RESEARCH ARTICLE

Rapid detection of internalizing diagnosis in young children enabled by wearable sensors and machine learning

Ryan S. McGinnis[1], Ellen W. McGinnis[2,3]*, Jessica Hruschak[4], Nestor L. Lopez-Duran[3], Kate Fitzgerald[4], Katherine L. Rosenblum[4], Maria Muzik[4]

1 Department of Electrical and Biomedical Engineering, University of Vermont, Burlington, VT, United States of America, **2** Department of Psychiatry, University of Vermont, Burlington, VT, United States of America, **3** Department of Psychology, University of Michigan, Ann Arbor, MI, United States of America, **4** Department of Psychiatry, University of Michigan, Ann Arbor, MI, United States of America

☯ These authors contributed equally to this work.
* ryan.mcginnis@uvm.edu

Abstract

There is a critical need for fast, inexpensive, objective, and accurate screening tools for childhood psychopathology. Perhaps most compelling is in the case of internalizing disorders, like anxiety and depression, where unobservable symptoms cause children to go unassessed–suffering in silence because they never exhibiting the disruptive behaviors that would lead to a referral for diagnostic assessment. If left untreated these disorders are associated with long-term negative outcomes including substance abuse and increased risk for suicide. This paper presents a new approach for identifying children with internalizing disorders using an instrumented 90-second mood induction task. Participant motion during the task is monitored using a commercially available wearable sensor. We show that machine learning can be used to differentiate children with an internalizing diagnosis from controls with 81% accuracy (67% sensitivity, 88% specificity). We provide a detailed description of the modeling methodology used to arrive at these results and explore further the predictive ability of each temporal phase of the mood induction task. Kinematical measures most discriminative of internalizing diagnosis are analyzed in detail, showing affected children exhibit significantly more avoidance of ambiguous threat. Performance of the proposed approach is compared to clinical thresholds on parent-reported child symptoms which differentiate children with an internalizing diagnosis from controls with slightly lower accuracy (.68-.75 vs. .81), slightly higher specificity (.88–1.00 vs. .88), and lower sensitivity (.00-.42 vs. .67) than the proposed, instrumented method. These results point toward the future use of this approach for screening children for internalizing disorders so that interventions can be deployed when they have the highest chance for long-term success.

Introduction

Nearly 1 out of every 5 children experience an internalizing disorder (19.6% anxiety, 2.1% depression) during childhood [1,2]. Anxiety and depression (collectively internalizing

disorders) are chronic conditions that start as early as the preschool years [3,4] and impair a child's relationships, development, and functioning [5–9]. If left untreated, childhood internalizing disorders predict later health problems including substance abuse [10,11], development of comorbid psychopathology [12–14], increased risk for suicide [15], and substantial functional impairment [16,17]. These negative long-term outcomes reveal the high individual and societal burden of internalizing disorders [18] and make clear the need for effective early interventions.

Thanks to greater neuroplasticity, interventions can be very effective in this population if disorders are identified early in development [19]. However, the current healthcare referral process usually involves parents reporting problem behaviors to their pediatrician and, if functionally impairing, the child is then referred to a child psychologist or psychiatrist for a diagnostic assessment. Children with internalizing disorders, where symptoms are inherently inward facing, are less likely than those with externalizing disorders to be identified by parents or teachers as needing professional assessment ([20]; for review see [21]), thus preventing or delaying their access to early intervention. Children under 6 have the highest rate of unmet needs [22]. For example, as little as 3% of 4 year-olds with a clinical diagnosis receive the necessary professional mental health intervention [23]. This points to the need for standardized screening tools for internalizing disorders in young children.

Even if referred, current diagnostic assessments have been shown to capture only the most severely impaired preschoolers, but miss a large number of children who may go on to develop additional clinical impairments [24,25]. Providers try to improve these assessments by considering multi-informant reports from children, parents, and teachers, but these also have limitations. For example, children under the age of eight are unreliable self-reporters [26–28], and parental report of child problems are often inaccurate [29–31] as the unobservable symptoms characteristic of internalizing disorders (e.g., thoughts and emotions), are difficult to identify and thus go underreported [31]. Parents who have an internalizing disorder themselves are known to over-report unobservable symptoms [32], increasing complexity of this problem. Thus, there is a clear and unmet need for objective markers of internalizing disorders that can be incorporated into new screening tools for all children.

Observational methods for assessing psychopathology 'press' for specific behaviors and affect [33] and have high research and clinical utility [34]. One example, known as a mood induction task, engages a child in a short laboratory-based activity meant to induce expected negative or positive emotions. To provide objective markers of psychopathology, researchers often utilize a behavioral coding technique on video recordings of the task, where at least two researchers watch the video recordings and assign scores based on child verbalizations or facial and body movements (e.g., see [35]). Behavioral coding has been shown to identify valid risk markers for childhood psychopathology using a variety of mood induction activities [36,37]. However, it has significant drawbacks that limit clinical utility, including the need for extensive training in a standardized coding manual and the hours required to watch and score video recordings of the task, while also consensus scoring a percentage of participants (often one out of five) to ensure reliability [38]. While it is clear that observational methods can provide objective markers of psychopathology, the complexity and resources required for behavioral coding prevent its use as a screening tool for childhood internalizing disorders.

New advances in wearable sensors present the opportunity to track child movement without the need for extensive training or time to watch and score task videos. Our previous work has described the use of a wearable inertial measurement unit (IMU), composed of a three-axis accelerometer and three-axis angular rate gyroscope, for tracking child motion during a standardized fear induction task [39–42]. Kinematical measures extracted from these data were associated with other known measures of risk for internalizing disorders and confirmed

expected temporal characteristics of the task in a small (N = 18) sample of children [40]. In a larger sample (N = 62), IMU data, but not behavioral codes, were associated with parent-reported child symptoms and exhibited statistically significant differences between children with and without an internalizing diagnosis [41]. This *instrumented* mood induction task provides an objective measure of child motion without the limitations of behavioral coding and, when taken with these preliminary results, has the potential to be used as a screening tool for childhood internalizing disorders.

Advancing the use of this instrumented mood induction task as a tool for identifying children with an internalizing disorder requires establishing a model of the complex relationship between kinematical measures extracted from wearable sensor data and diagnosis. A data driven approach, like machine learning, is ideally suited for this task, and has been leveraged for this use in a variety of conditions including, for example, Multiple Sclerosis [43–45], Parkinson's Disease [46,47], and Atrial Fibrillation [48]. In this case, the wearable sensor time series captured from each child during the mood induction task provides a complete, albeit high dimensional (i.e., an IMU sampling at 100 Hz for a 20-second task yields 12,000 data points), picture of their motion. However, by computing a smaller number of *features* (e.g., mean, kurtosis) that each explain a different pattern inherent to the data, the dimensionality of the high-dimensional time series can be reduced. This process of defining the set of features that is able to capture important aspects of the raw data is known as *feature engineering*. The process of *machine learning* is in then training a statistical model to recognize the relationship between these objective measures of a child's motion during the task and their diagnosis. This approach allows for the realization of much more complex relationships then would be possible from theory-based modeling alone. These efforts form one facet of the burgeoning field of digital medicine [49,50], but notably the use of these techniques for improving childhood mental health is just beginning, with efforts focusing primarily on improving access to care through mobile delivery methods [51]. Thus, the use of machine learning and wearable sensors for advancing the state of childhood mental health screening represents a novel contribution to the field of digital medicine.

To further investigate the potential of this instrumented mood induction task as a screening tool for childhood psychopathology, we explore the use of machine learning to develop statistical models for identifying children who have an internalizing disorder. Specifically, this paper builds upon two of our recent conference papers [39,42] by presenting a detailed description of the modeling methodology and performance, an analysis of kinematical measures most discriminative of internalizing diagnostic status, and a comparison to the performance of models trained on parent-reported child symptoms.

Method

Participants

Studies had approval from the University of Michigan Institutional Review Board (HUM00091788; HUM00033838). Participants included 63 children (57% female) and their primary caregivers (95.2% mothers). Participants were recruited from either an ongoing observational study (Bonding Between Mothers and Children, PI: Maria Muzik; *n* = 14) or from flyers posted in the community (*n* = 14) and psychiatry clinics (*n* = 35) to obtain a sample with a wide range of symptom presentations. Eligible participants were children between the ages of 3 and 8 who spoke fluent English and whose caregivers were 18 years and older. Exclusion criteria were suspected or diagnosed developmental disorder (e.g., autism), having a serious medical condition, or taking medications that affect the central nervous system. The resulting sample of children were aged between 3 and 7 years (M = 5.25 SD = 1.10), was 65% White

non-Latinx, and 82.5% lived in two-parent households. Twenty participants (32%) had an annual household income of greater than $100,000.

Multimodal assessments, including diagnostic interviews, were conducted for 62 of the children between August 2014 and August 2015. Based on these multimodal assessments and consensus coding, 21 participants were identified as having an internalizing diagnosis (current (n = 17), past (n = 4)) according to DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, 4th. Edition). Diagnostic details are provided in Table 1.

Procedure

Child and caregiver were brought into the university-based laboratory and provided written consent to complete a battery of tasks. Caregivers completed self- and parent-report questionnaires and a diagnostic interview to assess for child psychiatric diagnoses while children underwent a series of behavioral tasks in an adjacent room. Behavioral tasks were designed to elicit fear responses and positive affect. Participants were compensated for their time.

Herein, we consider a subset of data from the larger study by examining participant response to a single behavioral task designed to elicit fear (the 'Snake Task'), as well as the diagnostic interview and questionnaires used to assess internalizing symptoms and diagnoses. The Snake Task has been shown to induce anxiety and fear in young children [52,53]. This task is standardized and all research assistants were trained to carry out the task according to protocol. The total task duration was approximately 90 seconds, and task behaviors were conceptually segmented into three temporal phases [54]: 1) Potential Threat: The child was led into a novel, dimly lit room, unsure of what was inside while the administrator gave scripted statements to build anticipation such as "I have something in here to show you" and "Let's be quiet so it doesn't wake up". The administrator led the child slowly toward the back of the room where a terrarium was covered with a blanket, gesturing for them to follow until they paused within 1 foot of the terrarium; 2) Startle: The child was startled by the administrator rapidly uncovering the terrarium and bringing the fake snake from inside to the child's eye level several inches from their face; 3) Response Modulation: The child was encouraged to touch the snake if they wanted, to ensure it was fake, and was reassured verbally (e.g. "It's just a silly toy snake") as needed, remaining with the snake until the administrator gestured them to leave the room and end the task. Following the task, children transitioned to free play with the task administrator to regulate and debrief about their experience.

Questionnaires. The Child Behavior Checklist (CBCL) is a parent-completed questionnaire designed to assess child problem behaviors [55]. The scale consists of 120 items related to

Table 1. Diagnostic characteristics of the sample.

Primary Diagnoses	n = 22	Secondary Diagnoses	n = 10
Post-traumatic Stress Disorder	5	Specific Phobia	2
Anxiety/Depression Not Otherwise Specified	5	Social Anxiety Disorder	2
Adjustment Disorder	4	Depression	1
Separation Anxiety Disorder	4	Anxiety Not Otherwise Specified	1
Specific Phobia	3	Attention Deficit Hyperactive Disorder	2
Depression	1	Oppositional Defiant Disorder	2
Attention Deficit Hyperactive Disorder	1		

Number of subjects with specific primary and secondary diagnoses. Note that this includes both internalizing and externalizing diagnoses.

behavior problems across multiple domains. Items are scored on a three-point scale ranging from "not true" to "often true" of the child. Responses result in global T scores for externalizing, internalizing, and total problems, as well as a number of empirically based syndrome scales and disorder-based scales. Only scales available in both versions (ages 1.5–5 and 6–18) were used in subsequent analyses. The CBCL has well established validity and reliability (see [56]).

Subject demographic information was collected using a questionnaire that includes questions regarding child race, gender and family income.

Clinical interview. Trained clinical psychology doctoral students, or postdoctoral fellows, conducted a single structured clinical interview with each child's caregiver. The current study used a version of The Schedule for Affective Disorders and Schizophrenia for School-Age Children Present and Lifetime Version (K-SADS-PL) modified for use with preschool-aged children [57]. In this diagnostic interview, the clinician spent up to two hours with the caregiver assessing symptoms of past and current child psychiatric disorders. Interviewers received monthly (or more frequent) supervision by a licensed psychologist and psychiatrist, wherein all cases were reviewed by all clinicians and the supervisor. Final diagnoses were derived via clinical consensus using the best-estimate procedures [58] to integrate a holistic picture based on child and parent report, family history, and other self-report symptom checklists. It is worth noting that team-based assessment and consensus diagnosis are most often only conducted in research contexts, the resulting diagnoses can be considered a true gold-standard, and this practice is not representative of the diagnostic procedures used in the majority of clinical contexts.

Wearable sensor signal processing and feature extraction. During the behavioral battery, child motion was tracked using a belt-worn IMU (3-Space Sensor, YEI Technology, Portsmouth, OH, USA) secured around the waist at approximately the location of the body center of mass. Acceleration and angular velocity data were sampled by the device at approximately 300 Hz, down-sampled to 100 Hz, and low-pass filtered using a fourth-order Butterworth IIR filter with a cutoff frequency of 20 Hz in software prior to use. These data were fused to determine device orientation as a function of time using the complementary filtering approach described in [40,59]. Device orientation was used to resolve raw IMU measurements of acceleration and angular velocity in a world-fixed reference frame that has one axis directed vertically upwards. These data were further decomposed into vertical acceleration and angular velocity (av and ωv, respectively), and the vector magnitude of horizontal acceleration and angular velocity (ah and ωh, respectively). Orientation estimates were also used to compute tilt (α) and yaw (γ) angles of the participant as a function of time yielding six time series (ah, av, ωh, ωv, α, γ) for further analysis. The interested reader may refer to [40], for a detailed description of this approach.

Time series were segmented into the three conceptual phases [54]: Potential Threat (20 seconds, from 23 to 3 seconds prior to the moment of startle), Startle (6 seconds, from 3 seconds prior to 3 seconds post the moment of startle), and Response Modulation (20 seconds, from 3 seconds to 23 seconds post the moment of startle) and signal features were extracted from each. Signal features included mean, root mean square (RMS), skew, kurtosis, range, maximum, minimum, standard deviation, peak to RMS amplitude, signal power within specific frequency bands (i.e., 0–0.5 Hz, 0.5–1.5 Hz, 1.5–5 Hz, 5–10 Hz, 10–15 Hz, 15–20 Hz, all frequencies greater than 20 Hz), and the location and height of peaks in the power spectrum and autocorrelation of the signal. This yielded a total of 29 features from each of the six time series, or 174 total features, from each phase of the task. Signal processing and feature extraction were performed in MATLAB (Mathworks, Natick, MA, USA), using source code available from [60].

Statistical models for identifying internalizing diagnosis. A supervised learning approach was used to create binary classification models that relate features from the IMU-derived signals to internalizing diagnosis derived from the K-SADS-PL and clinical consensus. Models were created using features from each temporal phase of the task. Performance of the classifiers was established using leave-one-subject-out (LOSO) cross validation. In this approach, features from 61 of the 62 subjects were partitioned into a training dataset and converted to z-scores prior to performing Davies-Bouldin Index [61] based feature selection to yield the 10 features with zero mean and unit variance that best discriminate between diagnostic groups. These features were used to train a logistic regression for predicting internalizing diagnosis. The same 10 features were extracted, converted to z-scores based on parameters (e.g. mean, variance) from the training set, and used as input to the model for predicting the diagnosis of the one remaining test subject. This process was repeated until the diagnosis of each subject had been predicted. Logistic regression was chosen herein to protect against over-fitting given the relatively small (N = 62) sample, and because it requires minimal computational overhead for prediction enabling future deployment on resource-constrained devices.

We also examined the utility of the CBCL as a screening tool for internalizing diagnosis (according to the K-SADS-PL with clinical consensus) in this sample using previously-established clinical cutoffs (T score $\geq$ 70) for manualized use [55] and a more conservative cutoff (T score $\geq$ 55) suggested for improving screening efficiency [62].

Model performance was assessed in several ways. First, we examined classification performance by reporting accuracy, sensitivity and specificity with a score threshold for the logistic regression of 0.5. These metrics were computed following standard definitions [63]. Next, receiver operating characteristic (ROC) curves–which plot true positive rate against false positive rate for varying thresholds on the scores–were constructed for each classifier. Area under the ROC curve (AUC) was used to comment on the general discriminative ability of the classifiers [63]. Finally, a permutation test was conducted to examine these results in the context of results obtained by chance from this dataset [64]. To complete this test, we first approximated the distribution of possible error rates (error rate = number of incorrect predictions / total number of predictions = 1 –classification accuracy) for each classifier as a beta distribution parameterized by the number of incorrect predictions and the total number of observations, as indicated in [64,65], and randomly sampled 100 possible error rates from this distribution. Next, we repeated the model training process outlined previously for 100 random permutations of the diagnostic labels, computing the classification error rate for each. Finally, a paired-sample Mann-Whitney U-test was used to identify the temporal phases that yield classification models with error rates significantly different from those expected by chance from this dataset.

For models that reported significantly better error rates than those expected by chance, we further examined the features used as input to these models.

Results

Performance of the logistic regression models trained for detecting children with internalizing diagnoses based on wearable sensor data from each temporal phase of the snake task are reported in Table 2. Metrics include accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC). The model developed from data sample during the Potential Threat phase outperforms models from the Startle and Response Modulation phases (Accuracy: .81 vs. .58 and .52; Sensitivity: .67 vs. .33 and .29; Specificity: .88 vs. .71 and .63; AUC: .85 vs. .59 and .48). This performance difference is further revealed in the ROC curves of Fig 1, where models trained on data from the Potential Threat, Startle, and Response Modulation phases are shown in blue, red, and yellow, respectively. As indicated in the ROC

Table 2. Performance characteristics of models developed for detecting children with internalizing diagnoses from wearable sensor data during each phase of the mood induction task.

	Accuracy	Sensitivity	Specificity	AUC
Potential Threat	.81	.67	.88	.85
Startle	.58	.33	.71	.59
Response Modulation	.52	.29	.63	.48

Accuracy, sensitivity, specificity and area under the receiver operating characteristic curve (AUC) for models trained using data from each temporal phase of the mood induction task (Potential Threat, Startle, Response Modulation).

curves, changing the score threshold for the logistic regression can alter its specific performance metrics. In this case, changing the threshold from .5 to .375 for the Potential Threat model maintains the overall accuracy (.81), decreases specificity (.88 to .81), and importantly

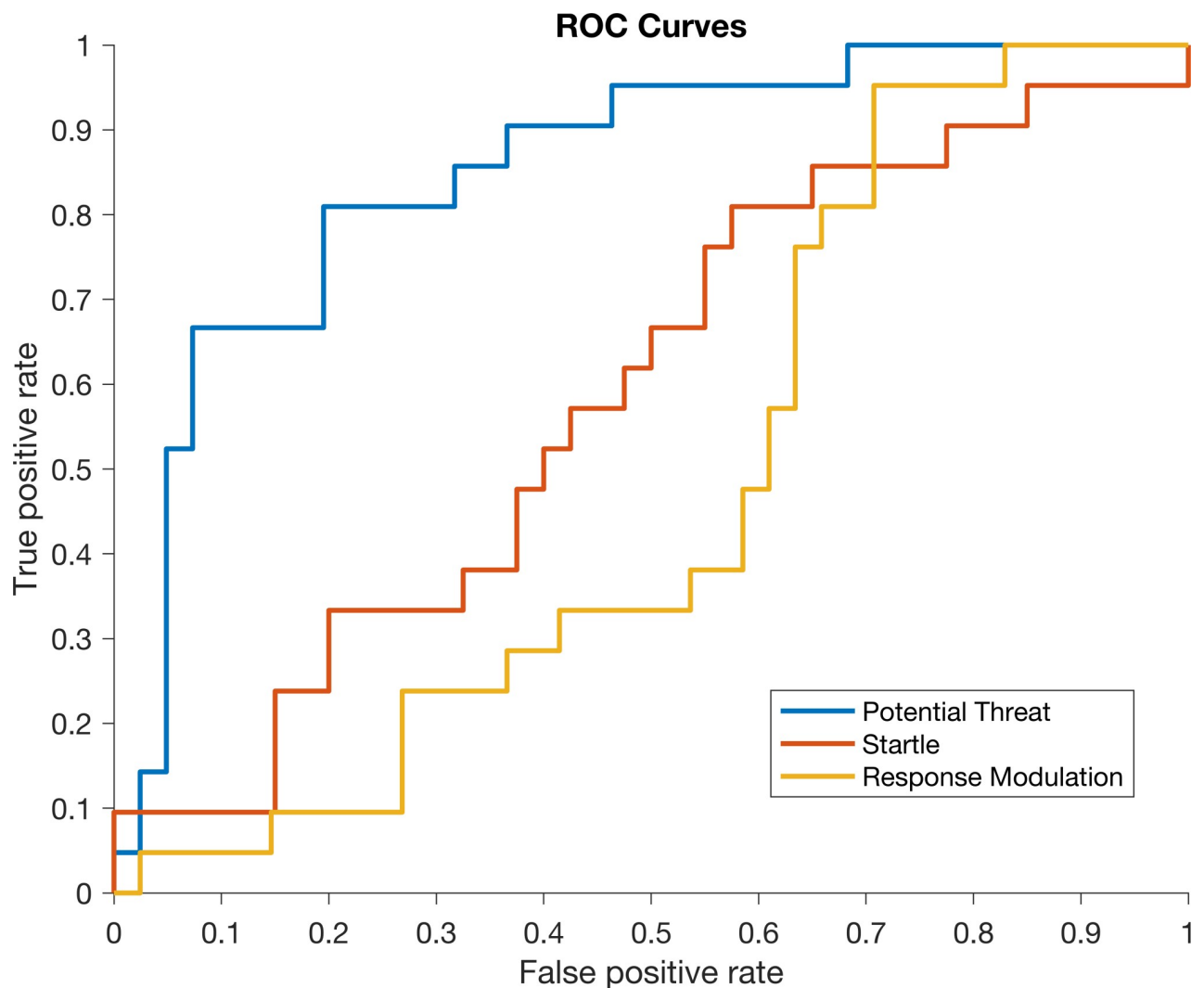**Fig 1. Receiver operating characteristic (ROC) curves for models trained to detect children with internalizing diagnoses.** Curves for logistic regressions trained on data from the potential threat, startle, and response modulation phases of the snake task are indicated in blue, red, and yellow, respectively. The model trained on data from the potential threat phase performs better than the other models.
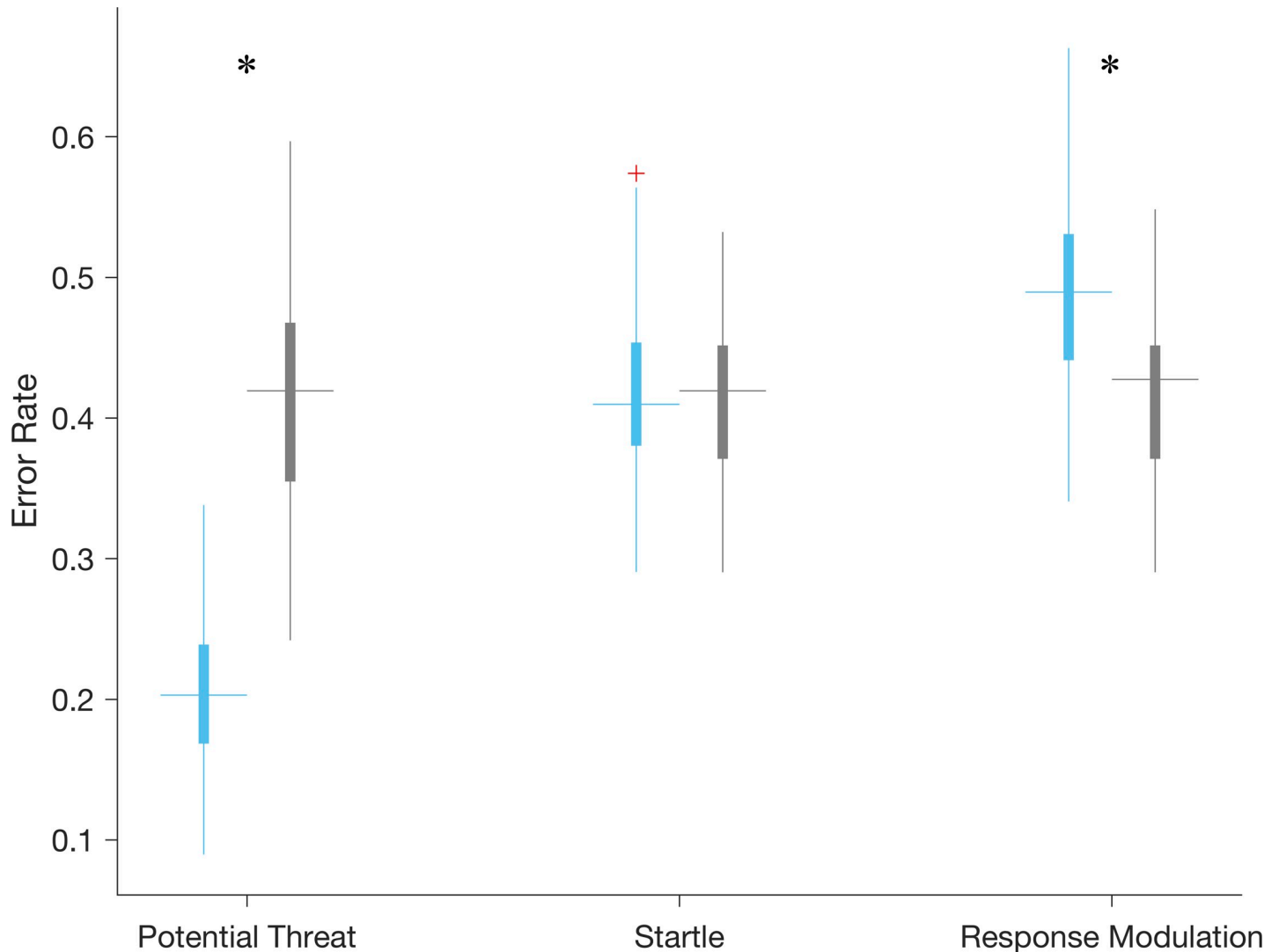
Fig 2. Boxplots of error rates for models trained to detect children with internalizing diagnoses compared to those due to chance for each temporal phase of the snake task. Error rates due to random chance determined via permutation test are shown in gray, while those from the actual data are in teal. Statistically significant differences are noted with an asterisk. The model trained on data from the potential threat phase is the only one to outperform random chance.

increases sensitivity (.67 to .81). This could be important when considering use of this method for screening children for internalizing disorders.

Results of the permutation test used for determining how the error rate of the classifiers trained on data from each phase compares to error rates expected by random chance from the dataset in are reported in the boxplot of Fig 2. Specifically, the distribution of error rates for the logistic regressions from each phase (teal) and the corresponding error rates achieved by chance (gray) are reported. Statistically significant differences between the median error rates achieved by the classifiers and by chance were observed in the Potential Threat (significantly lower error rate than expected by chance, p < .01) and Response Modulation (significantly higher error rate than expected by chance, p < .01) phases. Significant differences are indicated by asterisks in Fig 2.

The results of Fig 2 indicate that the classifier developed from data sampled during the Potential Threat phase is the only model that provides a statistically significant improvement
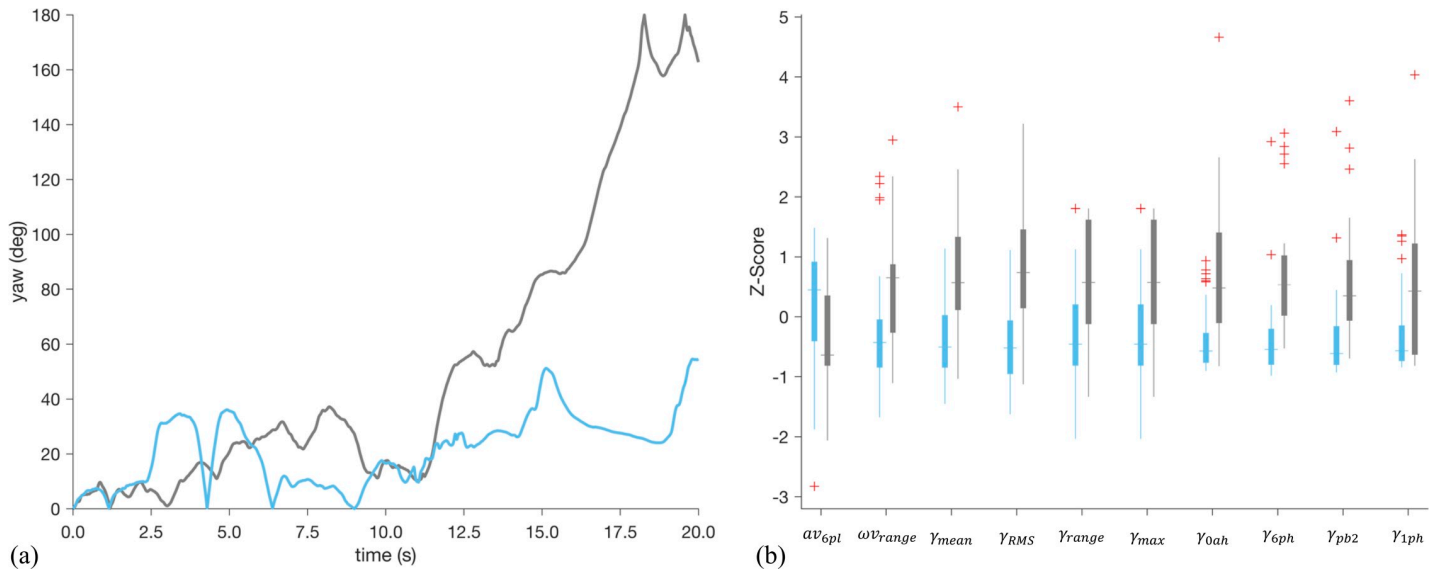
Fig 3. Yaw angle time series from selected subjects (a) and boxplots of selected features from all subjects (b). Time series data from a subject with an internalizing diagnosis is shown in gray, while that from a subject without is shown in teal. Similarly, gray boxplots correspond to data from subjects with a diagnosis while teal boxplots are from those without. The significant deviation in the yaw angle between subjects noted in (a) is reflected across all subjects in the boxplots of (b).

in classification error rate over random chance. To this end, we further examine the 10 features used as input to this model. Across the 62 iterations of the leave-one-subject-out cross validation, the following ten features were selected a minimum of 58 times: location of the 6th peak in the power spectrum of av (av_{6pl}), range of ωv (ωv_{range}), mean of γ (γ_{mean}), RMS of γ (γ_{RMS}), range of γ (γ_{range}), maximum of γ (γ_{max}), the height of the peak at zero lag in the autocorrelation of γ (γ_{0ah}), the heights of the 1st and 6th peaks in the power spectrum of γ (γ_{1ph}) and γ_{6ph}, respectively), and the signal power between 0.5–1.5 Hz in γ (γ_{pb2}). Nine of the ten features in this list are directly related to the yaw angle (γ) of the subject (ωv is essentially yaw angular velocity), and we therefore report the γ time series from representative subjects with (gray) and without (teal) an internalizing diagnosis in Fig 3A. There is a significant divergence in the yaw angles achieved by each subject beginning roughly half-way through the phase that leads the subject with a diagnosis to end the phase facing in the opposite direction of where they started ($\gamma \cong 180°$), which contrasts the subject without a diagnosis who ends the phase facing in roughly the same direction as when they started ($\gamma < 60°$). The differences noted in these representative subjects are consistent across the sample as evidenced by boxplots of the ten features used as input to the classifier presented in Fig 3B. Subjects with a diagnosis (gray) have higher values of ωv_{range}, γ_{mean}, γ_{RMS}, γ_{range}, and γ_{max}, all of which confirm the divergence noted in Fig 3A is consistent across the sample.

Performance of the logistic regression models identifying internalizing diagnosis using elevated T scores (T score $\geq$ 70 and T score $\geq$ 55) for the internalizing broadband and two DSM-oriented scales (anxiety problems, depressive problems) of the CBCL are reported in Table 3. Metrics include accuracy, sensitivity, specificity, and AUC.

Discussion

There is a significant need for a rapid and objective method for screening young children with internalizing disorders. We propose the use of data from a single wearable sensor during a 90-second fear induction task and machine learning to fulfill this need. Herein, we take an

Table 3. Performance characteristics of models developed for detecting children with internalizing diagnoses from parent-reported child problems measured by the CBCL.

	Accuracy		Sensitivity		Specificity		AUC
Cutoff	70	55	70	55	70	55	70 & 55
Internalizing T Score	.68	.73	.00	.26	1.00	.95	.76
Anxiety Problems T Score	.75	.73	.21	.42	1.00	.88	.75
Depressive Problems T Score	.70	.72	.05	.26	1.00	.93	.79

Accuracy, sensitivity, specificity, and area under the ROC curve (AUC) for logistic regression models on parent-reported internalizing problems over the 6 months leading up to participation in the study as measured by the Child Behavior Checklist. Also included are two subscales of total internalizing problems (Anxiety Problems, Depressive Problems) oriented to DSM-IV criteria [56].

https://doi.org/10.1371/journal.pone.0210267.t003

initial step toward this goal by training classifiers for detecting early indications of internalizing diagnoses using data sampled from each of three conceptually-segmented [54] temporal phases of a mood induction task, establishing their performance, and discussing the implications of these results. We further examine the specific features identified as being especially indicative of an internalizing diagnosis and discuss the behaviors described by these features in the context of internalizing disorders. The proposed approach is the first step toward creating an objective method for screening children for internalizing diagnoses rapidly and at low cost.

It is important to place these results in the context of previous work and existing diagnostic techniques. For example, these results compare favorably to our previous work, where k-nearest neighbor (k = 3) and logistic regression models were able to achieve 75% and 80% accuracy, respectively. The logistic regression employed herein achieves 81% accuracy, and we further report sensitivity (.67), specificity (.88), and ROC curves, quantities especially important when considering development of a tool used to screen for psychopathology. Of note, this logistic regression can be optimized for screening by adjusting the score threshold to yield an increase in sensitivity (.67 to .81) at the expense of a slight decrease in specificity (.88 to .81). Moreover, these results help to advance the use of wearable sensors and machine learning in childhood digital mental health, a burgeoning field that promises to improve access to, and speed of, mental healthcare. According to guides for classifying the accuracy of a diagnostic-screening test where AUC's under .60 are considered fail, .60-.70 are poor, .70-.80 are moderate, .80-.90 are good, and .90–1 are excellent [66], IMU-derived features during the Potential Threat phase (AUC = .85) are considered good.

Investigation into specific temporal phases of the mood induction task demonstrated that it was the Potential Threat phase that differentiated between children with and without internalizing diagnoses as evidenced by the results of Table 2 and Fig 1. Children with internalizing disorders tended to turn away ($\gamma \cong 180°$) from the ambiguous threat only when they were physically closest in the last 10 seconds of the phase (see Fig 3A). This may suggest that across depressive, anxiety and stress-related disorders, there is a shared anticipatory threat response which manifests physically in young children. Previous literature speculates that this type of response may be due to attention avoidance (attending away from the threat as shown in children with trauma and PTSD [67]) or emotional dysregulation (from attending to the threat in previous moments) [68].

Interestingly, the acute threat response during the Startle phase did not demonstrate clear differences by diagnostic status (e.g., see results of Figs 2 and 3 and Table 2). This could be due to heterogeneity of startle response across internalizing disorders, as previous research finds adults with moderate severity internalizing disorders (i.e., specific phobia) had heightened startles, healthy controls had low startle responses, and those with severe internalizing

disorders (i.e., GAD, PTSD) had blunted startle responses [69]. A similar phenomenon may exist in our data, however larger sample sizes are needed to better assess this possibility. Alternatively, the significantly heightened avoidance motion (i.e., γ) during Potential Threat without a significantly heightened motion during Startle could suggest a physiological manifestation of "It wasn't as bad as I thought it was going to be," a cognitive shift often seen in anxious children after exposure therapy [70]. Regardless of other phases, ambiguous threat avoidance during potential threat contexts appears to unify internalizing disorders and differentiate them from controls (e.g., see results of Table 2 and Fig 3B).

We compare psychometric properties of the questionnaire-based parent-reported CBCL and IMU-derived feature models on child internalizing diagnosis as determined via K-SADS-PL with clinical consensus. CBCL-derived models for both cutoffs (55 and 70) exhibited slightly lower classification accuracy (.68-.75 vs. .81), slightly higher specificity (.88–1.00 vs. .88), lower sensitivity (.00-.42 vs. .67), and slightly lower AUCs (.75-.79 vs. .85) compared to IMU-derived models during the Potential Threat phase. CBCL subscale psychometrics in our study are similar to those from much larger studies (e.g., see [66,71]). Notably, both in our study and paralleled in these previous studies is the varied sensitivity of the CBCL, with some samples exhibiting sensitivities as low as .00-.38 [72] and some as high as .44 to .86 [73]. Overall, CBCL internalizing psychometrics across studies suggest room for improvement in internalizing screening efficiency, especially as it was consistently worse than that of externalizing screening efficiency [71]. The IMU-based results presented herein yield a minimum 60% improvement in sensitivity over that observed from the CBCL suggesting that this supplemental objective data may be especially helpful for increasing sensitivity.

Overall, this paper describes a methodology requiring very limited computational resources (e.g. compute 10 features from 20 seconds of wearable sensor data, use as input to a logistic regression) which points toward future deployment of this technique for identifying young children with internalizing disorders using resource-constrained but ubiquitous devices like mobile phones. This new approach reduces the time required for diagnostic screening while also establishing high sensitivity–which can help to reduce barriers and better alert families to the need for child mental health services. While these results can likely be improved and extended, and should be replicated, this is an important first step in connecting often overlooked children [20,21] to the help they need to both mitigate their current distress and prevent subsequent comorbid emotional disorders and additional negative sequelae [12,14,74].

Limitations

This study is not without limitations. Future research should replicate and investigate our claims in a larger study with subjects at varying levels of risk for developing an internalizing disorder. Additionally, a larger sample size would allow examination of internalizing disorders without the presence of comorbid externalizing disorders, and also specific internalizing disorders to explore whether one disorder type yielded different motions than another. Future work may also explore alternative or additional device locations and more complex non-linear models for improving classification performance.

Conclusion

The results presented herein demonstrate that, when paired with machine learning, 20 seconds of wearable sensor data extracted from a fear induction task can be used to identify young children with internalizing disorders with a high level of accuracy, sensitivity, and specificity. These results point toward the future use of this approach for screening children for internalizing disorders.

Supporting information

S1 File. IMU features, CBCL scales, and clinical diagnosis.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Ryan S. McGinnis, Ellen W. McGinnis, Nestor L. Lopez-Duran, Kate Fitzgerald, Katherine L. Rosenblum, Maria Muzik.

Data curation: Ryan S. McGinnis, Ellen W. McGinnis, Jessica Hruschak.

Formal analysis: Ryan S. McGinnis, Ellen W. McGinnis.

Funding acquisition: Ellen W. McGinnis, Nestor L. Lopez-Duran, Kate Fitzgerald, Katherine L. Rosenblum, Maria Muzik.

Investigation: Ryan S. McGinnis, Ellen W. McGinnis, Jessica Hruschak, Nestor L. Lopez-Duran, Kate Fitzgerald, Katherine L. Rosenblum, Maria Muzik.

Methodology: Ryan S. McGinnis, Ellen W. McGinnis, Jessica Hruschak, Nestor L. Lopez-Duran, Kate Fitzgerald, Katherine L. Rosenblum, Maria Muzik.

Project administration: Ellen W. McGinnis, Jessica Hruschak, Kate Fitzgerald, Katherine L. Rosenblum, Maria Muzik.

Resources: Ryan S. McGinnis, Ellen W. McGinnis, Nestor L. Lopez-Duran, Kate Fitzgerald, Katherine L. Rosenblum, Maria Muzik.

Software: Ryan S. McGinnis, Ellen W. McGinnis.

Supervision: Ellen W. McGinnis, Nestor L. Lopez-Duran, Kate Fitzgerald, Katherine L. Rosenblum, Maria Muzik.

Validation: Ryan S. McGinnis, Ellen W. McGinnis.

Visualization: Ryan S. McGinnis, Ellen W. McGinnis.

Writing – original draft: Ryan S. McGinnis, Ellen W. McGinnis.

Writing – review & editing: Ryan S. McGinnis, Ellen W. McGinnis, Jessica Hruschak, Nestor L. Lopez-Duran, Kate Fitzgerald, Katherine L. Rosenblum, Maria Muzik.

References

1. Egger HL, Angold A. Common emotional and behavioral disorders in preschool children: presentation, nosology, and epidemiology. J Child Psychol Psychiatry. 2006; 47: 313–337. https://doi.org/10.1111/j.1469-7610.2006.01618.x PMID: 16492262

2. Bufferd SJ, Dougherty LR, Carlson GA, Klein DN. Parent-Reported Mental Health in Preschoolers: Findings Using a Diagnostic Interview. Compr Psychiatry. 2011; 52: 359–369. https://doi.org/10.1016/j.comppsych.2010.08.006 PMID: 21683173

3. Luby JL, Si X, Belden AC, Tandon M, Spitznagel E. Preschool depression: Homotypic continuity and course over 24 months. Archives of General Psychiatry. 2009; 66: 897–905. https://doi.org/10.1001/archgenpsychiatry.2009.97 PMID: 19652129

4. Tandon M, Cardeli E, Luby J. Internalizing Disorders in Early Childhood: A Review of Depressive and Anxiety Disorders. Child and Adolescent Psychiatric Clinics of North America. 2009; 18: 593–610. https://doi.org/10.1016/j.chc.2009.03.004 PMID: 19486840

5. Luby JL, Belden AC, Pautsch J, Si X, Spitznagel E. The clinical significance of preschool depression: Impairment in functioning and clinical markers of the disorder. Journal of Affective Disorders. 2009; 112: 111–119. https://doi.org/10.1016/j.jad.2008.03.026 PMID: 18486234

6. Towe-Goodman NR, Franz L, Copeland W, Angold A, Egger H. Perceived family impact of preschool anxiety disorders. J Am Acad Child Adolesc Psychiatry. 2014; 53: 437–446. https://doi.org/10.1016/j.jaac.2013.12.017 PMID: 24655653

7. Belden AC, Gaffrey MS, Luby JL. Relational Aggression in Children With Preschool-Onset Psychiatric Disorders. Journal of the American Academy of Child & Adolescent Psychiatry. 2012; 51: 889–901. https://doi.org/10.1016/j.jaac.2012.06.018 PMID: 22917202

8. Bufferd SJ, Dougherty LR, Carlson GA, Rose S, Klein DN. Psychiatric disorders in preschoolers: continuity from ages 3 to 6. Am J Psychiatry. 2012; 169: 1157–1164. https://doi.org/10.1176/appi.ajp.2012.12020268 PMID: 23128922

9. Whalen DJ, Sylvester CM, Luby JL. Depression and Anxiety in Preschoolers: A Review of the Past 7 Years. Child and Adolescent Psychiatric Clinics of North America. 2017; 26: 503–522. https://doi.org/10.1016/j.chc.2017.02.006 PMID: 28577606

10. Wittchen H-U, Sonntag H. Nicotine consumption in mental disorders: a clinical epidemiological perspective. European Neuropsychopharmacology. 2000; 10: 119. https://doi.org/10.1016/S0924-977X(00)80014-0

11. Compton SN, Burns BJ, Helen LE, Robertson E. Review of the evidence base for treatment of childhood psychopathology: internalizing disorders. J Consult Clin Psychol. 2002; 70: 1240–1266. PMID: 12472300

12. Bittner A, Egger HL, Erkanli A, Jane Costello E, Foley DL, Angold A. What do childhood anxiety disorders predict? J Child Psychol Psychiatry. 2007; 48: 1174–1183. https://doi.org/10.1111/j.1469-7610.2007.01812.x PMID: 18093022

13. Beesdo K, Bittner A, Pine DS, Stein MB, Höfler M, Lieb R, et al. Incidence of social anxiety disorder and the consistent risk for secondary depression in the first three decades of life. Arch Gen Psychiatry. 2007; 64: 903–912. https://doi.org/10.1001/archpsyc.64.8.903 PMID: 17679635

14. Cole DA, Peeke LG, Martin JM, Truglio R, D A. A longitudinal look at the relation between depression and anxiety in children and adolescents. Journal of Consulting and Clinical Psychology. 1998; 66: 451–460. https://doi.org/10.1037/0022-006X.66.3.451 PMID: 9642883

15. GOULD MS, KING R, GREENWALD S, FISHER P, SCHWAB-STONE M, KRAMER R, et al. Psychopathology Associated With Suicidal Ideation and Attempts Among Children and Adolescents. Journal of the American Academy of Child & Adolescent Psychiatry. 1998; 37: 915–923. https://doi.org/10.1097/00004583-199809000-00011 PMID: 9735611

16. Craske MG, Stein MB. Anxiety. The Lancet. 2016; 388: 3048–3059. https://doi.org/10.1016/S0140-6736(16)30381-6

17. Rapee RM, Schniering CA, Hudson JL. Anxiety Disorders During Childhood and Adolescence: Origins and Treatment. Annual Review of Clinical Psychology. 2009; 5: 311–341. https://doi.org/10.1146/annurev.clinpsy.032408.153628 PMID: 19152496

18. Konnopka A, Leichsenring F, Leibing E, König H-H. Cost-of-illness studies and cost-effectiveness analyses in anxiety disorders: A systematic review. Journal of Affective Disorders. 2009; 114: 14–31. https://doi.org/10.1016/j.jad.2008.07.014 PMID: 18768222

19. Luby JL. Preschool Depression: The Importance of Identification of Depression Early in Development. Current Directions in Psychological Science. 2010; 19: 91–95. https://doi.org/10.1177/0963721410364493 PMID: 21969769

20. Pavuluri MN, Luk SL, McGee R. Help-seeking for behavior problems by parents of preschool children: a community study. J Am Acad Child Adolesc Psychiatry. 1996; 35: 215–222. https://doi.org/10.1097/00004583-199602000-00015 PMID: 8720631

21. Mian ND. Little children with big worries: addressing the needs of young, anxious children and the problem of parent engagement. Clin Child Fam Psychol Rev. 2014; 17: 85–96. https://doi.org/10.1007/s10567-013-0152-0 PMID: 23949334

22. Kataoka SH, Zhang L, Wells KB. Unmet need for mental health care among U.S. children: variation by ethnicity and insurance status. Am J Psychiatry. 2002; 159: 1548–1555. https://doi.org/10.1176/appi.ajp.159.9.1548 PMID: 12202276

23. Lavigne JV, Lebailly SA, Hopkins J, Gouze KR, Binns HJ. The prevalence of ADHD, ODD, depression, and anxiety in a community sample of 4-year-olds. J Clin Child Adolesc Psychol. 2009; 38: 315–328. https://doi.org/10.1080/15374410902851382 PMID: 19437293

24. Luby JL, Mrakotsky C, Heffelfinger A, Brown K, Hessler M, Spitznagel E. Modification of DSM-IV Criteria for Depressed Preschool Children. The American Journal of Psychiatry. 2003; 160: 1169–1172. https://doi.org/10.1176/appi.ajp.160.6.1169 PMID: 12777277

25. Luby JL, Heffelfinger AK, Mrakotsky C, Brown KM, Hessler MJ, Wallis JM, et al. The clinical picture of depression in preschool children. Journal of the American Academy of Child & Adolescent Psychiatry. 2003; 42: 340–348. https://doi.org/10.1097/00004583-200303000-00015 PMID: 12595788

26. Kaminer Y, Feinstein C, Seifer R. Is there a need for observationally based assessment of affective symptomatology in child and adolescent psychiatry? Adolescence. 1995; 30: 483–489. PMID: 7676882

27. Garber J, Kaminski KM. Laboratory and performance-based measures of depression in children and adolescents. Journal of clinical child psychology. 2000; 29: 509–525. https://doi.org/10.1207/S15374424JCCP2904_5 PMID: 11126630

28. Chansky TE, Kendall PC. Social expectancies and self-perceptions in anxiety-disordered children. J Anxiety Disord. 1997; 11: 347–363. PMID: 9276781

29. Herjanic B, Reich W. Development of a structured psychiatric interview for children: Agreement between child and parent on individual symptoms. Journal of Abnormal Child Psychology: An official publication of the International Society for Research in Child and Adolescent Psychopathology. 1997; 25: 21–31. https://doi.org/10.1023/A:1025703323438

30. Kolko DJ, Kazdin AE. Emotional/behavioral problems in clinic and nonclinic children: Correspondence among child, parent and teacher reports. Journal of Child Psychology and Psychiatry. 1993; 34: 991–1006. https://doi.org/10.1111/j.1469-7610.1993.tb01103.x PMID: 8408380

31. Renouf AG, Kovacs M. Concordance between mothers' reports and children's self-reports of depressive symptoms: A longitudinal study. Journal of the American Academy of Child & Adolescent Psychiatry. 1994; 33: 208–216. https://doi.org/10.1097/00004583-199402000-00008 PMID: 8150792

32. De Los Reyes A, Augenstein TM, Wang M, Thomas SA, Drabick DAG, Burgers DE, et al. The validity of the multi-informant approach to assessing child and adolescent mental health. Psychol Bull. 2015; 141: 858–900. https://doi.org/10.1037/a0038498 PMID: 25915035

33. Lord C, Risi S, Lambrecht L, Cook EH Jr, Leventhal BL, DiLavore PC, et al. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. J Autism Dev Disord. 2000; 30: 205–223. PMID: 11055457

34. Mash EJ, Foster SL. Exporting analogue behavioral observation from research to clinical practice: useful or cost-defective? Psychol Assess. 2001; 13: 86–98. PMID: 11281042

35. Durbin CE, Hayden EP, Klein DN, Olino TM. Stability of laboratory-assessed temperamental emotionality traits from ages 3 to 7. Emotion. 2007; 7: 388–399. https://doi.org/10.1037/1528-3542.7.2.388 PMID: 17516816

36. Adrian M, Zeman J, Veits G. Methodological implications of the affect revolution: a 35-year review of emotion regulation assessment in children. J Exp Child Psychol. 2011; 110: 171–197. https://doi.org/10.1016/j.jecp.2011.03.009 PMID: 21514596

37. Martin M. On the Induction of Mood. Clin Psychol Rev. 1990; 10: 669–697. https://doi.org/10.1016/0272-7358(90)90075-L

38. Chorney JM, McMurtry CM, Chambers CT, Bakeman R. Developing and Modifying Behavioral Coding Schemes in Pediatric Psychology: A Practical Guide. J Pediatr Psychol. 2015; 40: 154–164. https://doi.org/10.1093/jpepsy/jsu099 PMID: 25416837

39. McGinnis RS, McGinnis EW, Hruschak J, Lopez-Duran N, Fitzgerald K, Rosenblum K, et al. Wearable Sensors and Machine Learning Diagnose Anxiety and Depression in Young Children. 2018 IEEE International Conference on Biomedical and Health Informatics (BHI). Las Vegas, NV; 2018.

40. McGinnis E, McGinnis R, Muzik M, Hruschak J, Lopez-Duran N, Perkins N, et al. Movements indicate threat response phases in children at-risk for anxiety. IEEE Journal of Biomedical and Health Informatics. 2017; 21: 1460–1465. https://doi.org/10.1109/JBHI.2016.2603159 PMID: 27576271

41. McGinnis EW, McGinnis RS, Hruschak J, Bilek E, Ip K, Morlen D, et al. Wearable sensors detect childhood internalizing disorders during mood induction task. PLOS ONE. 2018; 13: e0195598. https://doi.org/10.1371/journal.pone.0195598 PMID: 29694369

42. McGinnis RS, McGinnis EW, Hruschak J, Lopez-Duran N, Fitzgerald K, Rosenblum K, et al. Rapid Anxiety and Depression Diagnosis in Young Children Enabled by Wearable Sensors and Machine Learning. 2018 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Honolulu, HI; 2018.

43. McGinnis RS, Mahadevan N, Moon Y, Seagers K, Sheth N, Wright JA, et al. A machine learning approach for gait speed estimation using skin-mounted wearable sensors: From healthy controls to individuals with multiple sclerosis. PLoS One. 2017; 12. https://doi.org/10.1371/journal.pone.0178366 PMID: 28570570

44. Moon Y, McGinnis RS, Seagers K, Motl RW, Sheth N, Jr JAW, et al. Monitoring gait in multiple sclerosis with novel wearable motion sensors. PLOS ONE. 2017; 12: e0171346. https://doi.org/10.1371/journal.pone.0171346 PMID: 28178288

45. Sun R, Moon Y, McGinnis RS, Seagers K, Motl RW, Sheth N, et al. Assessment of Postural Sway in Individuals with Multiple Sclerosis Using a Novel Wearable Inertial Sensor. DIB. 2018; 2: 1–10. https://doi.org/10.1159/000485958

46. Son D, Lee J, Qiao S, Ghaffari R, Kim J, Lee JE, et al. Multifunctional wearable devices for diagnosis and therapy of movement disorders. Nature Nanotechnology. 2014; 9: 397–404. https://doi.org/10.1038/nnano.2014.38 PMID: 24681776

47. Patel S, Lorincz K, Hughes R, Huggins N, Growdon J, Standaert D, et al. Monitoring Motor Fluctuations in Patients with Parkinson's Disease Using Wearable Sensors. IEEE Trans Inf Technol Biomed. 2009; 13: 864–873. https://doi.org/10.1109/TITB.2009.2033471 PMID: 19846382

48. Lee J, Reyes BA, McManus DD, Maitas O, Chon KH. Atrial Fibrillation Detection Using an iPhone 4S. IEEE Transactions on Biomedical Engineering. 2013; 60: 203–206. https://doi.org/10.1109/TBME.2012.2208112 PMID: 22868524

49. Elenko E, Underwood L, Zohar D. Defining digital medicine. In: Nature Biotechnology [Internet]. 12 May 2015 [cited 21 May 2018]. https://doi.org/10.1038/nbt0715-683a

50. Topol EJ, Steinhubl SR, Torkamani A. Digital Medical Tools and Sensors. JAMA. 2015; 313: 353–354. https://doi.org/10.1001/jama.2014.17125 PMID: 25626031

51. Digital technology for treating and preventing mental disorders in low-income and middle-income countries: a narrative review of the literature [Internet]. [cited 21 May 2018]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5523650/

52. Calkins SD, Graziano PA, Berdan LE, Keane SP, Degnan KA. Predicting cardiac vagal regulation in early childhood from maternal-child relationship quality during toddlerhood. Dev Psychobiol. 2008; 50: 751–766. https://doi.org/10.1002/dev.20344 PMID: 18814182

53. Lopez-Duran NL, Hajal NJ, Olson SL, Felt BT, Vazquez DM. Individual differences in cortisol responses to fear and frustration during middle childhood. J Exp Child Psychol. 2009; 103: 285–295. https://doi.org/10.1016/j.jecp.2009.03.008 PMID: 19410263

54. Gross JJ. Emotion regulation: Affective, cognitive, and social consequences. Psychophysiology. 2002; 39: 281–291. https://doi.org/10.1017/S0048577201393198 PMID: 12212647

55. Achenbach TM, Howell CT, Quay HC, Conners CK. National survey of problems and competencies among four- to sixteen-year-olds: Parents' reports for normative and clinical samples. Monographs of the Society for Research in Child Development. 1991; 56: v–120. https://doi.org/10.2307/1166156

56. Achenbach TM, Rescorla LA. ASEBA School Age Forms and Profiles. Burlington, Vt: ASEBA. 2001; Available: http://www.childbehaviorchecklist.com/ordering/ASEBA%20Reliability%20and%20Validity-School%20Age.pdf

57. Gaffrey MS, Luby JL. Kiddie Schedule for Affective Disorders and Schizophrenia- Early Childhood Version (K-SADS-EC). St Louis, MO: Washington University School of Medicine. 2012;

58. Maziade M, Roy MA, Fournier JP, Cliche D, Mérette C, Caron C, et al. Reliability of best-estimate diagnosis in genetic linkage studies of major psychoses: results from the Quebec pedigree studies. Am J Psychiatry. 1992; 149: 1674–1686. https://doi.org/10.1176/ajp.149.12.1674 PMID: 1443244

59. McGinnis RS, Cain SM, Davidson SP, Vitali RV, McLean SG, Perkins NC. Validation of Complementary Filter Based IMU Data Fusion for Tracking Torso Angle and Rifle Orientation. Montreal, QC; 2014. p. V003T03A052. https://doi.org/10.1115/IMECE2014-36909

60. McGinnis R. GitHub repository child-mental-health [Internet]. 2017. Available: https://github.com/ryan-mcginnis/child-mental-health

61. Davies DL, Bouldin DW. A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1979; PAMI-1: 224–227. https://doi.org/10.1109/TPAMI.1979.4766909

62. Hudziak JJ, Copeland W, Stanger C, Wadsworth M. Screening for DSM-IV externalizing disorders with the Child Behavior Checklist: a receiver-operating characteristic analysis. J Child Psychol Psychiatry. 2004; 45: 1299–1307. https://doi.org/10.1111/j.1469-7610.2004.00314.x PMID: 15335349

63. Fawcett T. An Introduction to ROC Analysis. Pattern Recogn Lett. 2006; 27: 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

64. Santafe G, Inza I, Lozano JA. Dealing with the evaluation of supervised classification algorithms. Artif Intell Rev. 2015; 44: 467–508. https://doi.org/10.1007/s10462-015-9433-y

65. Isaksson A, Wallman M, Göransson H, Gustafsson MG. Cross-validation and Bootstrapping Are Unreliable in Small Sample Classification. Pattern Recogn Lett. 2008; 29: 1960–1965. https://doi.org/10.1016/j.patrec.2008.06.018

66. de la Osa N, Granero R, Trepat E, Domenech JM, Ezpeleta L. The discriminative capacity of CBCL/1½-5-DSM5 scales to identify disruptive and internalizing disorders in preschool children. Eur Child Adolesc Psychiatry. 2016; 25: 17–23. https://doi.org/10.1007/s00787-015-0694-4 PMID: 25715996

67. Pine DS, Mogg K, Bradley BP, Montgomery L, Monk CS, McClure E, et al. Attention Bias to Threat in Maltreated Children: Implications for Vulnerability to Stress-Related Psychopathology. AJP. 2005; 162: 291–296. https://doi.org/10.1176/appi.ajp.162.2.291 PMID: 15677593

68. Vasey MW, el-Hag N, Daleiden EL. Anxiety and the processing of emotionally threatening stimuli: distinctive patterns of selective attention among high- and low-test-anxious children. Child Dev. 1996; 67: 1173–1185. PMID: 8706516

69. McTeague LM, Lang PJ. THE ANXIETY SPECTRUM AND THE REFLEX PHYSIOLOGY OF DEFENSE: FROM CIRCUMSCRIBED FEAR TO BROAD DISTRESS. Depress Anxiety. 2012; 29: 264–281. https://doi.org/10.1002/da.21891 PMID: 22511362

70. journalist DC is a freelance, Ell senior content producer with content strategy firm, Partners R. Fear of Vomiting. In: Child Mind Institute [Internet]. [cited 21 May 2018]. Available: https://childmind.org/article/fear-of-vomiting/

71. Warnick EM, Bracken MB, Kasl S. Screening Efficiency of the Child Behavior Checklist and Strengths and Difficulties Questionnaire: A Systematic Review. Child and Adolescent Mental Health. 2008; 13: 140–147. https://doi.org/10.1111/j.1475-3588.2007.00461.x

72. Rishel CW, Greeno C, Marcus SC, Shear MK, Anderson C. Use of the Child Behavior Checklist as a Diagnostic Screening Tool in Community Mental Health. Research on Social Work Practice. 2005; 15: 195–203. https://doi.org/10.1177/1049731504270382

73. Aschenbrand SG, Angelosante AG, Kendall PC. Discriminant validity and clinical utility of the CBCL with anxiety-disordered youth. J Clin Child Adolesc Psychol. 2005; 34: 735–746. https://doi.org/10.1207/s15374424jccp3404_15 PMID: 16232070

74. Kendall PC, Safford S, Flannery-Schroeder E, Webb A. Child anxiety treatment: outcomes in adolescence and impact on substance use and depression at 7.4-year follow-up. J Consult Clin Psychol. 2004; 72: 276–287. https://doi.org/10.1037/0022-006X.72.2.276 PMID: 15065961