EarlyScreen:
Multi-scale Instance Fusion for Predicting Neural Activation and Psychopathology in Preschool Children

Manasa Kalanadhabhatta, Adelys Mateo Santana, Zhongyang Zhang, Deepak Ganesan, Adam S. Grabell, Tauhidur Rahman
Emotion Regulation in Early Childhood

Integral part of early childhood development.

Poor emotion regulation predicts range of psychological disorders (e.g., ADHD, mood and anxiety disorders).

Hard to distinguish from normative misbehavior (e.g., temper tantrums).
Current Diagnostic Practices

Clinical questionnaires (e.g., CBCL) filled by parents
- Modest diagnostic accuracy (sensitivity 0.66, specificity 0.83).
- Indirect/biased measure.

Comprehensive interactions or observations of the child
- Time-consuming – ADOS requires 2-3 hour-long sessions.
- Long waitlists and cost burdens.

Do not consider neurological underpinnings of symptoms
- Association between clinical irritability and low activation of lateral prefrontal cortex.
- No access to neuroimaging in the field.

EarlyScreen

Screening Tool for Emotion Dysregulation in Preschool Children

Direct Measurement of Child Behavior
- Record behavioral correlates including facial expressions and movement.
- Less burdensome and fewer obstacles for parents.

Rapid and Accessible
- Less than 10 minutes of behavioral data.
- Future home deployment on a tablet with a camera.

Neural Information in the Wild
- Predict IPFC activation during frustration.
- Additional modality for clinical diagnosis.
EarlyScreen Study

- 94 participants aged 3.5 to 5 years completed a clinically-validated frustration inducing task.
- Parents completed diagnostic questionnaires about child’s behavior.

Simultaneous neural and behavioral recording:
- Facial expressions, eye and head movement using video cameras.
Indicators of Early Psychopathology

Low lateral prefrontal cortex activation during frustration

- Measure change in oxyhemoglobin ($\Delta$HbO$_2$) during frustration blocks
- Lower $\beta$ indicates poor emotion regulation
  => higher risk of psychopathology

Clinical symptoms based on parental questionnaires

- Four subscales that measure most prevalent childhood problems
- Above clinical threshold on at least one scale
  => clinical categorization
Behavioral Correlates

- **Facial Expressions**
  - Presence and Intensity of 18 discrete Action Units
  - Positive and Negative expression categories

- **Eye Gaze**
  - Change in gaze angle

- **Head Movement**
  - Change in head position relative to camera

**Grouped into 6 Feature Sets**

- Action Units (AUs)
- Facial expressions
- Movement (eye and head)
- AU + movement
- AU + Facial expressions
- AU + Movement + Facial Expressions
Classification Baselines and Features

- **Evaluate feasibility**: can behavioral features predict neural activation and clinical risk?
- **Feature set and model selection**: 5-fold cross-validation area under ROC curve

---

Frames from 4-sec feedback segments

Sequence of video frames

Feature Extraction Module

Aggregate features from all feedback segments

Supervised Learning Module

*Neural Activation or Clinical Status Label*
Classification Baselines and Features

Neural Activation
Baseline performance: 0.80 AUROC
Best features: AU features
Best model: Random Forest

Clinical Status
Baseline performance: 0.77 AUROC
Best features: AU + Movement + Facial Expressions
Best model: AdaBoost
Improving Predictions

- **Limitation:** Single label per individual => small number of samples.
- **Opportunity:** Multiple trials per individual => opportunity to learn from fine- and coarse-grained data.

Multiple Instance Learning

![Diagram of instance learning](image)

Standard MIL Assumption:

\[
Y_i = \begin{cases} 
+1 & \text{if } \exists y_{ij} : y_{ij} = +1 \\
-1 & \text{if } \forall y_{ij} : y_{ij} = -1 
\end{cases}
\]

EarlyScreen Task Setting:

- No instance-level labels $y_{ij}$.
- Need to learn bag-level labels independently.
Proposal

- MIL pipeline with a bag representation module, learning transformation from instance-level feature space to bag-level space:
  \[ f : B_i \rightarrow B_i^\phi \]

- Use \( B_i^\phi \) as input to supervised model to extract \( Y_i \).

- Combine MIL pipeline with the baseline “single instance learning” pipeline.
Multi-scale Instance Fusion Framework

Single Instance Learning Pipeline

- Feature Extraction Module
- Aggregate features from all feedback segments
- Supervised Learning Module

Multiple Instance Learning Pipeline

- Feature Extraction Module
- Instance-level features
- Bag Representation Module
- Bag-level features
- Supervised Learning Module

Frames from all feedback segments

‘Bag’ containing feedback ‘instances’

Sequence of video frames

Feature Extraction Module

Class Probabilities

Weighted Fusion

Predicted Label
Classification Performance of MIF Models

**IPFC Activation:** AUROC of 0.85 (compared to 0.80 for SIL model)

Using AU features and polynomial minimax kernel for bag representation.

**Clinical Status:** AUROC of 0.80 (compared to 0.77 for SIL model)

Using AU + Movement + Facial Expressions and MInD mapping for bag representation.
Demographic Fairness

IPFC Activation

Clinical Status
Asked 60 mental health practitioners about

- current diagnostic practices.
- utility of EarlyScreen and concerns regarding deployment.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree</th>
<th>Somewhat disagree</th>
<th>Neither agree nor disagree</th>
<th>Somewhat agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>As a field, we need to improve the accuracy, efficiency, and convenience of how we diagnose early childhood mental illness.</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>41</td>
</tr>
<tr>
<td>A patient's biological data could someday improve the accuracy of their diagnosis.</td>
<td>0</td>
<td>2</td>
<td>11</td>
<td>31</td>
<td>9</td>
</tr>
<tr>
<td>Inexpensive and widely available neuroimaging data could someday improve diagnostic accuracy.</td>
<td>3</td>
<td>5</td>
<td>15</td>
<td>19</td>
<td>11</td>
</tr>
<tr>
<td>Continuous behavioral data collected in home settings could someday improve diagnostic accuracy.</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>24</td>
<td>23</td>
</tr>
<tr>
<td>In the future, using home-based games such as EarlyScreen can help provide useful diagnostic information that can add to currently available methods.</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>30</td>
<td>12</td>
</tr>
<tr>
<td>The preliminary accuracy of EarlyScreen’s models is encouraging for future tests and subsequent deployment as an additional diagnostic tool.</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>25</td>
<td>15</td>
</tr>
<tr>
<td>Apps such as EarlyScreen could be useful for collecting ecologically valid data in home settings.</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>31</td>
<td>13</td>
</tr>
<tr>
<td>I am concerned about the data privacy of such an application.</td>
<td>2</td>
<td>10</td>
<td>10</td>
<td>20</td>
<td>7</td>
</tr>
<tr>
<td>I am concerned about the ethical considerations behind such an application.</td>
<td>3</td>
<td>10</td>
<td>11</td>
<td>19</td>
<td>6</td>
</tr>
</tbody>
</table>
Next Steps

Testing on larger populations and in diverse conditions.

At-home deployment using a tablet-based game.

Integrating wearable sensor data for more accurate predictions.

Image source: Empatica (https://www.empatica.com/blog/considering-embrace2-for-your-child.html)
For more details, see our paper

“EarlyScreen: Multi-scale Instance Fusion for Predicting Neural Activation and Psychopathology in Preschool Children”