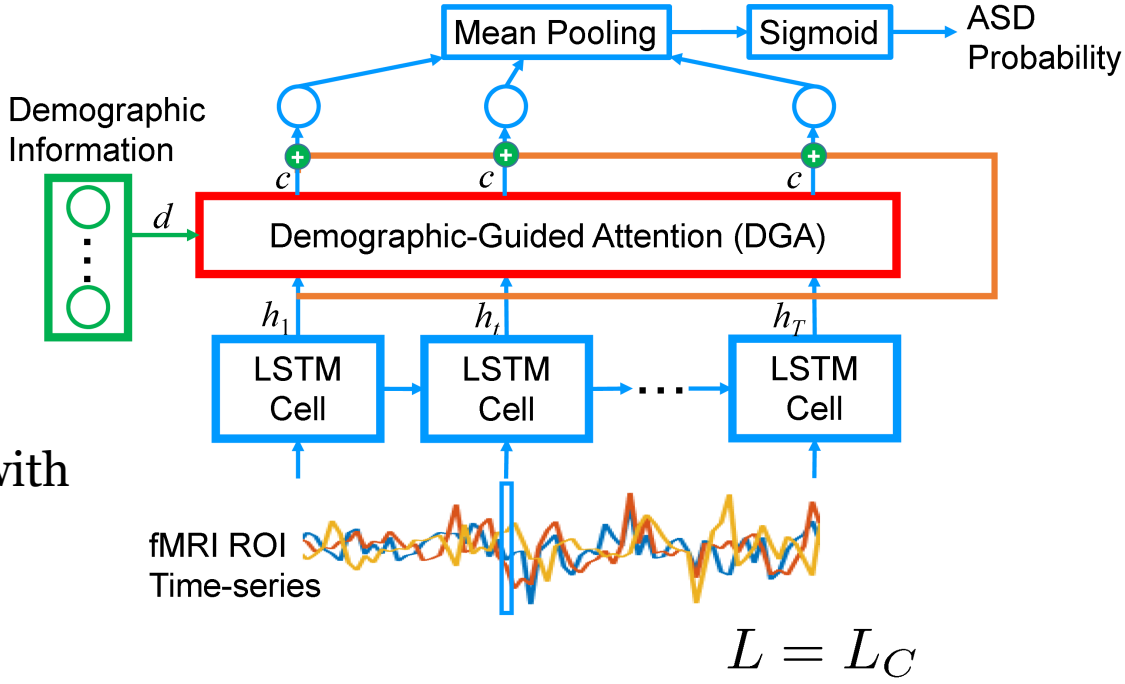
LSTM + DGA with 1 or 2 heads, context alone (no residual connection)



Classification of ASD vs. Healthy Control: Methods Compared

Model
Orig† [9]
LSTM [5]
DFuse [7]
DInit [6]
DGA1-C
DGA2-C
DGA1
DGA2
DGA2-QDL

LSTM + DGA with 1 or 2 heads (with residual connection)

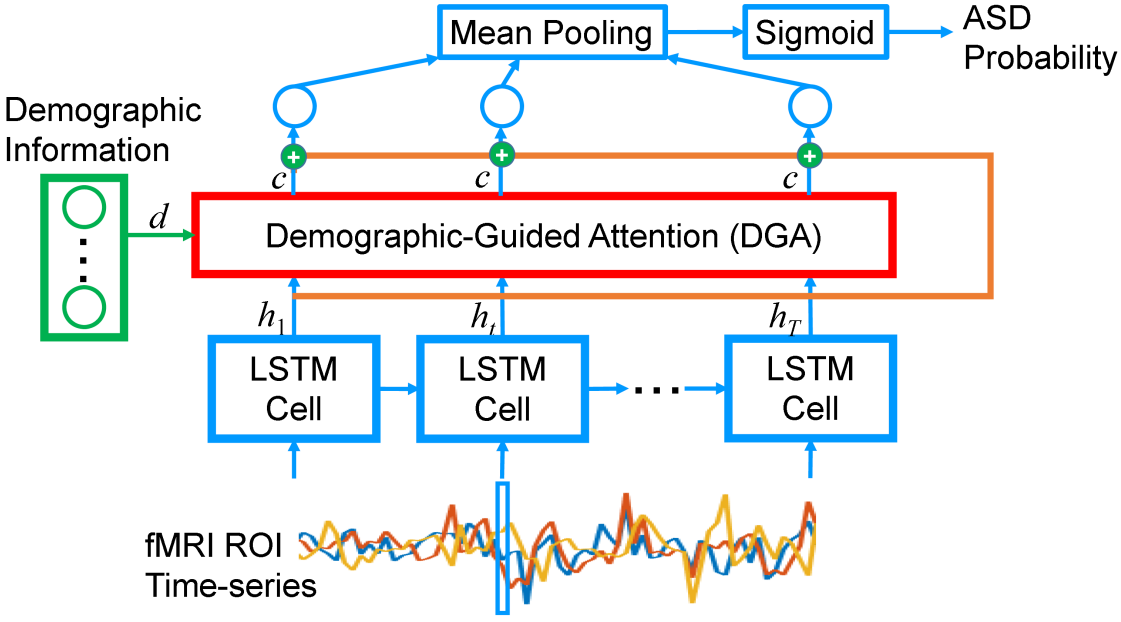


$$L = L_C$$

Classification of ASD vs. Healthy Control: Methods Compared

Model
Orig† [9]
LSTM [5]
DFuse [7]
DInit [6]
DGA1-C
DGA2-C
DGA1
DGA2
DGA2-QDL

Full model



$$L = L_C + \lambda L_{QD}$$

Classification of ASD vs. Healthy Control: Methods Compared

Model
Orig† [9]
LSTM [5]
DFuse [7]
DInit [6]
DGA1-C
DGA2-C
DGA1
DGA2
DGA2-QDL

Evaluation of implemented models

- Leave-one-site-out (LOSO) cross-validation (CV), repeated 5 times
- Averaged performance measures for each site across CV runs
- Paired two-tailed t-tests to compare models

DS2 Classification Results

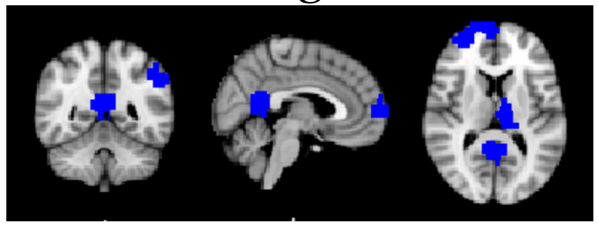
Table 2: DS2 Classification Results (N = 1035, 48.8% ASD)

Model	Leave-One-Site-Out				Weighted by # Subjects/Site		
	Mean (Std) ACC (%)	Mean (Std) TPR (%)	Mean (Std) TNR (%)	Mean (Std) AUC	Mean (Std) ACC (%)	Mean (Std) TPR (%)	Mean (Std) TNR (%)
Orig† [9]	65 (1.5)	69 (2.6)	62 (2.7)	-	65.4 (1.3)	68.1 (2.6)	62.3 (2.6)
LSTM [5]	63.6 (0.5)	55.2 (1.6)	71.9 (0.6)	0.709 (0.006)	65.6 (0.6)	58.2 (1.7)	72.7 (0.9)
DFuse [7]	65.5 (0.9) *	57.1 (0.6)	73.5 (1.6)	0.713 (0.006)	67.2 (0.6)	61.2 (1.2)	72.8 (1.0)
DInit [6]	65.8 (0.8) *	58.1 (0.4)	72.9 (1.4)	0.720 (0.009)	**67.5 (1.1)** *	61.8 (1.6) *	72.9 (3.2)
DGA1-C	65.6 (1.7) *	61.1 (1.6)	69.6 (1.1)	0.713 (0.011)	66.8 (1.6)	**64.1 (2.0)** *	69.3 (1.9)
DGA2-C	65.8 (0.9) *	52.6 (2.4)	**78.3 (1.7)** *	0.719 (0.009)	67.2 (1.2) *	55.9 (2.4)	**78.0 (0.8)** *
DGA1	66.1 (1.5) *	**61.3 (2.5)** *	70.4 (1.4)	0.719 (0.011)	67.4 (1.7) *	63.6 (2.3) *	70.9 (1.7)
DGA2	65.5 (1.0) *	54.3 (1.5)	76.5 (1.4) *	0.716 (0.015)	67.1 (1.4)	57.6 (1.3)	76.1 (2.3) *
DGA2-QDL	**66.4 (0.4)** *	58.0 (1.9) *	74.2 (2.0)	**0.722** (0.006)	67.4 (0.5) *	61.3 (1.7) *	73.1 (1.9)

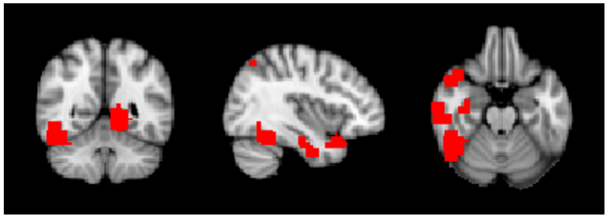
* Higher compared to LSTM with no demographics ($p < 0.05$)

† Taken from literature, reflects 1 round of LOSO CV

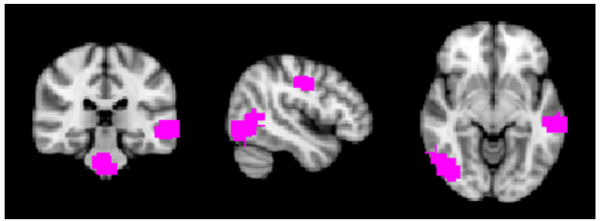
Networks with Demographic-guided Heterogeneity of Functional Processing



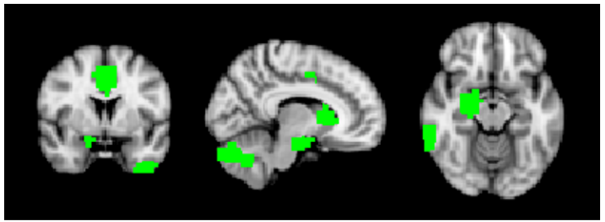
Autobiographical Memory, Default Mode



Face Recognition, Visual



Visual Perception, Face



Fearful, Happy

US Site

Yale Site

Different modes of response for functional network modulated by demographics may point to different mechanisms of ASD pathophysiology

Conclusions

- What we did:
  - Novel demographic-guided attention mechanism for modeling heterogeneity in neuropathophysiology
  - Achieved higher ASD classification performance on several ABIDE datasets under different preprocessing pipelines using LOSO CV
- What this means:
  - Improved generalization to data from new imaging sites
  - Different neural mechanisms may explain in part difficulty in classification and conflicting ASD literature
- What's next:
  - Include other phenotypic information (e.g., genetic, behavior scores)
  - Deeper analysis of changes in functional network patterns

Thank you!

- NIH Grants R01 MH100028 and R01 NS035193
- Contact: nicha.dvornek@yale.edu