Learning Generalizable Recurrent Neural Networks from Small Task-fMRI Datasets

Nicha C. Dvornek[a], Daniel Yang[c], Pamela Ventola[b], and James S. Duncan[a,d,e]

[a]Radiology & Biomedical Imaging and [b]Child Study Center, Yale School of Medicine, New Haven, CT, USA
[c]Autism and Neurodevelopmental Disorders Institute, George Washington University, Washington, DC, USA
[d]Biomedical Engineering and [e]Electrical Engineering, Yale University, New Haven, CT, USA

Yale SCHOOL OF MEDICINE

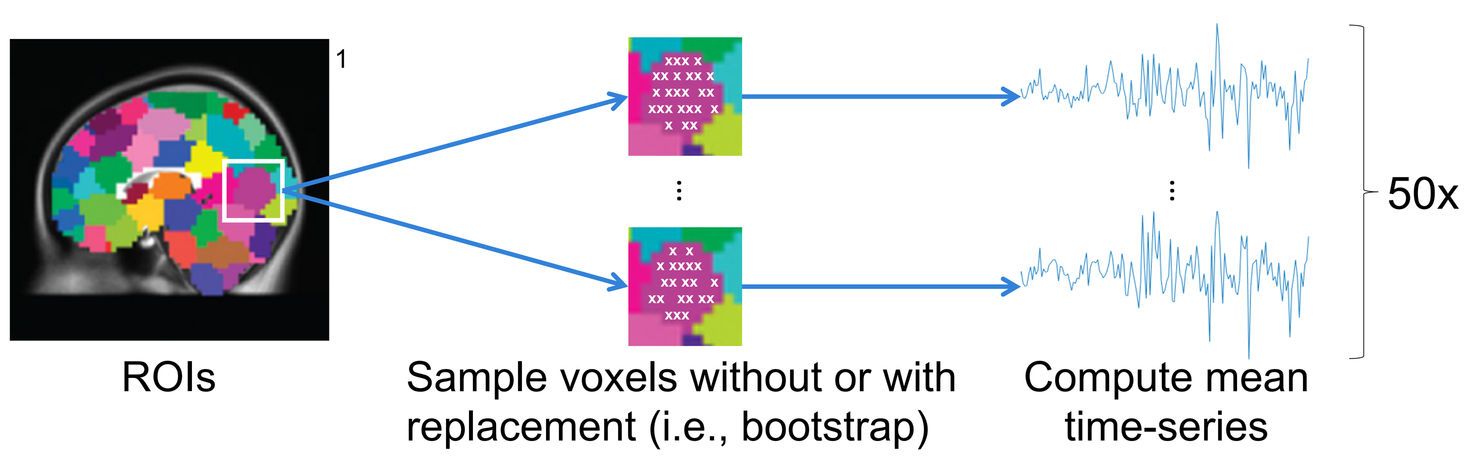
Background

- Deep learning has become state-of-the-art for many image analysis problems
- However, deep networks often require large datasets to learn effectively
- Challenge: Many medical image analysis problems have only small number of subjects available, e.g.:
  - Population constraints, e.g., disease, treatment conditions
  - Time-intensive data collection, e.g., fMRI
- Contributions: Develop approaches for deep learning from smaller fMRI datasets
  1. Data augmentation via resampling for ROI-based fMRI analysis
  2. Subject-specific initialization of LSTM using non-imaging information
  3. Model selection using criteria based on training loss

Methods: Data Augmentation by Resampling

- Standard data augmentation methods (random image croppings/rotations) not appropriate for fMRI time-series analysis
- Traditional fMRI ROI analysis extracts mean time-series from all voxels in ROI

→ Augment data by extracting mean time-series of randomly sampled voxels in ROI



Methods: LSTM-Based Network with Non-Imaging Information

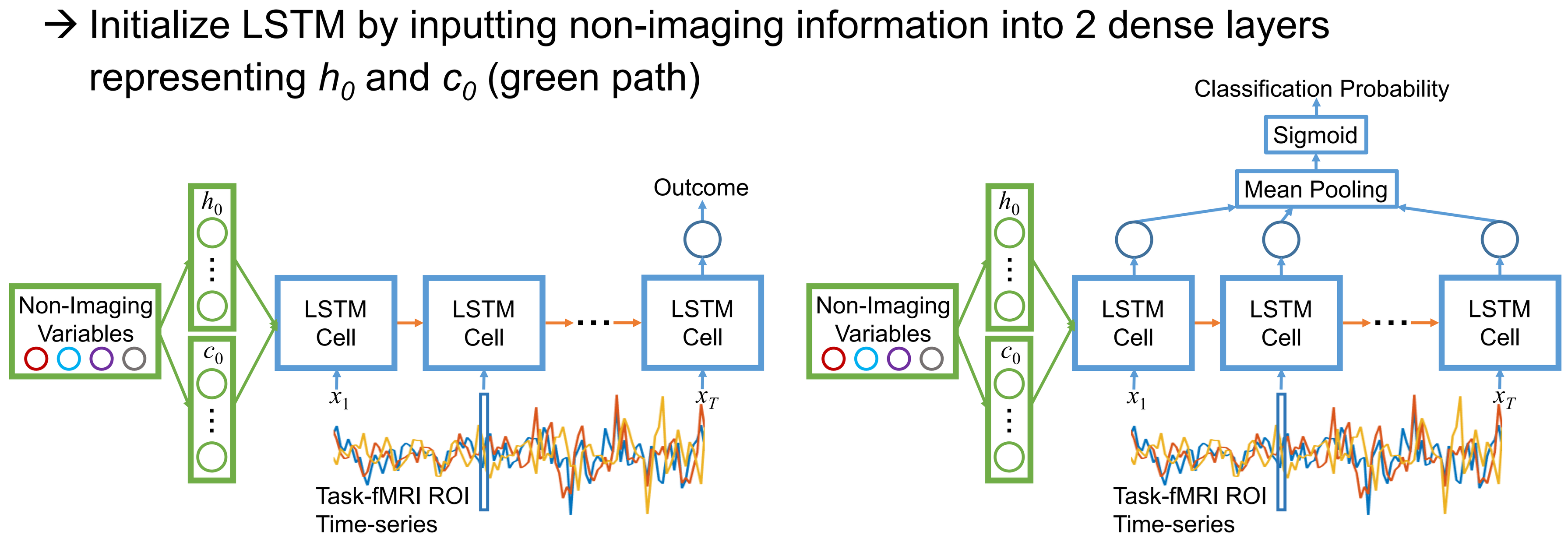
Base Network Architecture

- LSTM-based network predicts from ROI-summarized fMRI time-series
- Advantages of proposed LSTM-based model for smaller fMRI datasets:
  - Utilizes fMRI time-series data as inputs (recently proposed for classification[2])
  - ROI representation greatly reduces input dimension compared to raw fMRI data
  - Deep network with shared parameters across time → lower model complexity

LSTM Initialization

- LSTM cell contains hidden state h_t and cell state c_t
- Simple non-imaging information often available (e.g., age)

→ Initialize LSTM by inputting non-imaging information into 2 dense layers representing h_0 and c_0 (green path)



Regression Network: Analyze entire signal before making prediction

Classification Network: Output from each timestep used to make prediction

Methods: Model Selection from Training Loss

- Large datasets - choose best model from multiple training runs using validation set
- Small datasets - not enough for validation, want to use all data possible for training

→ Choose model $\hat{M}$ that learns *slowest* based on training loss criteria:

$$\hat{M} = \underset{M}{\operatorname{argmax}} \left[median\left(\Delta L_{M,s}\right) \frac{1}{L_M(0) \times s} \right]$$

Look for: Slow decline (will be negative) | High initial loss | Long relaxation time

$L_M(x)$ = training loss after epoch x for model M

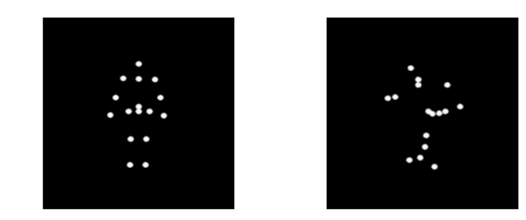
s = first epoch s.t. $L_M(s) < L_M(s)/e$

$\Delta L_{M,s}$ = first differences of loss curve from epoch 0 to epoch s

Experiments: Data and Preprocessing

Data Collection

- 21 children with autism spectrum disorder (ASD) + 19 typically-developing controls
- ASD subjects given 16 weeks Pivotal Response Therapy
- Baseline imaging:
  - MP-RAGE structural MRI
  - BOLD fMRI with biological motion perception task[3]
- Non-imaging information:
  - Baseline for all: age, sex, IQ, Social Responsiveness Scale (SRS) score
  - Post-treatment for ASD subjects: SRS score



Biological Motion | Scrambled Motion

Input Preprocessing

- fMRI images preprocessed using standardized pipeline[4]
- Standardized time-series extracted from each cerebral ROI of AAL atlas[5]
- Normalized each non-imaging variable to [-1,1]

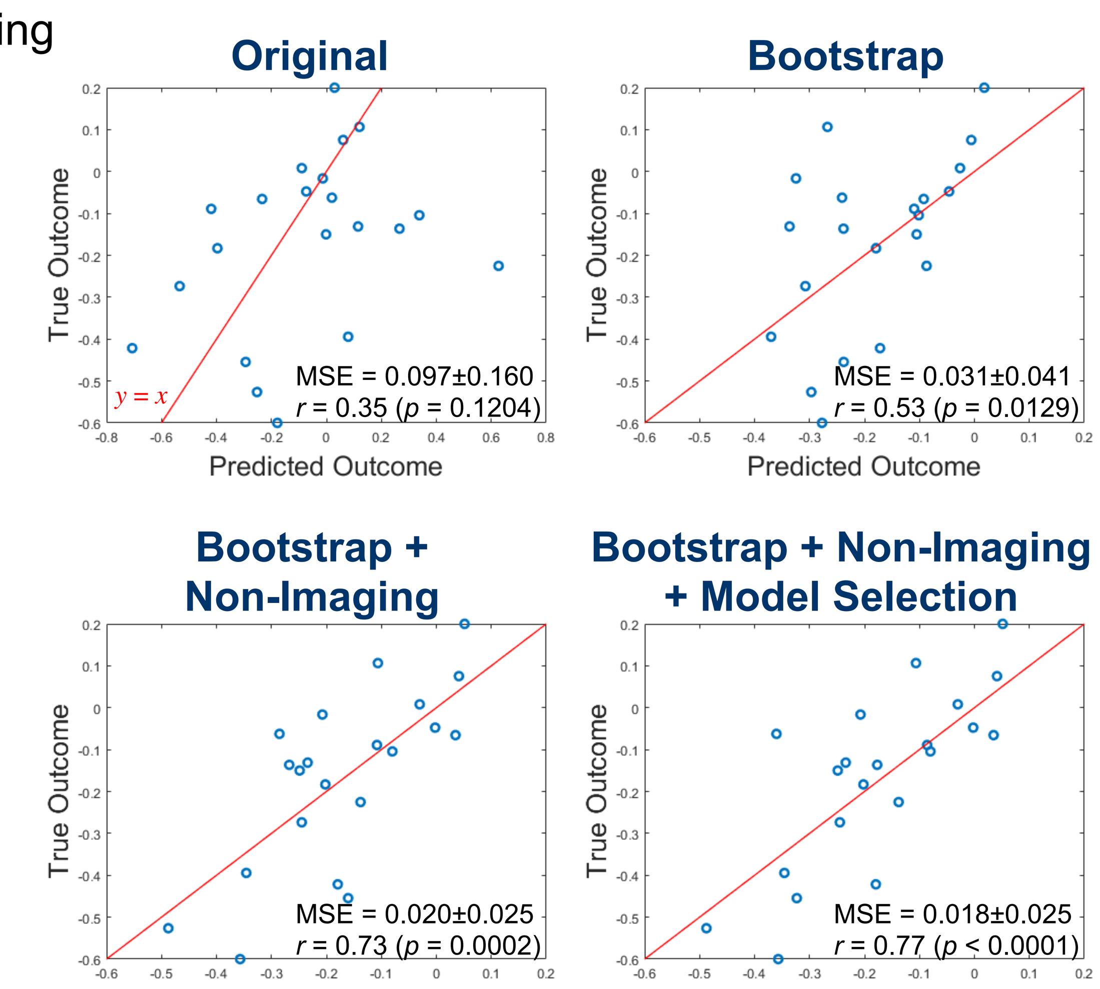
Experiments: Regression Task

Experimental Setup

- Goal: Predict treatment outcome (percent change in SRS after treatment) (N = 21)
- Evaluation: Leave-one-out cross-validation (CV), one-tailed paired t-tests (p < 0.05)

Results

- Augmented dataset 50x using: 1) Data repetition, 2) Standard Gaussian noise addition, 3) Proposed resampling
  - Data repetition did not significantly reduce MSE
  - Noise addition and resampling significantly reduced MSE compared to original dataset
  - No significant differences between any noise and sampling methods
- Non-imaging information in LSTM initialization performed better than in standard top-level multi-modal fusion[6]
- Model selection from 2 separate runs performed better than model bagging



Experiments: Classification Task

Experimental Setup

- Goal: Classify ASD vs. typically-developing subjects (N = 40)
- Evaluation: 10-fold CV repeated 10x, one-tailed paired t-tests (p < 0.05)

Results

Method	Mean (SD) Accuracy (%)	Mean (SD) TPR (%)	Mean (SD) TNR (%)
Original	51.8 (3.3)	56.1 (13.3)	55.1 (12.4)
Bootstrap	64.5 (5.1) *	70.7 (7.3) *	60.9 (11.1)
Bootstrap + Non-Imaging	67.5 (6.7) *	72.2 (9.2) *	64.6 (6.3) *
Bootstrap + Non-Imaging + Model Select	69.8 (5.5) *^†	75.1 (8.4) *	65.5 (6.8) *

* Significantly better than original dataset. ^ Significantly better than bootstrap dataset.
† Significantly better than at least one individual model.

Conclusions

- Our learning strategies for small datasets produced more generalizable models
  - Data augmentation via bootstrap sampling requires no parameter selection
  - LSTM initialization with non-imaging information incorporates more subject-specific variation at small cost
  - Model selection from training loss alone maximizes amount of data for learning

References 1. Craddock et al., *Hum. Brain Mapp.* 2012. 2. Dvornek et al., *MLMI* 2017. 3. Kaiser et al., *PNAS* 2010. 4. Pruim et al., *NeuroImage* 2015. 5. Tzourio-Mazoyer et al., *NeuroImage 2002.* 6. Dvornek et al., ISBI 2018.

This work was supported by NIH grants R01NS035193 and R01MH100028.

Contact: nicha.dvornek@yale.edu

MICCAI 2018 Granada SPAIN