2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)

m-Health and Autism: Recognizing Stress and Anxiety with Machine Learning and Wearables Data

Aaron J. Masino
Department of Anesthesiology & Critical Care Medicine
Perelman School of Medicine
University of Pennsylvania
Philadelphia, United States
masinoa@email.chop.edu

Daniel Forsyth
Department of Biomedical & Health Informatics
Children's Hospital of Philadelphia
Philadelphia, United States
forsythd@email.chop.edu

Heather Nuske
Department of Psychiatry
Perelman School of Medicine
University of Pennsylvania
Philadelphia, United States
hjnuske@upenn.edu

John Herrington
Department of Psychiatry
Perelman School of Medicine
University of Pennsylvania
Philadelphia, United States
herringtonj@email.chop.edu

Jeffrey Pennington
Department of Biomedical & Health Informatics
Children's Hospital of Philadelphia
Philadelphia, United States
penningtonj@email.chop.edu

Yelena Kushleyeva
Department of Biomedical & Health Informatics
Children's Hospital of Philadelphia
Philadelphia, United States

Christopher P. Bonafide
Department of Pediatrics
Perelman School of Medicine
University of Pennsylvania
Philadelphia, United States
bonafide@email.chop.edu

***Abstract*—Consumer-grade wearables provide physiological measurements which may inform m-health applications that predict adverse outcomes. Autism Spectrum Disorder (ASD) represents a compelling example. Many individuals with ASD present with challenging behaviors that are preceded by physiological changes. Physiological measures could, therefore, support real-time interventions to avert challenging behaviors in various social settings. However, no prior research has demonstrated a methodological approach to detect these changes using wearable device data. We sought to demonstrate a machine learning approach that uses wearables data to differentiate physiological states associated with stressful and non-stressful scenarios in children with ASD. In a controlled laboratory setting, we collected heart rate and RR interval measurements during rest and during activities designed to mimic stress using a consumer-grade wearable device. Our analysis included 38 participants (22 ASD, 16 non-ASD). Following outlier removal, we extracted 20 statistical features from data collected during each patient's rest and stressful periods. Using nested leave-one-out cross-validation over 76 sample periods (38 rest / 38 stress), we trained and evaluated logistic regression (LR) and support vector machine (SVM) classifiers to label each validation sample as a rest or stressful period. The SVM and LR models achieved 93% and 87% accuracy, respectively. These results suggest that machine learning models combined with wearables data may support real-time m-health intervention applications.**

Keywords— machine learning, m-Health, wearables, autism

I. Introduction

The rapidly expanding market of consumer-grade wearable devices has made acquisition of health data in non-clinical settings widely available at low cost. Wearables generate physiological parameters such as heart rate (HR), heart rate variability (HRV), oxygen saturation, and skin temperature that could be processed to detect early physiological changes that precede adverse events. Autism Spectrum Disorder (ASD) provides a compelling example of the potential benefits wearable devices could offer to children. Many children with ASD present with challenging behaviors, such as aggression directed towards other, objects, or themselves (i.e., self-injury) [1,2]. Challenging behaviors in children with ASD have a devastating impact on family wellbeing [3], predict teacher burnout [4], and are the most frequent reason youth with ASD are hospitalized [5]. These behaviors are difficult to predict as children with ASD often cannot communicate distress [6]. Single-subject research on adults with ASD performed in laboratory settings shows that changes in physiologic parameters induced by stress predict challenging behaviors [7,8]. Moreover, recent work by our team and that of others has found that physiological indices of stress, such as HR increase, were associated with onset of challenging behavior episodes [9,10]. Wearable monitoring devices, therefore, can provide actionable decision support to caregivers allowing them to intervene before the challenging behaviors escalate and cause harm.

In prior research, a study of 5 children with ASD showed that lower HRV measures derived from consumer-grade wearables were associated with engagement in sociocognitive tasks [11]. Smartwatch accelerometer data was used in a machine learning algorithm to detect hand flapping, painting, and sibbing (self-injury) in a study that included 15 subjects (2 with ASD) [12]. In a similar study of 2 individuals with ASD, smartwatch accelerometer, heart rate, and gyroscope data was used to train a machine learning algorithm to detect the moment of behavioral crises [13]. To date, however, no prior study has demonstrated a methodological approach using machine learning to predict challenging behaviors before they happen. As a step toward that objective, we sought to demonstrate a machine learning approach that uses data from wearables to differentiate physiological states associated with stressful and non-stressful scenarios in a larger cohort of children than the studies above, including subjects with and without ASD.

2372-9198/19/$31.00 ©2019 IEEE
DOI 10.1109/CBMS.2019.00144

In the remainder of this paper, we describe the development of two machine learning models tasked with differentiating rest and stressful task scenarios given statistical features of HR and beat to beat (RR) interval data collected from wearable devices in individuals with and without ASD. We present an evaluation of model performance that suggests machine learning models are able to recognize these two states with high reliability using wearables data.

II. Methods

All participants provided informed consent following study procedures approved by the Institutional Review Board at the University of Pennsylvania and Children's Hospital of Philadelphia.

A. Study Setting & Participants

We implemented a prospective observational study with a cohort of 32 children with ASD and 23 children without ASD. Of these, continuous HR and RR data was collected for 44 children (26 with ASD and 18 without ASD). Diagnoses were confirmed with a research-reliable clinician using the Autism Diagnostic Observation Schedule – Second Edition [14]. The study was performed in a controlled psychology laboratory setting at Children's Hospital of Philadelphia.

B. Data Collection & Processing

All study subjects were fitted with a consumer-grade physiological data monitor with a chest strap electrocardiographic sensor. The same make and model device was used for all subjects. After an initial rest period of 7 minutes, during which children viewed a relaxing video designed to obtain resting state physiology (*InScapes*) [15], subjects were engaged in two tasks designed to mimic stressful scenarios. These stress tasks (Transparent Box and Tangrams/Tangoes) were taken from the middle childhood version of the Laboratory Temperament Assessment Battery [16]. In the Transparent Box task, children were given incorrect keys to the box for 4 minutes (and then the correct keys were given). In the Tangrams task, children were asked to complete two sets of tangram puzzles, each with an extra piece (making them unsolvable). They were given 5 minutes, timed with a sand timer, to complete the puzzles (and then they were shown the correct configuration).

HR and RR interval data was collected during the entire observation period which included the initial rest and task periods (see Fig. 1). For the analysis in this study, data collected during the two task periods were concatenated to form a single sample of task period observations for each participant. Following standard Bluetooth low-energy protocols, the monitor buffered HR and RR interval data and transmitted it to the recording computer at a rate of approximately one transmission per second. Sample rest periods contained a mean of 586 HR and 584 RR interval measurements per study subject. The concatenated sample stress periods contained a mean of 789 HR and 792 RR interval measurements per subject.

Although study subjects were asked to remain seated during the observation session, they were not prevented from moving around. Therefore, to alleviate the possibility that our models might detect movement, rather than stress, induced physiological changes, we excluded data from subjects that were out of their seat for more than 15% of the task or rest periods. This resulted in removal of data for 6 subjects in this analysis. Our final dataset consisted of HR and RR interval data for 38 participants (22 with ASD) collected during one rest period and the concatenated task-induced stress periods (i.e. 76 total sample periods).

Preliminary data analysis indicated the presence of potential outliers. As such values could potentially bias machine learning models, we removed values less than the lower bound, *LB*, or greater than the upper bound, *UB*, defined by

$$LB = [Q1 - 3 * IQR] \quad (1)$$

$$UB = [Q3 + 3 * IQR] \quad (2)$$

where *Q1* and *Q3* are the middle value of the first and second half, respectively, of the ordered data and *IQR* is the interquartile range. Values for *LB* and *UB* were individually calculated for each study participant for both HR and RR intervals using data collected for the entire observation period. For the 38 participants in the final dataset, this procedure led to the removal of a mean of 2.1 RR rest period measurements. No HR

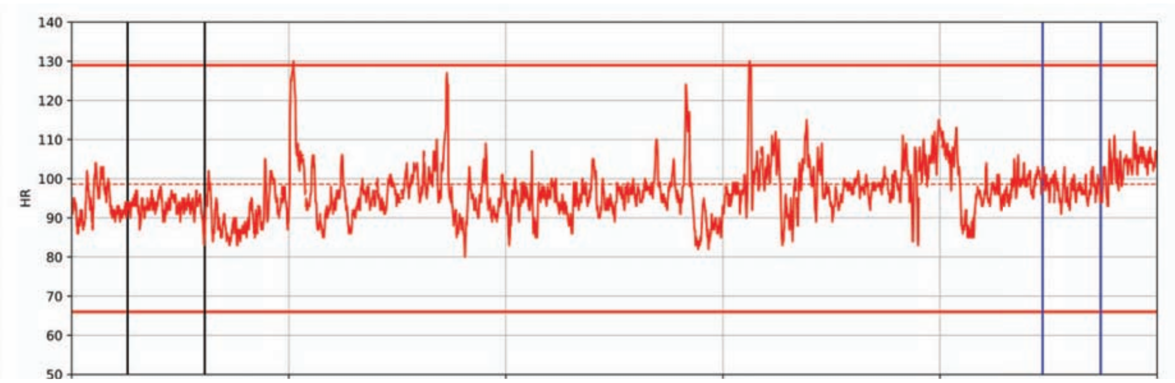

Fig. 1. Sample HR measurements collected from consumer-grade wearable. X-axis is time in minutes from session start. Vertical black lines (left) indicate rest period boundaries. Vertical blue lines (right) indicate one of the task period boundaries. Horizontal redlines indicate outlier boundaries.

Funded by Children's Hospital of Philadelphia Foerderer Award for Excellence.

measurements were removed during the rest period. For the task period, a mean of 8.7 HR and 5.2 RR measurements were removed.

C. *Feature Engineering*

For model input, HR and RR observations collected during the rest and task periods were converted to time-domain statistical features of HRV. Specifically, we calculated the maximum, minimum, mean, standard deviation, skew, and kurtosis for each patient for both HR and RR interval measurements collected during each of the 76 sample periods. We also calculated the root mean square of successive differences between normal heartbeats (RMSSD), the number of successive NN intervals (intervals between normal RR peaks) that differ from each other by more than 20 (NN20) or 50 (NN50) milliseconds (ms), the percentage of successive NN intervals that differ by more than 20 (pNN20) or 50 (pNN50) ms, the coefficient of variation of NN intervals (CVNN), and the coefficient of variation of successive differences of NN intervals (RMSSD / mean NN interval). In accordance with standard machine learning practice, all features were adjusted to have a standard normal distribution.

D. *Machine Learning Analysis*

We trained two machine learning classifier models to label input data as either occurring during a rest or stress (i.e. task) period: (1) logistic regression (LR) with L2 regularization; and (2) support vector machine (SVM) with a radial basis function (RBF) kernel. Due to our limited data size, we chose to train and evaluate each model using a nested leave-one-out (LOO) cross-validation (CV) procedure [17]. In nested LOO-CV, there is an outer evaluation loop and an inner hyperparameter selection and training loop. For each outer loop iteration, 1 sample is held out for evaluation. The remaining *N-1* samples are used in the inner loop to perform a grid search over a range of hyperparameters (Table I) with selection of the optimal parameters through a second LOO-CV performed in the inner loop. The optimal parameters and the *N-1* training samples are used to train the model which is then used to output a label for the current held-out sample. This process is repeated *N* times, so each model provides a label for all samples in the dataset. We used the *Python scikit-learn* library for all training and analysis [18].

TABLE I. HYPER-PARAMETER GRID SEARCH RANGES FOR LOGISTIC REGRESSION (LR) AND SUPPORT VECTOR MACHINE (SVM) CLASSIFIERS

Model	Parameter	Values Tested
LR	Inverse regularization, C	0.01, 0.1, 1, 10, 100
SVM	Inverse regularization, C	0.01, 0.1, 1, 10, 100
	Kernel coefficeint, γ	0.01, 0.1, 1, 10, 100

III. RESULTS

We implemented two machine learning classifiers with the objective of discriminating rest and stressful task scenarios given input statistical features derived from physiological data collected from a consumer-grade wearable device. The performance metrics by class label (rest *vs.* stress) obtained through LOO-CV are given in Table II. The LR and SVM models achieved overall accuracy of 87% and 93%, respectively. Considering only the 22 individuals with ASD, the LR and SVM models achieved 84% and 91% accuracy, respectively. The SVM model confusion matrix is given in Table III. Receiver operating characteristic (ROC) and corresponding area under the curve (AUC) are given in Fig. 2. McNemar's test indicates that the difference in performance of the two models is not statistically significant. For the LR model, the optimal regularization parameter, C, was 10 for 66 LOO-CV iterations and 100 for the remaining 10. For the SVM model, the optimal hyperparameters were C=100 and γ=0.01 for every iteration.

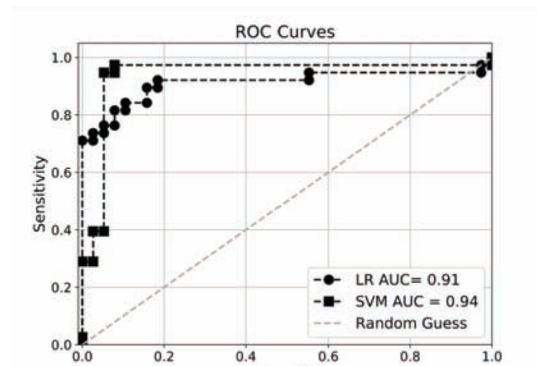

Fig. 2 Receiver operating characteristic (ROC) curves and area under the curve (AUC) for the LR and SVM models.

TABLE II. LR AND SVM CLASSIFIER PERFORMANCE BY CLASS (I.E. TREATING REST OR STRESS AS THE POSTIVE CLASS, RESPECTIVELY) CALCULATED FROM VALIDATION PREDICTIONS IN NESTED LOO-CV. METRICS: F1 – F1 SCORE, S – SENSITIVITY, P – PRECISION. VALUES IN SQUARE BRACKETS INDICATE 95% CONFIDENCE INTERVALS CALCULATED BY THE WILSON SOCRE INTERVAL. ROWS UNDER *ALL* INDICATE RESULTS FOR ENTIRE COHORT. ROWS UNDER *ASD ONLY* INDICATE RESULTS FOR INDIVIDUALS WITH ASD.

	All (N=38)		
Model - class	***F1***	***S***	***P***
LR - rest	0.86 [0.78, 0.94]	0.84 [0.76, 0.92]	0.89 [0.82, 0.96]
LR – stress	0.87 [0.79, 0.95]	0.89 [0.82, 0.96]	0.85 [0.77, 0.93]
SVM – rest	0.93 [0.87, 0.99]	0.89 [0.82, 0.96]	0.97 [0.93, 1]
SVM – stress	0.94 [0.89, 0.99]	0.97 [0.93, 1]	0.90 [0.83, 0.97]
	ASD Only (N=22)		
Model - class	***F1***	***S***	***P***
LR - rest	0.84 [0.76, 0.92]	0.82 [0.73, 0.91]	0.86 [0.78, 0.94]
LR – stress	0.84 [0.76, 0.92]	0.86 [0.78, 0.94]	0.83 [0.74, 0.91]
SVM – rest	0.90 [0.83, 0.97]	0.86 [0.78, 0.94]	0.95 [0.90, 1]
SVM – stress	0.91 [0.85, 0.97]	0.95 [0.90, 1]	0.88 [0.81, 0.95]

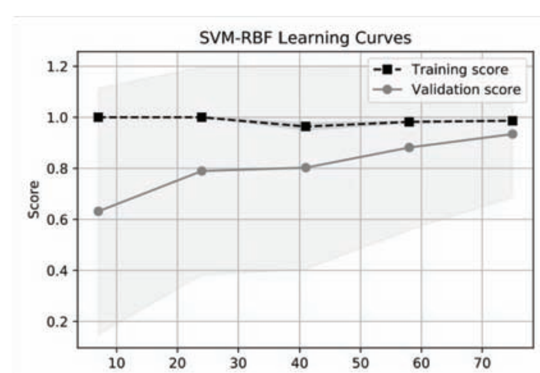

Fig. 3. SVM model learning curves. Symbols indicate mean value over 10 CV folds, shaded region indicates one standard deviation.

TABLE III. SVM MODEL CONFUSION MATRIX. STRESS IS CONSIDERED THE POSITIVE CLASS. *PREDICTED* REFERS TO MODEL OUTPUT. *ACTUAL* REFERS TO THE TRUE LABEL. NUMERICAL VALUES INDICATE NUMBER OF VALIDATION SAMPLES FROM THE ENTIRE COHORT (N=76) THAT BELONG TO THE GIVEN CELL. NUMBERS IN PARENTHESES INDICATE COUNTS FOR ASD INDIVIDUALS (N=22).

	Predicted	
Actual	***Positive***	***Negative***
Positive	37 (21)	1 (1)
Negative	4 (3)	34 (19)

To gain insight on the importance of the different HRV input features, we examined the logistic regression model coefficients. Table IV indicates the mean, minimum, and maximum coefficient values for the features with a mean magnitude greater than 1.0 as taken over all iterations of the nested LOO-CV procedure. Seven features had coefficients that changed sign (i.e. between positive and negative) during the procedure: CVSD; RMSSD; HR kurtosis; RR standard deviation, maximum, kurtosis, and skew. However, these coefficients all had a mean magnitude less than 0.50 and likely did not influence the classifier output.

Learning curves for the SVM model are shown in Fig. 3. The y-axis indicates the F1 score (harmonic average of sensitivity and precision) for which the optimal value is 1.0. The training and validation curves approach the same asymptote indicating variance (over-fitting) is not a significant factor. Similarly, the validation curve approaches the optimal metric value indicating a lack of bias, i.e. there is sufficient model capacity. A similar result (not shown) was observed for the logistic regression model.

TABLE IV. LOGISTIC REGRESSION COEFFICENTS FOR FEATURES WITH MEAN MAGNITUDE OF COEFFICIENT GREATER THAN 1.0 OVER N=76 ITERATIONS OF NESTED LOO-CV.

Feature	Mean [Min, Max]
pNN20	-5.85 [-14.96, -5.33]
NN20	4.87 [4.45, 11.36]
NN50	2.38 [1.97, 4.45]
Median RR	2.30 [1.43, 7.03]
Mean HR	1.95 [0.89, 8.29]
Mean RR	1.65 [1.15, 4.57]
pNN50	-1.64 [-3.80, -1.07]
Maximum HR	1.45 [0.83, 3.04]
Minimum HR	-1.25 [-4.78, -0.42]
CVNN	-1.21 [-4.26, -0.63]

IV. DISCUSSION & CONCLUSION

Our results indicate that, in a controlled setting, machine learning classifier models can utilize data from consumer-grade wearable devices to differentiate physiological states associated with rest from those associated with task-induced stress in individuals with ASD. Considering only the 22 individuals with ASD, the LR and SVM models developed in this study achieved 84% and 91% accuracy, respectively. The few previous studies that have investigated approaches to utilize wearable device data as a potential intervention tool for individuals with ASD are limited to identification of the moment of crisis [13] or physical behavior following crises [12]. To our knowledge, this is the first work to demonstrate a method that can identify stress induced physiological state changes that may precede challenging behaviors in individuals with ASD.

Notably, as reflected by the performance metrics (Table II) and the SVM confusion matrix (Table III), model performance was essentially unchanged when applied to a nearly balanced cohort of ASD and non-ASD individuals. On that cohort, the LR and SVM models achieved 87% and 93%, accuracy, respectively. This result seems to imply that both ASD and non-ASD individuals experience similar physiological changes during stressful scenarios relative to their respective baselines, otherwise one would expect differences in model performance between the two groups. This suggests that it may be possible to identify stress in children through physiological changes independent of diagnoses.

Classifier model inputs included 20 statistical measures of temporal variations in HR and RR interval. One measure of the relative importance of individual features is the magnitude of coefficients obtained for the LR model. Coefficients with relatively larger coefficient magnitudes typically are more important in determination of model output. Our results indicate that 10 features had a mean coefficient magnitude greater than 1.0 (Table IV) and are considered important relative to the remaining 10 features with mean magnitude less than 1.0. Furthermore, the 6 to 4 ratio of positive to negative coefficient values suggests a complex interaction between the important features as it relates to the classification of overall physiological state. Our results appear to corroborate prior research that demonstrates association between changes in cardiac function and scenarios that involve stress [7,8,11]. We also observed that the coefficients of 7 features (none of the important features) changed signs during the nested LOO-CV iterations indicating that these features were associated with the rest class in some iterations and the stress class in others. Noting that the mean magnitude for all of these feature coefficients was less than 0.5, it may be that these features did not affect classification and thus a stable value could not be learned. Alternatively, the instability may suggest the presence of population subtypes where the variables interact differently relative to physiological response. Future research on a larger cohort is required to evaluate these possibilities.

As with many machine learning classifiers, the models presented here produce a numeric output, which can be interpreted as the probability that the current physiological state indicates the presence of stress. By default, the input is classified as stress if the estimated probability is greater than a threshold of 0.5. However, the threshold is an adjustable parameter. The ROC curves indicate that the threshold can be tuned to reduce false positives at the expense of increasing false negatives, or vice versa. In a real-world application, this tuning could be performed dynamically after model training and deployment to accommodate individual preference or varying social scenarios. For example, some individuals may find alerts distracting, similar to clinical alert fatigue [19], and therefore wish to reduce false positives by increasing the decision threshold. Similarly, in social settings such as school, where the perceived consequences of adverse outcomes are higher, a care-provider may be more willing to tolerate false alarms to avert escalation of challenging behaviors and therefore reduce the threshold to avoid false negatives.

The models presented here include several strengths. In particular, they achieve high performance with minimal computational requirements which would allow deployment on most consumer wearable devices. Additionally, they rely only on HR and RR interval data that are typically available from wearables. The structure of the models easily allows for retraining to accommodate additional features such as movement, oxygen saturation, and skin temperature that may be available from other sensors often integrated into current consumer wearables. Such information could further increase model performance. These findings support the use of consumer-grade wearable devices for tracking stress in clinical populations, including but not limited to children with ASD. These data may help to provide moment-to-moment care-giver decision support. We envision scenarios that involve a coordinated sequence of events across multiple m-health technologies: (1) wearable device detects an impending challenging behavior; (2) data sent to and processed by a cloud based application; (3) alert sent to care provider smartwatch; (4) provider acts to avert behavior escalation.

There are important limitations to this study. Most notably, these results are obtained for a balanced dataset (equal numbers of rest and stress samples). In a real-world setting, the actual rate of exposure to stress with the potential to lead to challenging behaviors may be much different than the rate of non-stressful exposure periods. The difference is likely to vary by individual and social setting. Consequently, a deployed model will have to accommodate variable and potentially large class imbalance. It is well known that class imbalance is a significant challenge for machine learning methods [20,21]. Overcoming this concern may require adjustment of the model training methods, e.g. to use class weighting. Additionally, more training samples may be required. This study included only 76 samples from 38 individuals. As indicated by the learning curves (Fig. 3), the sample size was sufficient to avoid significant model over-fitting and bias. This may not be the case if the real-world class balance is very far from one-to-one. Also, the introduction of more physiological features (e.g. temperature) would likely increase the necessary training sample size to avoid over-fitting.

An additional limitation relates to the controlled setting of this study wherein participants were seated during the task induced stress periods. While this replicates a typical classroom setting, it constrained our models to identifying changes that were minimally affected by motion artifacts and movement-induced physiological changes. A deployed application will have to contend with these factors which could potentially increase the false positive rate. Presumably, accelerometer input could address this limitation to some extent by informing the model about different characteristics of movement. This is a subject of our future research.

In conclusion, our results demonstrate that a machine learning approach can be used to recognize physiological states associated with stress that may lead to challenging behavior in individuals with ASD. These methods may therefore be valuable as part of a consumer application that alerts individuals with ASD or their caregivers of an impending challenging behavior. If utilized in settings with high incidence of stress, such an application could allow for intervention to avoid such behavior. As described above, challenges relative to class imbalance may be expected in translating such models from clinical to real-world settings. Therefore, further research that includes performance of a trial in realistic social settings is required to fully assess the utility of the approach described here.

Acknowledgment

We would like to thank the families who generously gave their time for this study. We would also like to thank the research assistants who helped with the testing, recruiting and behavioral coding of children who participated in this study, including Emma Finkel, Anushua Bhattacharya, Amanda Dennis, Ilona Jileaeva, Devin Murphy, Elizabeth Daniels and Zabryna Atkinson-Diaz.

References

[1] 1. Matson JL, Wilkins J, Macken J. The Relationship of Challenging Behaviors to Severity and Symptoms of Autism Spectrum Disorders. J Ment Health Res Intellect Disabil [Internet]. 2008 Dec 29 [cited 2019 Feb 3];2(1):29–44. Available from: http://www.tandfonline.com/doi/abs/10.1080/19315860802611415

[2] 2. McTiernan A, Leader G, Healy O, Mannion A. Analysis of risk factors and early predictors of challenging behavior for children with autism spectrum disorder. Res Autism Spectr Disord [Internet]. 2011 Jul 1 [cited 2019 Feb 3];5(3):1215–22. Available from: https://www.sciencedirect.com/science/article/pii/S1750946711000158

[3] 3. Lecavalier L, Leone S, Wiltz J. The impact of behaviour problems on caregiver stress in young people with autism spectrum disorders. J Intellect Disabil Res [Internet]. 2006 Mar 1 [cited 2019 Jan 22];50(3):172–83. Available from: http://doi.wiley.com/10.1111/j.1365-2788.2005.00732.x

[4] 4. Hastings RP, Brown T. Coping Strategies and the Impact of Challenging Behaviors on Special Educators' Burnout. Ment Retard [Internet]. 2002 Apr 14 [cited 2019 Jan 22];40(2):148–56. Available from: http://www.aaiddjournals.org/doi/abs/10.1352/0047-6765%282002%29040%3C0148%3ACSATIO%3E2.0.CO%3B2

[5] 5. Siegel M, Gabriels RL. Psychiatric hospital treatment of children with autism and serious behavioral disturbance. Child Adolesc Psychiatr Clin N Am [Internet]. 2014 Jan 1 [cited 2019 Jan 22];23(1):125–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24231172

[6] 6. Nuske HJ, Vivanti G, Dissanayake C. Are emotion impairments unique to, universal, or specific in autism spectrum disorder? A comprehensive review. Cogn Emot [Internet]. 2013 Sep [cited 2019 Jan 22];27(6):1042–61. Available from: http://www.tandfonline.com/doi/abs/10.1080/02699931.2012.762900

[7] 7. Barrera FJ, Violo RA, Graver EE. On the form and function of severe self-injurious behavior. Behav Interv [Internet]. 2007 Feb 1 [cited 2019 Jan 19];22(1):5–33. Available from: http://doi.wiley.com/10.1002/bin.228

[8] 8. Kuiper MWM, Verhoeven EWM, Geurts HM. Heart rate variability predicts inhibitory control in adults with autism spectrum disorders. Biol Psychol [Internet]. 2017 Sep 1 [cited 2019 Jan 22];128:141–52. Available from: https://www.sciencedirect.com/science/article/pii/S0301051117301254

[9] 9. Nuske HJ, Finkel E, Hedley D, Parma V, Tomczuk L, Pellecchia M, et al. Heart rate increase predicts challenging behavior episodes in preschoolers with autism. Stress. (in press).

[10] 10. Goodwin MS, Erdoğmuş D, Ioannidis S, Özdenizci O, Cumpanasoiu C, Tian P, et al. Predicting Imminent Aggression Onset in Minimally-Verbal Youth with Autism Spectrum Disorder Using Preceding Physiological Signals. In: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare - PervasiveHealth '18 [Internet]. New York, New York, USA: ACM Press; 2018 [cited 2019 Feb 3]. p. 201–7. Available from: http://dl.acm.org/citation.cfm?doid=3240925.3240980

[11] 11. Di Palma S, Tonacci A, Narzisi A, Domenici C, Pioggia G, Muratori F, et al. Monitoring of autonomic response to sociocognitive tasks during treatment in children with Autism Spectrum Disorders by wearable technologies: A feasibility study. Comput Biol Med [Internet]. 2017 Jun 1 [cited 2019 Jan 20];85:143–52. Available from: https://www.sciencedirect.com/science/article/pii/S001048251630083X

[12] 12. Amiri A, Peltier N, Goldberg C, Sun Y, Nathan A, Hiremath S, et al. WearSense: Detecting Autism Stereotypic Behaviors through Smartwatches. Healthcare [Internet]. 2017 Feb 28 [cited 2019 Jan 22];5(1):11. Available from: http://www.mdpi.com/2227-9032/5/1/11

[13] 13. Torrado JC, Gomez J, Montoro G, Torrado JC, Gomez J, Montoro G. Emotional Self-Regulation of Individuals with Autism Spectrum Disorders: Smartwatches for Monitoring and Interaction. Sensors [Internet]. 2017 Jun 11 [cited 2019 Jan 22];17(6):1359. Available from: http://www.mdpi.com/1424-8220/17/6/1359

[14] 14. Lord C, Risi S, Lambrecht L, Cook, Jr. EH, 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 [Internet]. 2000 [cited 2019 Feb 3];30(3):205–23. Available from: http://link.springer.com/10.1023/A:1005592401947

[15] 15. Vanderwal T, Kelly C, Eilbott J, Mayes LC, Castellanos FX. Inscapes: A movie paradigm to improve compliance in functional magnetic resonance imaging. Neuroimage [Internet]. 2015 Nov 15 [cited 2019 Feb 3];122:222–32. Available from: https://www.sciencedirect.com/science/article/pii/S1053811915006898

[16] 16. Goldsmith HH, Reilly HH, Lemery KL, Longley S, Prescott A. The laboratory temperament assessment battery: Middle childhood version. Unpubl manuscript, Univ Wisconsin-Madison. 2001;

[17] 17. Mani S, Ozdas A, Aliferis C, Varol HA, Chen Q, Carnevale R, et al. Medical decision support using machine learning for early detection of late-onset neonatal sepsis. J Am Med Informatics Assoc [Internet]. 2014 Mar 1 [cited 2014 Mar 24];21(2):326–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24043317

[18] 18. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in {P}ython. J Mach Learn Res. 2011;12:2825–30.

[19] 19. Kane-Gill SL, O'Connor MF, Rothschild JM, Selby NM, McLean B, Bonafide CP, et al. Technologic Distractions (Part 1). Crit Care Med [Internet]. 2017 Sep [cited 2019 Feb 3];45(9):1481–8. Available from: http://insights.ovid.com/crossref?an=00003246-201709000-00008

[20] 20. Haibo He, Garcia EA. Learning from Imbalanced Data. IEEE Trans Knowl Data Eng [Internet]. 2009 Sep;21(9):1263–84. Available from: http://ieeexplore.ieee.org/document/5128907/

[21] 21. Batista GEAPA, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explor Newsl [Internet]. 2004 Jun 1;6(1):20. Available from: http://portal.acm.org/citation.cfm?doid=1007730.1007735