



Towards Accurate and Scalable Mental Health Screening Technologies for Young Children

Manasa Kalanadhabhatta
University of Massachusetts Amherst
Amherst, MA, USA
manasak@cs.umass.edu

Deepak Ganesan
University of Massachusetts Amherst
Amherst, MA, USA
dganesan@cs.umass.edu

Tauhidur Rahman
University of California San Diego
San Diego, CA, USA
trahman@ucsd.edu

ABSTRACT

Mental, emotional, and behavioral disorders are highly prevalent in preschool-aged children and can significantly affect their social-emotional development and adaptive functioning. However, identifying signs of problematic behavior at this age is extremely challenging due to several structural and phenomenological barriers. This work leverages mobile and wearable devices to build accurate, usable, and scalable assessment tools that can be deployed in home settings to screen for common disorders in young children. It describes the development of novel screening algorithms that utilize behavioral and neurophysiological signals recorded during brief, naturalistic tasks, and presents stakeholder perspectives toward the usability and clinical utility of such screening tools.

CCS CONCEPTS

• **Human-centered computing** → Ubiquitous and mobile computing; • **Applied computing** → Health informatics.

KEYWORDS

mental health; behavioral disorders; diagnostic assessments; brain sensing; wearable computing; mobile health

ACM Reference Format:

Manasa Kalanadhabhatta, Deepak Ganesan, and Tauhidur Rahman. 2023. Towards Accurate and Scalable Mental Health Screening Technologies for Young Children. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing (UbiComp/ISWC '23 Adjunct)*, October 08–12, 2023, Cancun, Quintana Roo, Mexico. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3594739.3610763>

1 INTRODUCTION

Nearly one in five children in the United States are impacted by mental, emotional, or behavioral disorders such as attention-deficit / hyperactivity disorder (ADHD), autism spectrum disorder (ASD), disruptive behavior disorder, etc. [18]. Poor mental health impairs children's social-emotional development, diminishes learning outcomes, and persists into later stages of development [12].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
UbiComp/ISWC '23 Adjunct, October 08–12, 2023, Cancun, Quintana Roo, Mexico
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0200-6/23/10...\$15.00
<https://doi.org/10.1145/3594739.3610763>

A core feature of some of the most commonly diagnosed pediatric psychological disorders is poor emotion regulation, or the inability to appropriately modulate the duration, valence, or intensity of an emotional experience, particularly in response to negative emotional challenges such as frustration [5]. It is also the most common reason young children are referred to psychological services [3]. Therefore, there is particular interest in identifying and treating emotion dysregulation at an early age, when it has a profound effect on children's capacity to function adaptively in school, home, and social environments.

1.1 Problem Statement

Screening for emotion dysregulation-related disorders, especially at the preschool age, can ensure that children receive timely interventions that can improve long-term outcomes [18]. However, signs of problematic emotion regulation are often difficult to distinguish from normative misbehavior (e.g., temper tantrums) in young children [39]. This creates a “when to worry” problem, where caregivers lack guidelines to determine the severity and clinical significance of a child's behavior. Attaining a formal diagnosis typically requires families to overcome several barriers in order to seek specialized psychological care, including navigating a severely under-resourced mental healthcare system, logistical and financial burdens, stigma and/or poor mental health literacy among caregivers, etc [8]. Moreover, “gold standard” diagnostic tools require specialized training and lengthy sessions to administer [33] and have limited accuracy [41]. These screening gaps have a significant detrimental impact on children with treatable disorders, with about half of them failing to receive the required care [42]. The development of novel, highly scalable assessments that can be delivered in routine primary health care or home settings is imperative to expanding access to care [12].

1.2 Thesis Overview

My thesis posits that mobile and wearable devices can enable accurate, usable, and scalable early mental health screening via multi-modal sensing. In particular, I aim to demonstrate that this can be achieved during brief, naturalistic tasks that are familiar and less burdensome for children than traditional methods of assessment. To support this hypothesis, I will describe the development and evaluation of two at-home screening tools that use a frustration-inducing tablet-based game and in-the-wild parent-child interactions respectively as an assessment medium to detect emotion dysregulation.

Figure 1 presents an overview of the proposed screening tools. The first, called EarlyScreen [26], records children's facial expressions and movements in response to frustration via videos and uses these to predict emotion dysregulation-related brain activity

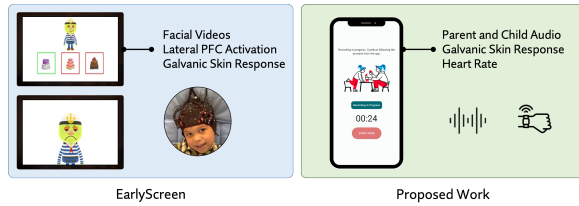


Figure 1: Overview of the assessment tasks and behavioral/physiological signals used to predict common disorders in the two at-home screening tools described in this work.

and disorder risk. My work also shows how brain activity and galvanic skin response during emotion regulation can help improve the accuracy of psychopathology prediction. Additionally, based on iterative feedback from diverse stakeholders involved in child mental health care, I propose a second screening tool that makes assessment more naturalistic, repeatable, and scalable. This approach utilizes child and parent speech and other dyadic behavior, as well as physiological signals from both individuals, during an at-home parent-child play session to identify behavioral disorders.

1.3 Expected Contribution

Taken together, my work will make three important contributions toward the development of accurate and scalable screening technologies: (a) establish the feasibility of using both in-lab and in-the-wild behavioral and neurophysiological signals to detect emotion dysregulation in children at an earlier developmental stage compared to existing research, (b) demonstrate how multimodal approaches can improve screening accuracy, and (c) examine the perspectives of diverse stakeholders in pediatric mental health to derive insights that inform the development of future tools.

2 RELATED WORK

2.1 Monitoring Health and Wellbeing in Early Childhood

There has long been a focus on designing mobile and ubiquitous technologies for children and families, with a large body of prior work focusing on children’s health and development. Baby Steps was designed to encourage parents to frequently record their child’s developmental progress and share it with their pediatrician [28]. BebeCODE implemented a multi-informant approach to development tracking via a mobile application [35]. Smart toys have been developed to monitor vital signs [37], fine motor skills [30], and traumatic stress [40] in young children via tangible user interfaces. While these tools help caregivers monitor developmental progress and potentially identify concerns, there is limited work that focuses on *automatically screening* children for behavioral and emotional disorders. Boccanfuso et al. used play patterns and affective responses of children while interacting with an emotional robot to differentiate between children with autism spectrum disorder (ASD) and those developing typically [7]. Chong et al. use eyeglass-based cameras to capture children’s faces during adult-child interactions to screen for ASD via eye-contact detection [9]. Jiang et al. have attempted to screen children between 7 and 13 years of age for ADHD using a platform consisting of wearables, physical devices,

and large touchscreen interfaces [24]. However, most approaches require specialized equipment that hinders scalability, and screening tools for younger children are virtually non-existent. This results in missed opportunities for effectively intervening at an early age and improving long-term outcomes, which my work aims to fill by developing naturalistic assessments for preschool-aged children.

2.2 Using Behavioral and Neurophysiological Signals to Screen for Disorders

There is an opportunity for next-generation screening tools to support clinical diagnoses by not only detecting psychopathological disorders but also the abnormalities in the neurobiological mechanisms that drive them. Neuroimaging work over the past few decades has identified neural mechanisms underpinning the emotion regulation response and driving symptoms of psychopathology in preschool-aged [21]. Decreased neural activation in the lateral Pre-Frontal Cortex (LPFC) has been linked to poorer emotion regulation and higher aggressive behavior [10]. However, clinicians are unable to use these insights in their diagnostic process due to the difficulty of measuring neural activation levels in the wild. Prior research suggests that neural activation during emotion regulation could be indirectly measured via behavioral correlates such as facial expressions [11]. Particular expressions during emotion regulation have been shown to be correlated with neural activity in the lateral and medial prefrontal cortex and the amygdala [20, 32] – regions of the human brain known to be associated with emotion regulation [6, 34]. Eye gaze and bodily movement-related measures have also been identified as potential biomarkers for ADHD diagnosis [29], and gaze fixations have also been shown to predict neural activation during emotion regulation [36]. EarlyScreen, the video-based screening tool I describe in this work, leverages facial expressions and head and eye movements to predict both psychopathological disorder status as well as underlying LPFC activation levels to support clinical diagnosis.

Over the past few decades, behavioral assessments have also been increasingly incorporated into mental health assessment batteries in pediatric clinical settings, in addition to self- or parent-reported measures and interviews [2]. There is particular interest in naturalistic observations as alternative mediums of assessment, and structured parent-child play sessions, which are coded using the Dyadic Parent-Child Interaction Coding System (DPICS; [16]), have emerged as a commonly used assessment medium. Recent work has revealed associations between DPICS child codes and parent-reported behavioral concerns [13], motivating further research into diagnosis using both child as well as parent codes. While there is some research on using phone-based audio to categorize parent vocal behaviors into DPICS codes [22, 23], these works have significant gaps such as low concordance with manual coding and exclusive focus on parent speech. Moreover, while physiological signals of both participants during parent-child interaction have been shown to be associated with disruptive and attention-seeking behavior [17] as well as parent-child emotional attunement and ASD symptoms [4], prior mobile sensing work has largely ignored these co-occurring physiological changes. My work therefore proposes combining audio and physiological modalities to extract clinically relevant dyadic patterns and screen for disorders.

3 METHODOLOGY

3.1 EarlyScreen Study

94 participants aged 3.5 to 5 years old completed a clinically validated emotion regulation task in a laboratory setting (see [26] for more details). All procedures were approved by the Institutional Review Board of the University of Massachusetts Amherst. Participants played the Incredible Cake Kids [19] game where they were asked to select the “most delicious” cake for customers of a virtual bakery. The task consisted of 30 trials during which a virtual customer provided predetermined positive (e.g., happy) or negative (e.g., grumpy) feedback on the child’s choice of cake. The task consisted of three negative (four negative and one positive trial grouped together) and three positive blocks (four positive and one negative trial).

The participants’ caregivers completed four clinical questionnaires to determine the severity of the most commonly diagnosed early psychological disorders. These included the ADHD Inattention and ADHD Hyperactivity subscales from the *ADHD Rating Scale-5 Home Version* [14], the temper loss subscale from the *Multi-dimensional Assessment Profile for Disruptive Behavior* [38], and the externalizing symptom scale from the *Child Behavior Checklist* [1] for ages 1.5 to 5. To determine ground-truth labels, participants were categorized as *clinical* if they scored above the clinical threshold on at least one of the subscales, and *non-clinical* otherwise.

Three sources of behavioral and neurophysiological signals were recorded during the emotion regulation task. These included (i) facial expressions using video cameras, (ii) neural activation in the IPFC via functional Near-infrared Spectroscopy (fNIRS), and (iii) arousal of the autonomic nervous system via galvanic skin response (GSR). Since low IPFC activation during frustration is indicative of poor emotion regulation, EarlyScreen first attempts to use *only* video-based behavioral features to predict both disorder risk and low/normal IPFC activation. Then, features from all three modalities are combined to improve the accuracy of clinical risk prediction.

3.2 Predicting Psychopathology and Neural Activation Levels

EarlyScreen aims to utilize behavioral features extracted from videos to predict (i) neural activation levels in the PFC, as well as (ii) psychopathological disorder status among preschool children. The key challenge in doing so is the limited number of labels (one per individual) in our dataset, providing only coarse-grained subject-level information. At the same time, the frustration-inducing task comprises of up to 30 trials (15 positive and 15 negative feedback trials), which were recorded in the form of facial videos. Behavioral responses from each of these trials can be thought of as independent episodes of data containing fine-grained information.

To utilize both these coarse- and fine-grained patterns, my work proposes a novel Multi-scale Instance Fusion (MIF) framework that draws on the Multiple-Instance Learning (MIL) paradigm. It combines (i) an *MIL Pipeline* learning subject-level features from each instance of feedback using a bag representation module with (ii) a corresponding *Single Instance Learning (SIL) Pipeline* operating on coarse-grained features extracted by aggregating all feedback segments. Figure 2 shows our proposed MIF architecture, which is an

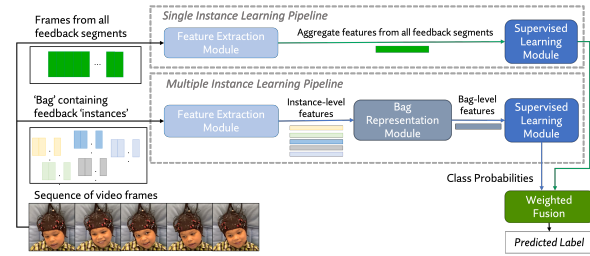


Figure 2: The proposed Multi-scale Instance Fusion (MIF) framework in EarlyScreen.

ensemble model consisting of these two components fused together in order to obtain the final predicted label. The bag representation module and the fusion weights are selected via hyperparameter tuning using nested cross-validation (see [26] for details).

3.3 Improving Predictions via Multimodal Supervised Contrastive Learning

As an extension of EarlyScreen, my follow-up work investigates whether fNIRS and GSR signals can be combined with video-based features to improve the accuracy of disorder screening. To this end, I proposed a multi-task supervised contrastive learning approach to learn multimodal embeddings that can be leveraged for downstream classification tasks [25]. The embedding model takes as input frame-level video features and the fNIRS and GSR time series from each trial in the EarlyScreen task. It attempts to minimize the supervised contrastive loss [27] between trials from similar individuals (clinical/non-clinical) as well as trials with the same feedback type (positive/negative). My work evaluates whether multimodal embeddings extracted in this manner outperform handcrafted features in classifying clinical individuals (for details, refer to [25]).

3.4 Stakeholder Perspectives on Pediatric Mental Health Screening Tools

In order to ensure the widespread adoption of screening tools, it is necessary to take into consideration the needs and perspectives of different groups of stakeholders (e.g., parents, pediatricians, child psychologists, preschool teachers, etc.) involved in supporting children’s mental health care. To this end, my work examines the perspectives of two key stakeholders – clinicians/mental health professionals and parents/primary caregivers – towards using mental health screening or diagnostic applications for young children. I conducted a survey of sixty mental health professionals and interviewed 25 parents of preschool-aged children to understand the major concerns and information gaps they currently encounter, their perception of the benefits and drawbacks of psychological screening technologies for children, and the functional requirements of screening technologies that are most important to them.

4 EVALUATION

4.1 Preliminary Findings

The MIF framework proposed in EarlyScreen achieves a significant improvement in classification performance both for predicting low vs. normal PFC activation levels as well as above vs. below

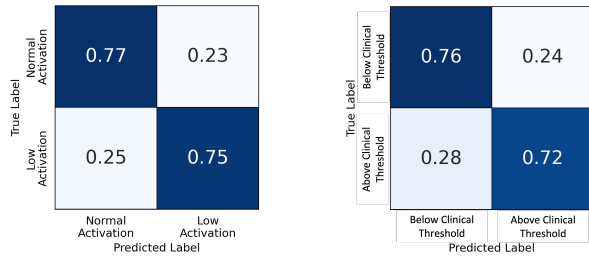


Figure 3: Normalized confusion matrices showing the prediction performance of neural activation and disorder risk classifiers in EarlyScreen.

clinical risk threshold as compared to “single instance” baselines. Figure 3 shows the normalized confusion matrices for both models. EarlyScreen is able to detect low PFC activation with an area under the receiver operating characteristic (ROC) curve of 0.85, and a sensitivity and specificity of 0.75 and 0.77 respectively. This work is the first to attempt to predict an objective measure of emotion regulation and produce proof-of-concept results with an accuracy comparable to current clinical diagnostic tools [41]. Similarly, EarlyScreen identifies children scoring above the clinical threshold on behavioral assessment scales with an area under the ROC curve (AUROC) of 0.80, a sensitivity of 0.72, and a specificity of 0.76. The model’s performance is comparable to recent approaches that predict clinical symptoms using behavioral data (e.g., [31]) using less than ten minutes of behavioral data.

Table 1 shows the AUROC and F1 scores for predicting clinical vs. non-clinical status using multimodal signals. It can be observed that combining handcrafted features leads to a drop in performance relative to models trained on a single modality, with GSR features outperforming others. However, multimodal embeddings extracted using a multitask supervised contrastive learning approach (proposed in [25]) result in improved classification performance, with an AUROC of 0.90 and an F1 score of 0.91.

In terms of stakeholder perspectives, over 75% of the clinicians surveyed felt the need to “improve the accuracy, efficiency, and convenience of how early childhood mental illness is diagnosed”. A majority of them felt that “biological and behavioral data could improve diagnosis”, and nearly 85% of respondents noted that “home-based diagnostic tools can make clinical services accessible to more families”. Most mental health practitioners appreciated EarlyScreen’s ability to collect ecologically-valid data. Similar sentiments were echoed by the parents interviewed about EarlyScreen. They noted the benefits of screening tools that “felt natural to children and not like an evaluation”. However, some parents voiced concerns about exposing children to screens and the impact of practice effects that might make longitudinal monitoring of symptoms with EarlyScreen less reliable. Clinicians also emphasized the need to control for parent involvement and other environmental factors when conducting assessments outside the clinic.

4.2 Future Work

Based on the encouraging classification performance of EarlyScreen’s clinical risk prediction models as well as insights from parents and

Table 1: Multimodal psychopathology prediction performance using handcrafted features vs. embeddings extracted via multitask supervised contrastive learning.

	Features	AUROC	F1 Score
Handcrafted Features	GSR	0.78	0.67
	Video	0.64	0.67
	fNIRS	0.74	0.67
	All	0.62	0.57
Multimodal Embeddings	All	0.90	0.91

clinicians, future work toward my dissertation will focus on developing a screening tool that (i) uses a screen-free, familiar, and naturalistic interaction as the assessment medium, (ii) can be deployed in the home and used for repeated assessments over time, and (iii) encourages and accounts for parent involvement.

The proposed tool will include a smartphone application that guides parents through a semi-structured play session (see Figure 1). This will be based on the SpecialTime activity used in parent-child interaction therapy [15] and will consist of three ten-minute phases: (i) child-led play, where the child chooses any activity and the parent is asked to play with them following their lead, (ii) parent-led play, where the parents choose the game and keep the child playing according to their rules, and (iii) clean-up, where parents direct the child to put away all toys on their own. The study app will record parent and child audio throughout the session. Simultaneous physiological and movement data will be collected via wearable devices on the parent and child’s wrists. The audio and wearable data will be used to develop novel machine learning algorithms that detect the dynamics of parent-child dyadic interactions and predict possible risks to the child’s behavioral health.

5 CONCLUSION

Mobile and wearable technology has the potential to enable accurate and highly scalable screening for mental health disorders in young children. In this work, I describe the development of video-based and multimodal approaches for predicting emotion dysregulation-related behaviors based on children’s responses during a clinically validated frustration-inducing game. Further, I present insights from user studies evaluating the utility of home-based screening tools for diverse stakeholders. Finally, I propose future work that incorporates feedback from prior studies in order to develop a screening tool with improved scalability and clinical utility. This work fulfills the urgent need for low-burden tools that provide instant and accurate information about young children’s mental health risk status, empowering caregivers and healthcare providers in their decision-making and significantly improving the developmental trajectories of high-risk children.

ACKNOWLEDGMENTS

This work was supported by the NIMH K23 MH11708 and the NIMH R21 MH126326 grants. The authors thank Dr. Adam Grabell and the SEED Lab at the University of Massachusetts Amherst for their invaluable contributions to this work.

REFERENCES

- [1] Thomas M Achenbach and Leslie A Rescorla. 2000. *Manual for the ASEBA preschool forms and profiles*. Vol. 30. Burlington, VT: University of Vermont, Research center for children, youth, & families.
- [2] Helen Aspland and Frances Gardner. 2003. Observational measures of parent-child interaction: an introductory review. *Child and Adolescent Mental Health* (2003).
- [3] Shelli Avenevoli, Joseph C Blader, and Ellen Leibenluft. 2015. Irritability in Youth: An Update. *Journal of the American Academy of Child and Adolescent Psychiatry* 54, 11 (2015), 881–883.
- [4] Jason K Baker, Rachel M Fenning, Mariann A Howland, Brian R Baucom, Jacquelyn Moffitt, and Stephen A Erath. 2015. Brief report: A pilot study of parent-child biobehavioral synchrony in autism spectrum disorder. *Journal of autism and developmental disorders* 45 (2015), 4140–4146.
- [5] Russell A Barkley. 1997. *ADHD and the nature of self-control*. Guilford Press.
- [6] Robert JR Blair. 2016. The neurobiology of impulsive aggression. *Journal of child and adolescent psychopharmacology* 26, 1 (2016), 4–9.
- [7] Laura Boccanfuso, Erin Barney, Claire Foster, Yeojin Amy Ahn, Katarzyna Chawarska, Brian Scassellati, and Frederick Shic. 2016. Emotional robot to examine different play patterns and affective responses of children with and without ASD. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 19–26.
- [8] Elizabeth H Bringewatt and Elizabeth T Gershoff. 2010. Falling through the cracks: Gaps and barriers in the mental health system for America's disadvantaged children. *Children and Youth Services Review* 32, 10 (2010), 1291–1299.
- [9] Eunji Chong, Katha Chanda, Zhefan Ye, Audrey Southerland, Nataniel Ruiz, Rebecca M Jones, Agata Rozga, and James M Rehg. 2017. Detecting gaze towards eyes in natural social interactions and its use in child assessment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–20.
- [10] Emil F Coccaro, Chandra Sekhar Sripada, Rachel N Yanowitch, and K Luan Phan. 2011. Corticolimbic function in impulsive aggressive behavior. *Biological psychiatry* 69, 12 (2011), 1153–1159.
- [11] Pamela M Cole. 1986. Children's spontaneous control of facial expression. *Child development* (1986), 1309–1321.
- [12] Pamela Y Collins, Vikram Patel, Sarah S Joestl, Dana March, Thomas R Insel, Abdallah S Daar, Isabel A Bordin, E Jane Costello, Maureen Durkin, Christopher Fairburn, et al. 2011. Grand challenges in global mental health. *Nature* 475, 7354 (2011), 27–30.
- [13] Allison M Cotter and Elizabeth Brestan-Knight. 2020. Convergence of parent report and child behavior using the Dyadic Parent-Child Interaction Coding System (DPICS). *Journal of Child and Family Studies* 29 (2020), 3287–3301.
- [14] George J DuPaul, Robert Reid, Arthur D Anastopoulos, Matthew C Lambert, Marley W Watkins, and Thomas J Power. 2016. Parent and teacher ratings of attention-deficit/hyperactivity disorder symptoms: Factor structure and normative data. *Psychological Assessment* 28, 2 (2016), 214.
- [15] SM Eyberg and B Funderburk. 2011. Parent-child interaction therapy protocol. Gainesville, FL: PCIT International (2011).
- [16] Sheila M Eyberg. 2013. *Dyadic parent-child interaction coding system (DPICS): Comprehensive manual for research and training*. PCIT International, Incorporated.
- [17] Nia Fogelman, Julie Schwartz, Tara M Chaplin, Ania M Jastreboff, Wendy K Silverman, and Rajita Sinha. 2022. Parent stress and trauma, autonomic responses, and negative child behaviors. *Child Psychiatry & Human Development* (2022), 1–10.
- [18] Centers for Disease Control and Prevention. 2023. *Improving Access to Children's Mental Health Care*. Retrieved May 6, 2023 from <https://www.cdc.gov/childrensmentalhealth/access.html>
- [19] Adam S Grabell, Theodore J Huppert, Frank A Fishburn, Yanwei Li, Christina O Hlutowksy, Hannah M Jones, Lauren S Wakschlag, and Susan B Perlman. 2019. Neural correlates of early deliberate emotion regulation: young children's responses to interpersonal scaffolding. *Developmental cognitive neuroscience* 40 (2019), 100708.
- [20] Adam S Grabell, Theodore J Huppert, Frank A Fishburn, Yanwei Li, Hannah M Jones, Aimee E Wilett, Lisa M Bemis, and Susan B Perlman. 2018. Using facial muscular movements to understand young children's emotion regulation and concurrent neural activation. *Developmental science* 21, 5 (2018), e12628.
- [21] Adam S Grabell, Yanwei Li, Jeff W Barker, Lauren S Wakschlag, Theodore J Huppert, and Susan B Perlman. 2018. Evidence of non-linear associations between frustration-related prefrontal cortex activation and the normal: abnormal spectrum of irritability in young children. *Journal of abnormal child psychology* 46, 1 (2018), 137–147.
- [22] Bernd Huber, Richard F Davis III, Allison Cotter, Emily Junkin, Mindy Yard, Stuart Shieber, Elizabeth Brestan-Knight, and Krzysztof Z Gajos. 2019. Special-Time: Automatically detecting dialogue acts from speech to support parent-child interaction therapy. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 139–148.
- [23] Inseok Hwang, Chungkuk Yoo, Chanyou Hwang, Dongsun Yim, Youngki Lee, Chulhong Min, John Kim, and June-hwa Song. 2014. TalkBetter: family-driven mobile intervention care for children with language delay. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 1283–1296.
- [24] Xinlong Jiang, Yiqiang Chen, Wuliang Huang, Teng Zhang, Chenlong Gao, Yunbing Xing, and Yi Zheng. 2020. WeDA: Designing and evaluating a scale-driven wearable diagnostic assessment system for children with ADHD. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–12.
- [25] Manasa Kalanadhabhatta, Adrelys Mateo Santana, Deepak Ganesan, Tauhidur Rahman, and Adam Grabell. 2022. Extracting Multimodal Embeddings via Supervised Contrastive Learning for Psychological Screening. In *2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 1–8.
- [26] Manasa Kalanadhabhatta, Adrelys Mateo Santana, Zhongyang Zhang, Deepak Ganesan, Adam S Grabell, and Tauhidur Rahman. 2022. Earlyscreen: Multi-scale instance fusion for predicting neural activation and psychopathology in preschool children. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022).
- [27] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. 2020. Supervised contrastive learning. *Advances in Neural Information Processing Systems* 33 (2020), 18661–18673.
- [28] Julie A Kientz, Rosa I Arriaga, and Gregory D Abowd. 2009. Baby steps: evaluation of a system to support record-keeping for parents of young children. In *Proceedings of the sigchi conference on human factors in computing systems*. 1713–1722.
- [29] Astar Lev, Yoram Braw, Tomer Elbaum, Michael Wagner, and Yuri Rassovsky. 2020. Eye Tracking During a Continuous Performance Test: Utility for Assessing ADHD Patients. *Journal of Attention Disorders* (2020), 1087054720972786.
- [30] Svetlana Mironcika, Antoine de Schipper, Annette Brons, Huub Toussaint, Ben Kröse, and Ben Schouten. 2018. Smart toys design opportunities for measuring children's fine motor skills development. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*. 349–356.
- [31] Philipp Mock, Maike Tibus, Ann-Christine Ehls, Harald Baayen, and Peter Gerjets. 2018. Predicting ADHD risk from touch interaction data. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*. 446–454.
- [32] Susan B Perlman, Beatriz Luna, Tyler C Hein, and Theodore J Huppert. 2014. fNIRS evidence of prefrontal regulation of frustration in early childhood. *Neuroimage* 85 (2014), 326–334.
- [33] Joseph S Raiker, Andrew J Freeman, Guillermo Perez-Algorta, Thomas W Frazier, Robert L Findling, and Eric A Youngstrom. 2017. Accuracy of Achenbach scales in the screening of attention-deficit/hyperactivity disorder in a community mental health clinic. *Journal of the American Academy of Child & Adolescent Psychiatry* 56, 5 (2017), 401–409.
- [34] Philip Shaw, Argyris Stringaris, Joel Nigg, and Ellen Leibenluft. 2014. Emotion dysregulation in attention deficit hyperactivity disorder. *American Journal of Psychiatry* 171, 3 (2014), 276–293.
- [35] Seokwoo Song, Juho Kim, Bumsoo Kang, Wonjeong Park, and John Kim. 2018. BebeCode: Collaborative child development tracking system. In *Proceedings of the 2018 chi conference on human factors in computing systems*. 1–12.
- [36] Carien M van Reekum, Tom Johnstone, Heather L Urry, Marchell E Thuro, Hillary S Schaefer, Andrew L Alexander, and Richard J Davidson. 2007. Gaze fixations predict brain activation during the voluntary regulation of picture-induced negative affect. *Neuroimage* 36, 3 (2007), 1041–1055.
- [37] Emanuel Vonach, Marianne Ternek, Georg Gerstweiler, and Hannes Kaufmann. 2016. Design of a health monitoring toy for children. In *Proceedings of the The 15th International Conference on Interaction Design and Children*. 58–67.
- [38] Lauren S Wakschlag, Seung W Choi, Alice S Carter, Heide Hullsiek, James Burns, Kimberly McCarthy, Ellen Leibenluft, and Margaret J Briggs-Gowan. 2012. Defining the developmental parameters of temper loss in early childhood: implications for developmental psychopathology. *Journal of Child Psychology and Psychiatry* 53, 11 (2012), 1099–1108.
- [39] Lauren S Wakschlag, Patrick H Tolan, and Bennett L Leventhal. 2010. Research Review: 'Ain't misbehavin': Towards a developmentally-specified nosology for preschool disruptive behavior. *Journal of Child Psychology and Psychiatry* 51, 1 (2010), 3–22.
- [40] Xiyue Wang, Kazuki Takashima, Tomoaki Adachi, Patrick Finn, Ehud Sharlin, and Yoshifumi Kitamura. 2020. AssessBlocks: Exploring Toy Block Play Features for Assessing Stress in Young Children after Natural Disasters. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–29.
- [41] Erin M Warnick, Michael B Bracken, and Stanislav Kasl. 2008. Screening efficiency of the Child Behavior Checklist and Strengths and Difficulties Questionnaire: A systematic review. *Child and Adolescent Mental Health* 13, 3 (2008), 140–147.
- [42] Daniel G Whitney and Mark D Peterson. 2019. US national and state-level prevalence of mental health disorders and disparities of mental health care use in children. *JAMA pediatrics* 173, 4 (2019), 389–391.