Biologically Representative Machine-Learning-Based Emotion Recognition Circuit for Modeling Social Impairment Disorders

Annika Sandholm[1], Gavin Tardif[2], Mark Cai[3], Travis Hamerling[4], Alexandre Douieb[5]

Shady Side Academy, 423 Fox Chapel Rd, Pittsburgh, PA 15238[1]; Eastside Preparatory School 10613 NE 38th Pl, Kirkland, WA 98033[2]; Council Rock High School North, 62 Swamp Road, Newtown, PA 18940[3]; Bronxville High School, 177 Pondfield Rd, New York, NY 10708[4], Lycée Français de New York, 505 East 75th Street, New York, NY 10021[5]

INTRODUCTION

- **Emotion recognition is a cognitive ability allowing humans to interpret facial cues and audio tones**
- **Autism Spectrum Disorder (ASD)** and **Schizophrenia (SZA)** significantly impair this ability, affecting 75 million and 24 million people, respectively
- Despite growing research, the neural basis for altered emotion recognition in schizophrenia and ASD remains **poorly understood**
- Current artificial intelligence models can classify emotions, but **most lack biological plausibility**

Purpose: **Model how ASD and Schizophrenia affect emotional perception via emotion-aware and biologically realistic machine learning**

We created a **4-step circuit** that can classify one of **6** emotions based on audiovisual data

We used a novel **Spiking Neural Network (SNN)** to mimic the brain's computation

We then modified parameters to simulate ASD and Schizophrenia
- better visualize and understand errors
- can qualitatively determine the effects of certain parameters
- step toward computational psychiatry - predicting behaviors of clinical populations

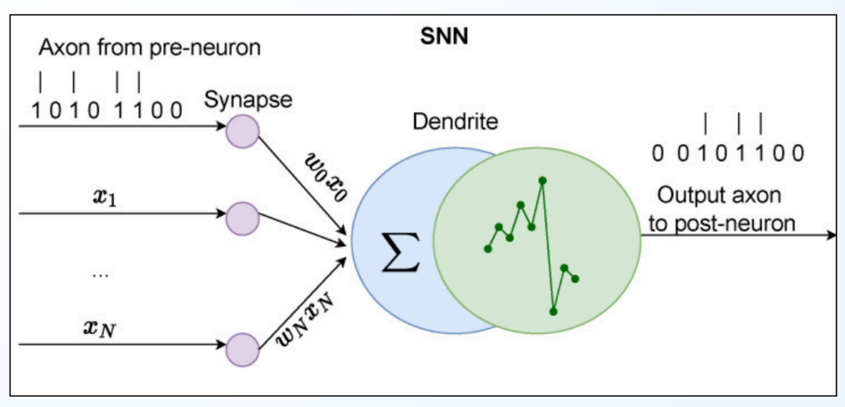


Figure 1. Representation of the architecture of a Spiking Neural Network[1]

DATASET

- Used **CREMA-Dataset** (Cao et. al., 2014)[2]: contains labeled audio and visual recordings of actors expressing 6 different emotions - happiness, anger, disgust, sadness, fear, and neutral
- 7441 audiovisual recordings, average length of **2.6 seconds**
- **91 actors** - 48 male and 43 female - consists of a variety of ethnicities
- **12 different sentences**
- Video clips in the **.flv** format, audio clips in the **.wav** format

METHODS

1. We divided the emotion recognition circuit in the brain into **4 components:**
   a. Visual feature extraction using **Facenet, a Convolutional Neural Network (CNN)**, mimicking the **visual cortex**
   b. Audio feature extraction (MFCCs, energy, pitch) using **Librosa**, mimicking the **auditory cortex**
   c. Features standardized and combined in the **Superior Temporal Sulcus (STS)** layer
   d. Combined features are passed to our **amygdala**, a **Spiking Neural Network (SNN)** that **classifies the final emotion**
      i. Used SNNTorch and PyTorch, Cross Entropy Loss Function, Adam Learning Algorithm, learning rate=0.001
      ii. Two hidden layers with 512 and 256 nodes, respectively
      iii. 50 epochs, batch size=128
      iv. 2nd-Order Integrate-and-Fire Neurons with Synaptic Conductance
2. Performed a **5x5-fold cross-validation** to test the variance in the network's performance across runs
3. Built on model system to simulate Autism and Schizophrenia models by altering biological

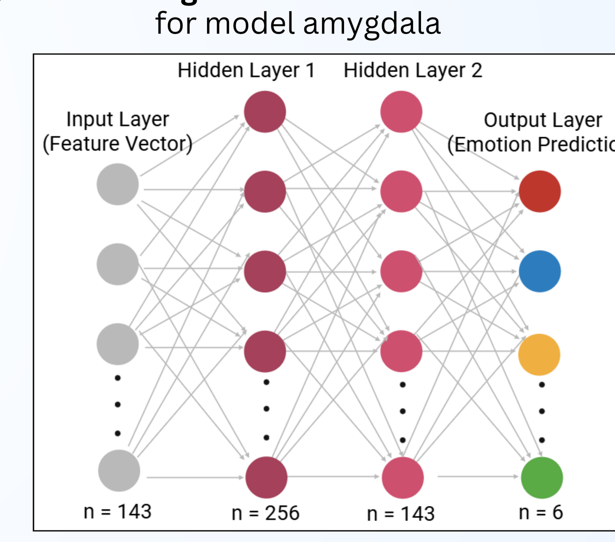
Autism Spectrum Disorder (ASD)

Hyperactivation in the amygdala, issues in sensory integration[3,4], impaired plasticity, detached temporal processing[5], and overconnectivity[6]

Schizophrenia (SZA)

Hypoactivation in the amygdala (blunted emotional response), uncontrolled synaptic pruning[8], and dysregulated and stochastic synaptic transmissions[7]

Figure 2. SNN architecture for model amygdala



Our Multisystem Model:

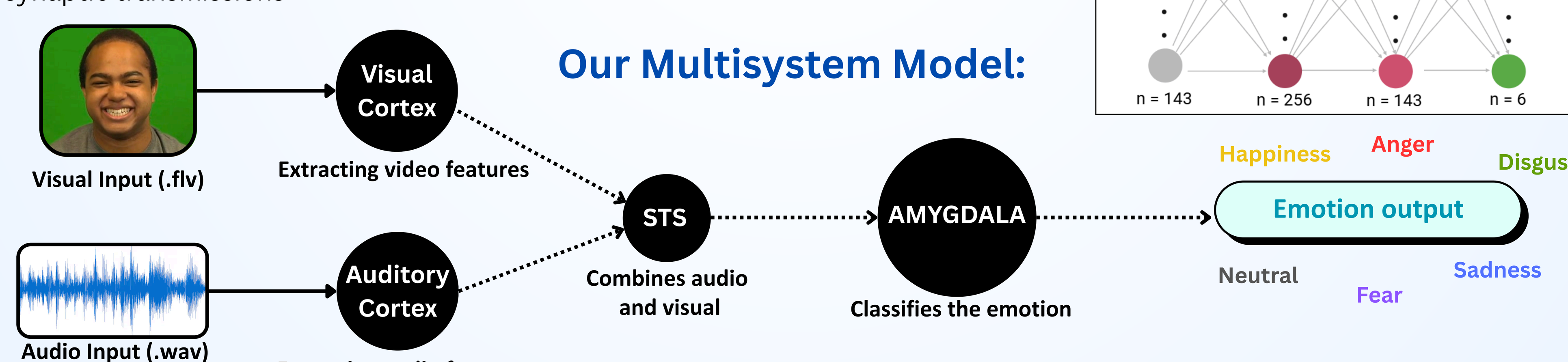
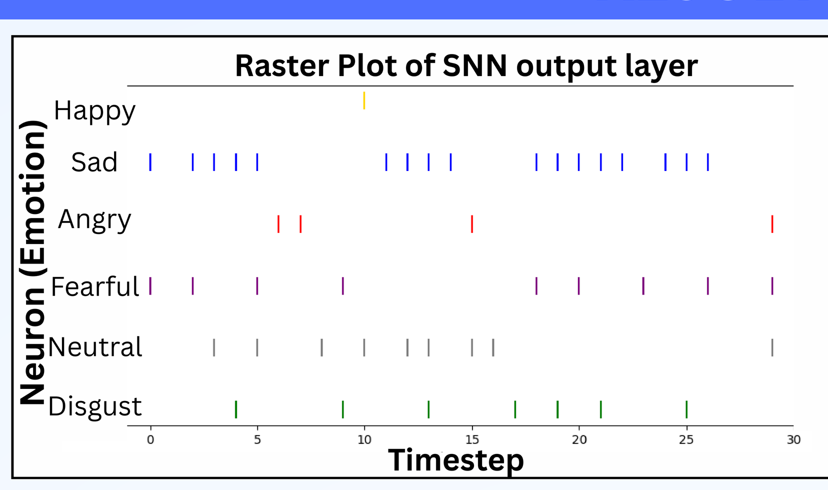


Figure 3. 4-Step Multisystem model

RESULTS

Example Output

Figure 4. Raster Plot showing spiking of the 6 output neurons



True Label:	Sad
Predicted class:	Sad
Happy:	0.00%
Sad:	99.93%
Angry:	0.00%
Fearful:	0.03%
Neutral:	0.03%
Disgust:	0.00%

Figure 5. Softmax output showing the predicted probabilities for each emotion.

Baseline Model

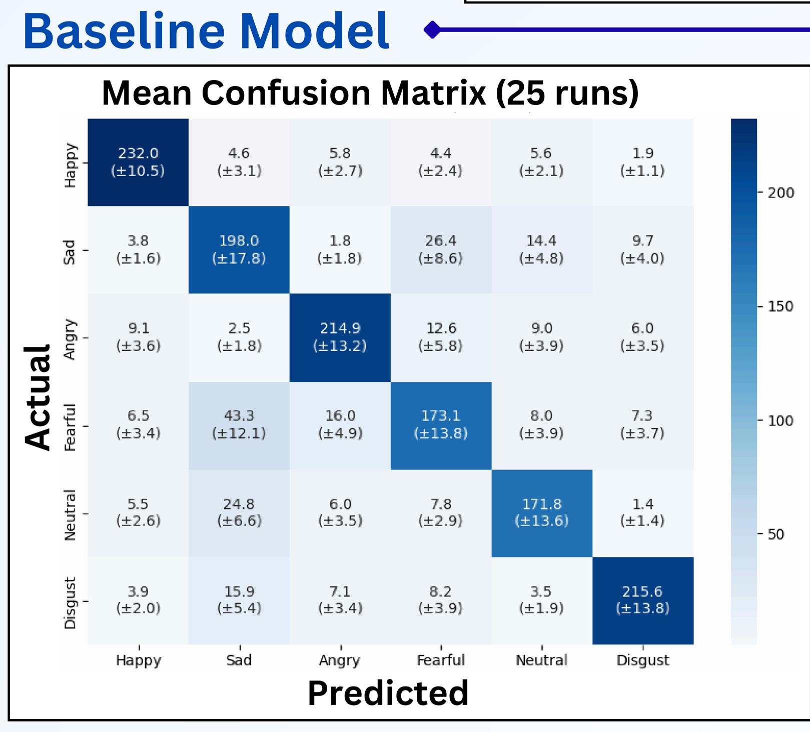


Figure 6. Mean confusion matrix over 25 runs

Accuracy (%): 81.97 ± (1.26)

Class	F1-Score (Mean ± Std Dev)
Happy	0.901 ± 0.018
Sad	0.729 ± 0.023
Angry	0.850 ± 0.013
Fearful	0.711 ± 0.025
Neutral	0.799 ± 0.02
Disgust	0.869 ± 0.016

Figure 7. F1 (function of precision and recall) score for each emotion.

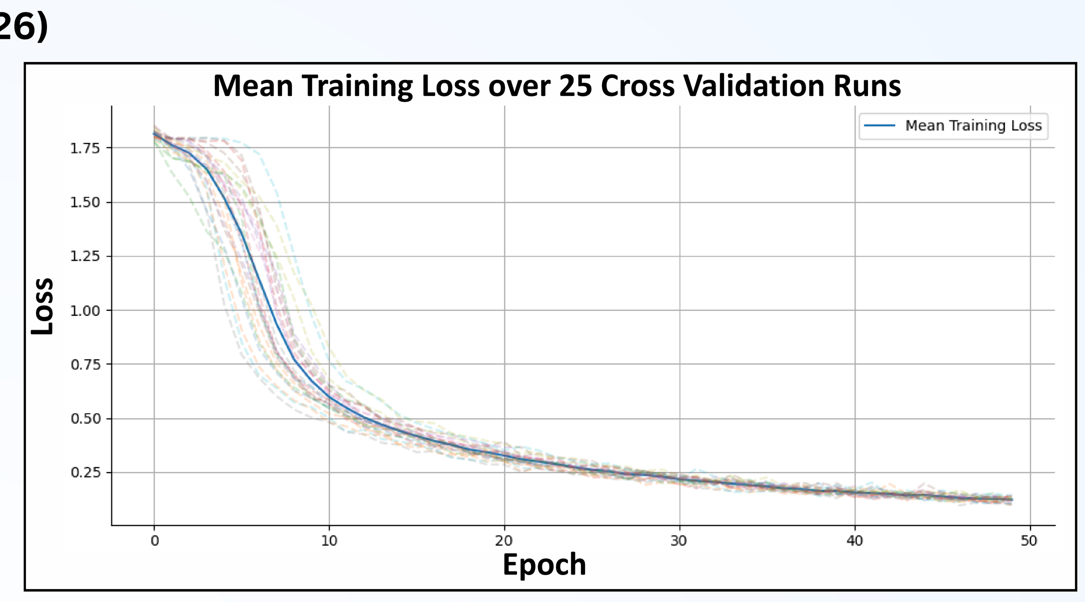


Figure 8. Mean training loss over 50 epochs

ASD

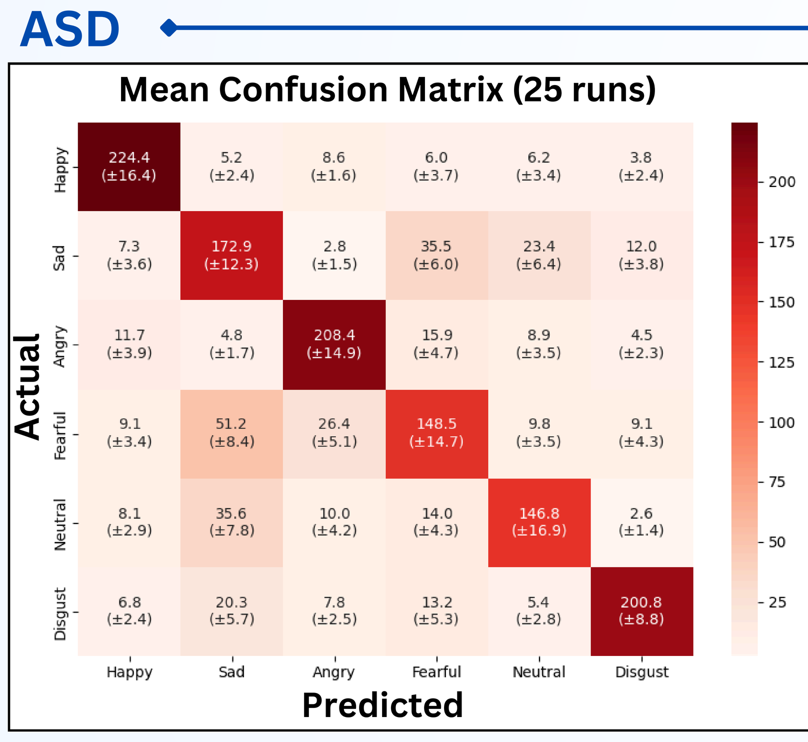


Figure 9. Mean confusion matrix of the ASD model over 25 runs

Accuracy (%): 74.14 ± (1.11)

Class	F1-Score (Mean ± Std Dev)
Happy	0.8602 ± 0.0190
Sad	0.6355 ± 0.0175
Angry	0.8040 ± 0.0173
Fearful	0.6087 ± 0.0271
Neutral	0.7014 ± 0.0301
Disgust	0.8246 ± 0.0124

Figure 10. Mean F1 score for each emotion of ASD model

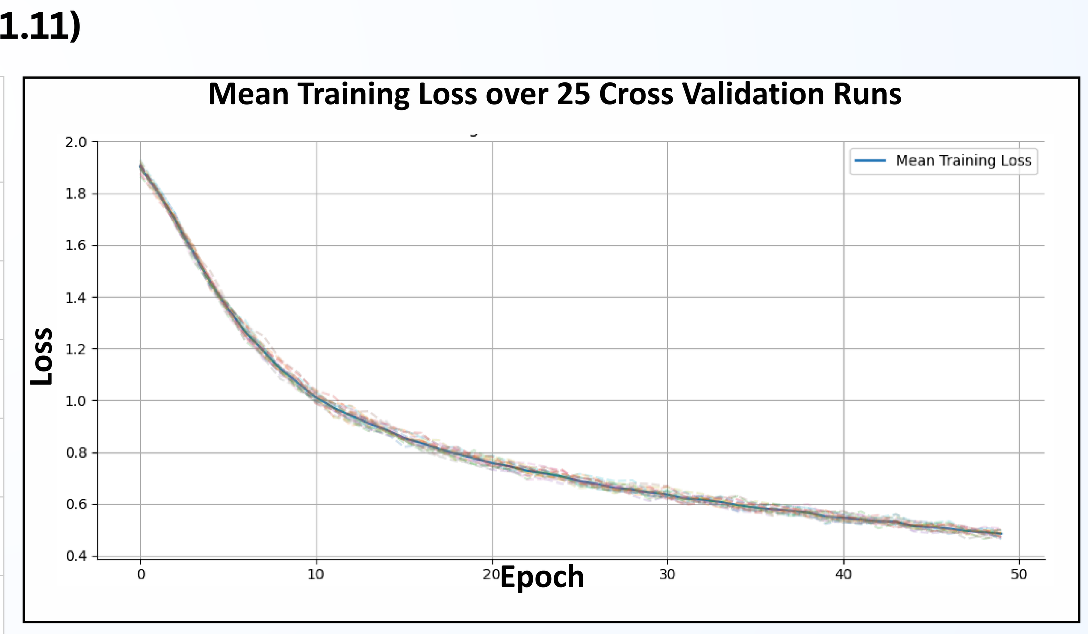


Figure 11. Mean training loss of ASD model over 50 epochs

Schizophrenia

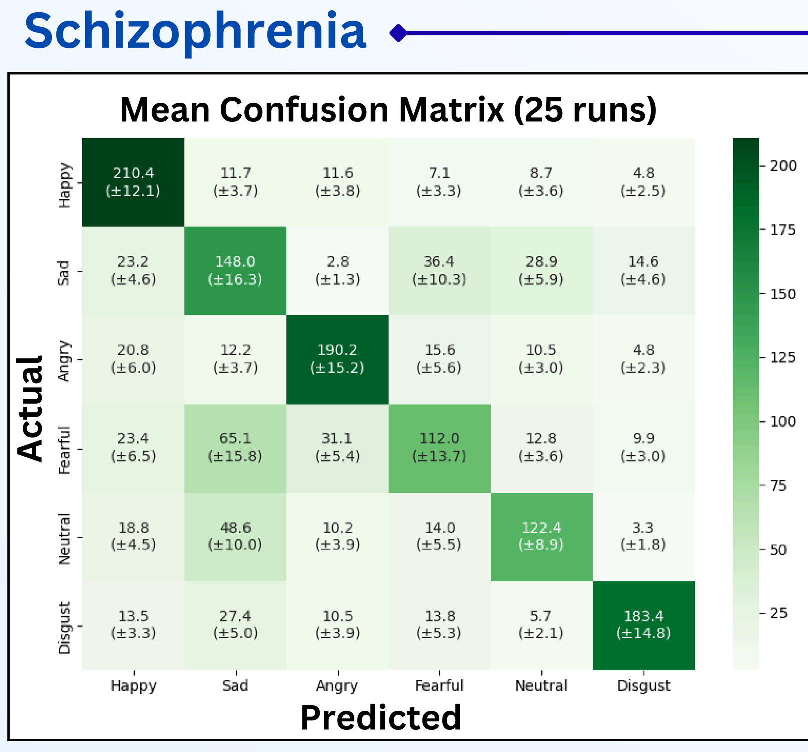


Figure 12. Mean confusion matrix of the SZA model over 25 runs

Accuracy (%): 73.70 ± (1.18)

Class	F1-Score (Mean ± Std Dev)
Happy	0.7457 ± 0.0188
Sad	0.5210 ± 0.0284
Angry	0.7444 ± 0.0199
Fearful	0.4929 ± 0.0309
Neutral	0.6021 ± 0.0251
Disgust	0.7714 ± 0.0223

Figure 13. Mean F1 score for each emotion of SZA model

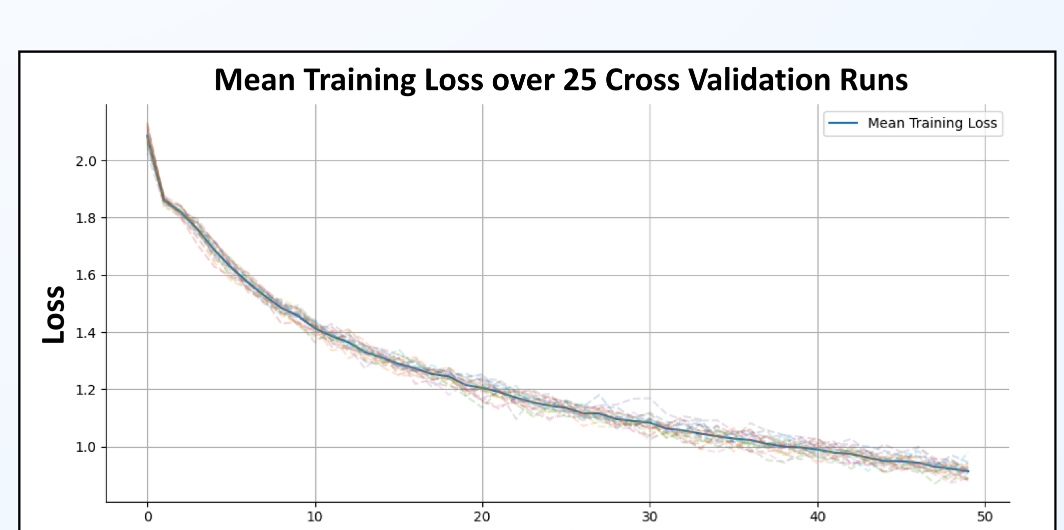


Figure 14. Mean training loss of SZA model over 50 epochs

DISCUSSION

Conclusions

- The biologically-grounded model reaches high performance: **81.97 ± (1.26) % accuracy** on never-before-seen clips
- The Autism Spectrum Disorder Simulation and the Schizophrenia Simulation both performed **significantly lower** in accuracy over 25 runs ($p < 0.001$ for both)
- Most confusion occurs between **Fearful ↔ Sad** and **Sad ↔ Neutral** across all 3
  - On the other hand, **Happy and Disgust** are consistently classified with the least errors
- All 3 models are **equally variable** according to their SD's
- Autism Spectrum Disorder
  - **Uniform declines** across all 6 emotions
  - Out of the changed parameters, the impaired sensory integration had the most impact on accuracy
- Schizophrenia
  - F1 declines are more heterogeneous, with **Sad and Fearful suffering the worst** (−0.21 to −0.22), while Angry and Disgust are comparatively less impacted (−0.10)
    - Mirrors biological findings[10]
- Both the ASD and Schizophrenia training loss graphs have not converged by epoch 50
  - Baseline graph converges sooner, meaning it finishes learning about the training data quicker[9]
  - Reflects shorter development time for neurotypical patients

Limitations

- **Simplified brain architecture**: 917 nodes total and only 4 components
- **Supervised Learning**: Neuron layers built on top of traditional node layers
- **Generalizability**: High accuracy on CREMA-D may not translate to other datasets or more naturalistic stimuli (e.g., real conversations, body language)

Future Work/Applications

- **Help therapists, educators, and caregivers** have a better understanding of how neurodivergent individuals process faces/emotions
- Incorporate **more biologically realistic learning algorithms**
  - E.g. unsupervised learning, synaptic plasticity, synaptic pruning
- Combine with in vivo studies to **quantify the extent/effect of traits associated with autism** and how they relate to general cognitive processing

ACKNOWLEDGEMENTS

We are sincerely grateful to the **RISE program** at Boston University for offering us this valuable research experience. Our thanks go to **Karla Montejo**, **Shankar Ramachandran, Pier Bressan**, and the entire team of **teaching fellows** for their insightful guidance and support. Lastly, we deeply appreciate **our families** for their unwavering encouragement throughout our scientific endeavors.

REFERENCES