

Biosensor Prediction of Aggression in Youth with Autism using Kernel-based Methods

Tales Imbiriba
talesim@ece.neu.edu
ECE, Northeastern University
Boston, Massachusetts

Stratis Ioannidis
ioannidis@ece.neu.edu
ECE, Northeastern University
Boston, Massachusetts

Diana Catalina Cumpanasoiu
cumpanasoiu.d@husky.neu.edu
HS, Northeastern University
Boston, Massachusetts

Deniz Erdoğan
erdogmus@ece.neu.edu
ECE, Northeastern University
Boston, Massachusetts

James Heathers
j.heathers@northeastern.edu
HS, Northeastern University
Boston, Massachusetts

Matthew S. Goodwin
m.goodwin@northeastern.edu
HS, Northeastern University
Boston, Massachusetts

ABSTRACT

Aggression to others by youth with autism is a significant problem since their difficulties self-reporting distress can lead to behaviors that appear to occur without warning. To address this issue, we recently demonstrated that biosensor data combined with linear classification algorithms (i.e., ridge-regularized logistic regression) can be used to predict aggression up to 1 minute before it occurs using 3 minutes of data from the past with an average area under the curve (AUC) of 0.71-0.84 depending on whether population versus individual models are used. In the present study, we both extend and enhance these prior results through the use of principal component analysis and a nonlinear kernel-based classifier (Support Vector Machines). Our results illustrate that these newly applied methods yield significant improvements, predicting aggression up to 3 minutes before it occurs with an average AUC of 0.98 in both population *and* individual models. Furthermore, we extend our prior work by evaluating aggression prediction performance across varying observed aggression intensities and find that moderate and high intensity aggression episodes are detectable with 2 to 5% higher average AUC than low-intensity aggression episodes.

CCS CONCEPTS

• Applied computing → Health informatics; • Computing methodologies → Supervised learning by classification.

KEYWORDS

Aggression, ASD, SVM, Kernel-Methods, Classification.

ACM Reference Format:

Tales Imbiriba, Diana Catalina Cumpanasoiu, James Heathers, Stratis Ioannidis, Deniz Erdoğan, and Matthew S. Goodwin. 2020. Biosensor Prediction of Aggression in Youth with Autism using Kernel-based Methods. In *The 13th Pervasive Technologies Related to Assistive Environments Conference (PETRA '20)*, June 30-July 3, 2020, Corfu, Greece. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3389189.3389199>

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 ACM 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.

PETRA '20, June 30-July 3, 2020, Corfu, Greece

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-7773-7/20/06...\$15.00

<https://doi.org/10.1145/3389189.3389199>

1 INTRODUCTION

Autism spectrum disorder (ASD) is an increasingly prevalent (1 in 59 children by age eight) childhood disorder [4] and is associated with high health care costs [2]. One of the most frequently observed problem behaviors in youth with ASD is aggression [14], which is a primary cause for referral to behavioral healthcare services [3]. Physical aggression – including scratching, biting, hitting, and throwing objects at others – is particularly debilitating in youth with ASD because it can occur without warning, sometimes long after any observable trigger. Unpredictable aggression creates a barrier to accessing community, therapeutic, medical, and educational services since caregivers are understandably afraid to put the individual with ASD into potentially stressful environments that might lead to aggression without warning.

Recent work performed by our team demonstrates that linear classification techniques (i.e., ridge-regularized logistic regression) can be used to predict aggression to others in ASD youth up to 1 minute before it occurs using 3 minutes of prior biosensor data with an average area under the curve of 0.71 for a population model and 0.84 for individual models [7–9, 16]. While promising, this prior work only considered linear classification techniques that do not account for nonlinear relations between extracted features and aggressive behavior onset. In the present study, we explore *Support Vector Machines* (SVMs) as a way to capture such nonlinear relations. Additionally, we consider the contribution of Principal Component Analysis (PCA) to reduce data redundancy and evaluate whether it improves SVM performance. Finally, we extend our prior work by evaluating aggression prediction performance across varying observed aggression intensities when LR and SVM approaches are used.

This paper is organized as follows. Section 2 presents the data set and our preprocessing methodology. Section 3 presents general SVM theory and our specific aggression prediction classifier. Experimental results are discussed in Section 4 and final remarks, including ongoing work, are presented in Section 5.

2 DATA SET AND PREPROCESSING

2.1 Participants

Twenty ADOS-2 [15] confirmed, behaviorally unstable, ASD youth in the Developmental Disorders Unit at Spring Harbor Hospital in Portland, Maine contributed data in this study. The sample on

average was 10.8 years old (SD = 3.10, range = 6–17), 75% male ($n = 15$), 95% Caucasian ($n = 19$), and 90% non-Hispanic ($n = 18$). Mean nonverbal IQ (NVIQ) of the sample assessed at hospital admission was 66.1 (SD = 18.99, range = 31 – 98), with 65% of participants ($n = 13$) meeting criteria for intellectual disability (NVIQ ≤ 70). Severity of behavior in the sample was also assessed at intake using the caregiver-rated Aberrant Behavior Checklist (ABC [1]) Irritability subscale, and resulted in a mean of 28.11 (SD = 8.55, range = 16 – 42).

2.2 Data Acquisition

Data recorded in this Institutional Review Board approved study included naturalistic observational sessions in a specialized ASD psychiatric inpatient unit. Participants wore an E4 biosensor (Empatica Inc., United States) on their wrist that records peripheral physiological arousal (blood volume pulse [BVP] and inter-beat-interval [IBI] at 64 Hz using photoplethysmography and electrodermal activity [EDA] at 4 Hz) and motion activity (8 g at 32 Hz using a 3-axis accelerometer [ACC_x , ACC_y , and ACC_z]) while inpatient research staff concurrently coded time-stamped naturalistic observations of operationally defined aggression to others (e.g., hitting, kicking, biting, scratching, grabbing, pulling). Inter-rater reliability for observed aggression onset and offset between two research staff from the inpatient site yielded 0.90 percent agreement and a corresponding Cohen's Kappa of 0.79, wherein a maximum tolerance of 2sec onset or offset difference between raters was considered an agreement (see aggression duration mean and standard deviation in Table 1, i.e., average duration of observed aggression in the sample was 28sec, much longer than a 2sec difference between raters) in 20% of data randomly selected from our corpus. Research staff conducted these data collection observations with minimal interference to participants' regularly scheduled daily routines over the course of their inpatient stay. Since the data collection sessions were observational and naturalistic, researchers did not interact with participants or clinical staff before, during, or after aggression episodes, they only provided clinical staff with assistance donning and doffing the E4 (see [8] for more detail about our data collection procedures).

Sixty-nine independent naturalistic observational sessions were collected (Table 1), including $m(sd) = 3.45(2.84)$ observational sessions per participant and totaling 87hrs ($m(sd) = 4.35\text{hrs}(4.8\text{hrs})$ per participant). Out of 548 total aggressions observed with concurrent E4 data, mean and standard deviation of aggression frequency and duration was 27(34) episodes in a four-hour period of 28sec(32sec) average length, respectively.

2.3 Data Preprocessing

Naturalistic observations yielded labeled time-series data via six biosignal sources (i.e., BVP , IBI , EDA , ACC_x , ACC_y , ACC_z). Statistical analyses on these physiological and motion activity signals were performed through extracted time-series features offline. In bins of $T_0 = 15$ seconds, the following features were calculated: first, last, maximum, minimum, mean and median value, amount of unique values, and the sum, standard deviation, and variance of values falling in a bin. Thus, for every 15-second bin a time-dependent feature vector $\mathbf{z}_t \in \mathbb{R}^{d_z}$, with $d_z = 60$, $t = kT_0$, $k \in \mathbb{Z}^+$, and a

label $y_{z,t} \in \{-1, +1\}$ indicating if an aggression episode occurred ($y_{z,t} = 1$) or not ($y_{z,t} = -1$) was created.

The objective of our work is to predict future aggression episodes by processing biosensor data from the past. To this end, given a time index $k > 0$, we construct the dataset by concatenating feature vectors from the past and compose their respective labels by observing if an aggression episode happens in the future. Thus, given past and future window lengths $\tau_p = n_{b_p}T_0$ and $\tau_f = n_{b_f}T_0$ with n_{b_p} and n_{b_f} being the number of bins used in the past and future, respectively, the final features are created as $\mathbf{x}_k = [z_k^\top, \dots, z_{k-n_{b_p}}^\top, \sigma_{z_k}^\top]^\top$, where $z_k = z_{kT_0}$, and $[\sigma_{z_k}]_i$, $i = 1, \dots, d_z$, is the standard-deviation of the vector $[z_{k,i}, \dots, z_{k-n_{b_p},i}]$. The dataset is then composed by pairs $(\mathbf{x}_\ell, y_\ell)$, $\ell = 1, \dots, N$, $\mathbf{x}_\ell \in \mathbb{R}^d$, $d = (n_{b_p} + 1)d_z$, $y \in \{-1, +1\}$, and $\ell = k - n_{b_p}$.

Next, to quantify the intensity of observed aggression episodes, we used the accelerometer (ACC_x , ACC_y , ACC_z) time-series to build a new feature time-series $NACC_t = (ACC_{x,t}^2 + ACC_{y,t}^2 + ACC_{z,t}^2)^{1/2}$ that quantifies the square root of the energy of XYZ movements over time. This time-series was also averaged for every 15-second bin resulting in a unidimensional temporal feature $y_k^{(NACC)}$. Cluster analysis on this new feature ($y_k^{(NACC)}$) was used to provide more detailed labels for observed aggression episodes, wherein three clusters sorting into *low*, *moderate*, and *high* intensities were derived for each participant.

3 SVM-BASED AGGRESSION PREDICTOR

In this section we present our SVM-based predictor. Kernel methods, including SVMs, have been used in a variety of domains including classification, regression, detection of nonlinearity [12, 13], dictionary learning [10, 11], among others [18, 20]. Nonlinear SVMs implicitly map input features into a *Reproducing Kernel Hilbert Space* (RKHS), where maximal-margin linear classifiers are then employed. Kernel functions play a central role in this process as they allow both mapping of data into the RKHS and computation of inner products of two mapped features by evaluating a real function in the original feature space, thus allowing one to implicitly work in a possible infinite dimensional Hilbert space.

3.1 SVM for Classification

Given a training set $\{\mathbf{x}_\ell, y_\ell\}_{\ell=1}^N$, nonlinear SVMs aim to find the maximal-margin separating hyperplane by casting the convex problem in the RKHS [5, 6, 20]. By virtue of the kernel reproducing property, the solution to this problem is a linear combination of kernel functions of the training data:

$$\begin{aligned} f(\mathbf{x}) &= \text{sgn} \left(\sum_{\ell \in \mathcal{S}} y_\ell \alpha_\ell \kappa(\mathbf{x}_\ell, \mathbf{x}) + b \right) \\ &= \text{sgn}(h(\mathbf{x}) + b) \end{aligned} \quad (1)$$

where $\text{sgn}(\cdot)$ is the sign function, $h(\mathbf{x}) = \sum_{\ell \in \mathcal{S}} y_\ell \alpha_\ell \kappa(\mathbf{x}_\ell, \mathbf{x})$, $\kappa: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is the kernel function, α_ℓ , $\ell = 1, \dots, N$ are the Lagrange multipliers introduced in the dual formulation of SVMs [6, 20], and \mathcal{S} is the index set of support vectors.

Probabilistic estimates can be obtained by using a sigmoid function $p(h(\mathbf{x})|y = 1) = [1 + \exp(Ah(\mathbf{x}) + B)]^{-1}$ whose parameters (A , B)

Table 1: Naturalistic data collection descriptive statistics of participants.

Participant	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	Group	Mean	SD
Number of Sessions	5	3	3	2	2	2	8	9	2	1	1	1	2	6	1	1	10	1	5	4	69	3.45	2.84
Total Obs. Duration*	9.33	3.97	3.77	3.02	1.25	0.57	7.02	8.47	2.72	1.43	0.22	1.38	1.6	8.47	0.52	1.02	20.48	0.78	5.1	5.87	86.99	4.35	4.80
Number of Aggression Episodes	72	7	13	8	9	6	35	30	50	1	2	3	9	39	1	8	130	2	76	47	548	27.40	33.84
Mean Agg. Duration†	9	102	77	11	19	9	19	18	50	3	1	15	19	6	51	7	103	4	7	22	–	27.6	31.93

* Total observation durations are presented in hours. † Mean aggression durations are presented in seconds.

can be obtained via maximum likelihood estimation [17]. Finally, defining the hypotheses \mathcal{H}_0 and \mathcal{H}_1 as being “non-aggression” and “aggression,” respectively, the SVM detector reduces to:

$$p(y = 1|h(\mathbf{x})) = \frac{p(h(\mathbf{x})|y = 1)p(y = 1)}{\sum_{j=\pm 1} p(h(\mathbf{x})|y = j)p(y = j)} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\leq}} \xi \quad (2)$$

where $\xi \in \mathbb{R}_+$ is the detection threshold.

The literature suggests a variety of kernel functions with different properties [19, 20]. In the present work, we consider the Gaussian kernel due to its smoothness and universal property [19]. The Gaussian kernel is given by:

$$\kappa(\mathbf{x}, \mathbf{x}') = \exp\left(-\|\mathbf{x} - \mathbf{x}'\|^2 / (2\sigma^2)\right) \quad (3)$$

where the parameter $\sigma^2 > 0$ controls the kernel bandwidth.

3.2 Individual and Population Models

In order to make direct comparisons between our previously published results and current approach, we followed the exact methodology presented in [7–9, 16] wherein we evaluate prediction accuracy at both individual and population levels at varying past and future time periods. At the individual level we learn a model specific to each participant. At the population level we learn a model from all participants. By comparing both approaches we can evaluate relative trade-offs between a more flexible strategy, which requires training a model for each participant, versus a less flexible one, in which all participants use the same model.

4 EXPERIMENTS

In this section we present results obtained with the dataset described in Section 2 and compare performance of the previous logistic regression (LR) strategy in [9] with the SVM-based classifier discussed in Section 3.

During all experiments, features extracted as described in Section 2 were normalized to have zero mean and a standard deviation of one. We also performed stratified cross-validation (CV) with four folds and six repetitions. In individual models, normalization constants (mean and standard deviation) were estimated using data from other participants. In the population model, these constants were estimated using the training set in each CV split. We implemented the same strategy in analyses involving PCA to ensure

that test data were never used during the training phase, and that enough data was available to perform the estimation process. More specifically, in the population model PCA parameters were computed as eigenvectors associated with the largest eigenvalues from the estimated covariance matrix, using only the training set at each CV iteration. In these models, test data is projected in the space spanning these eigenvectors. For individual models, we create a separate PCA model for each participant. To estimate PCA parameters we used data from all participants except for the model “owner.” This approach was adopted to ensure that: (1) enough data is available to estimate the data covariance matrix and, as a consequence, generate meaningful PCA parameters and (2) both population and individual test data within each CV iteration are never used in the pre-processing stage. Fig. 1 presents percentage explained by the largest eigenvalue as the number of eigenvalues increase. A clear knee in the curve is visible before 10 eigenvalues, thus we performed PCA with 10 principal components.

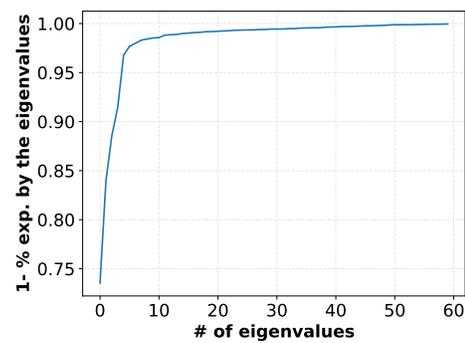


Figure 1: PCA: 1 - percentage explained by the eigenvalues.

In our preliminary tests we found that PCA greatly improved classification performance of SVMs while degrading performance of the LR classifier. Thus, we only use PCA in conjunction with SVMs in our subsequently reported results.

4.1 Performance Metrics

We used *receiver operating characteristic* (ROC) curves and the *area under the curve* (AUC) as our primary performance metrics in all experiments, accompanied by confidence intervals (± 1 standard deviation). ROC curves plot the probability of false alarms (P_{FA}) versus the probability of accurate predictions (P_D) at varying prediction thresholds ξ in (2). An ideal classifier presents $P_D = 1$ for a $P_{FA} = 0$, and thus, an AUC = 1.

4.2 Results

We present results from two different experiments in this section. In the first experiment, we compare our previous LR-based classifier to our new SVM-based classifier at varying past observation τ_p and prediction τ_f window lengths using only binary “aggression” and “non-aggression” labels. In the second experiment, we use the more detailed intensity of aggression labels, obtained by cluster analyses on $\gamma_k^{(NACC)}$ features, to assess classifier performance as τ_p is reduced.

Results of the first experiment are summarized in Table 2 for $\tau_p, \tau_f \in [180, 120, 60]$ seconds. Fig. 2 present ROCs for the LR-based (top panel) and SVM-based (bottom panel) classifiers for both individual (colored dashed lines) and population (blue solid line) models for $[\tau_p, \tau_f] = [180, 60]$.

Table 2: Average AUC for individual and population models.

Logistic Regression			
Average AUC for individual models			
τ_p (sec)	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$
180	0.84 ± 0.0369	0.84 ± 0.0327	0.86 ± 0.0293
120	0.83 ± 0.0444	0.85 ± 0.0314	0.85 ± 0.0310
60	0.80 ± 0.0550	0.82 ± 0.0447	0.82 ± 0.0453
Average AUC for population model			
180	0.67 ± 0.0118	0.64 ± 0.0066	0.64 ± 0.0068
120	0.67 ± 0.0092	0.64 ± 0.0075	0.63 ± 0.0079
60	0.66 ± 0.0099	0.63 ± 0.0061	0.62 ± 0.0106
Support Vector Machine + PCA			
Average AUC for individual models			
180	0.97 ± 0.0044	0.98 ± 0.0161	0.98 ± 0.0122
120	0.94 ± 0.0256	0.94 ± 0.0256	0.94 ± 0.0231
60	0.76 ± 0.0598	0.77 ± 0.0487	0.78 ± 0.0557
Average AUC for population model			
180	0.97 ± 0.0044	0.98 ± 0.0026	0.98 ± 0.0015
120	0.94 ± 0.0038	0.94 ± 0.0038	0.95 ± 0.0037
60	0.81 ± 0.0087	0.82 ± 0.0089	0.81 ± 0.0077

Results in Table 2 are aggression prediction accuracy levels observed at different window lengths for both classifiers (LR and SVM+PCA). At both individual and population levels, and across both models, overall prediction performance degrades as τ_p decreases, yet remains constant as τ_f increases. When comparing individual- versus population-trained models within each classifier,

it is evident that individual models have a greater impact for LR classifiers than SVMs models. Specifically, LR produces approximately 20% variability in accuracy between individual and population models. These results contrast with SVM classifier performance wherein almost constant accuracy is observed across both model levels, except for $\tau_p = 60$, where the population model provides an accuracy gain of approximately 6%. These results suggest that the flexibility of the SVM-based classifier at the population level minimizes model bias and yields lower variance relative to those obtained with individual models (see the ± 1 standard deviation intervals provided in Table 2). Comparing results between classifiers, greater performance gains in the SVM-based method is evident at $\tau_p \geq 120$, fluctuating between 9% and 13% for individual models and between 28% and 34% for the population model. At $\tau_p = 60$, SVMs continue to produce superior results to LR at the population level (ranging from 18% to 23%), but prediction accuracy decreases by roughly 6% at the individual level. These findings suggest that SVM benefits more from observing data further in the past than LR, and that it is possible to predict aggression with very high accuracy (up to 98%) three minutes into the future using SVM. We highlight $\tau_p = 180$ and $\tau_f = 60$ since these window lengths enabled the highest prediction accuracy in our prior LR work [7–9, 16].

Finally, Fig. 2 presents LR (top panel) and SVM (bottom panel) ROCs. In each panel, ROCs for individual (dashed lines) and population (solid blue line) models are displayed for $\tau_p = 180$ and $\tau_f = 60$. Again, we observe better performance with the SVM-based predictor relative to LR. For instance, if we design our prediction system with a P_{FA} of 0.1 the P_D at the population level is roughly 0.9 and 0.3 for SVM and LR models, respectively. Although LR results improve at the individual level resulting in a $P_D \approx 0.7$, its variance across different individuals is much larger than in SVM models.

The second experiment is summarized in Table 3 where AUCs are presented for *Low*, *Moderate*, and *High* aggression intensities, performed only at the individual level. Specifically, the ROC curves were constructed by computing prediction probabilities (P_D 's) for each individual aggression intensity class, resulting in the AUCs presented in Table 3. One drawback to this strategy is that we had to discard participants with very few observed aggressions in each class in order to perform stratified CV. Overall, 7 of our 20 participants did not present with more than 4 instances of any aggression intensity label and thus were removed. Results for the remaining 13 participants are summarized in Table 3, where the average AUC at the individual level is presented for $\tau_f = 120$.¹

Table 3: Average AUC for individual models ($\tau_f = 120$).

τ_p (sec)	Low	Moderate	High
180	0.96 ± 0.0292	0.98 ± 0.0183	0.97 ± 0.0224
120	0.92 ± 0.0447	0.94 ± 0.0390	0.94 ± 0.0512
60	0.71 ± 0.1252	0.74 ± 0.1138	0.75 ± 0.0914

Reviewing Table 3, we notice that conclusions made in the previous experiment regarding degradation of prediction performance

¹We present results only for $\tau_f = 120$ since $\tau_f = 180$ resulted in even fewer participants available for analysis and results with $\tau_f = 60$ led to equivalent results.

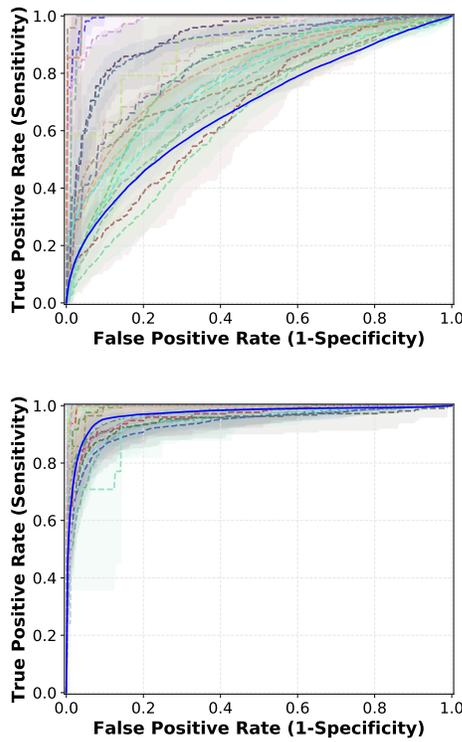


Figure 2: ROC curves obtained with the LR (top) and SVM (bottom) classifiers for the population (blue solid line) and individual (dashed lines) models for $\tau_p = 3$ and $\tau_f = 1$ minutes.

as τ_p decreases holds. Furthermore, we note that prediction performance decreases are noticeable across all aggression intensities. However, prediction accuracy for high and moderate aggression intensities is roughly 2 to 5% higher than results for low intensity.

5 CONCLUSIONS

This paper employed a kernel-based nonlinear classification methodology and principal component analysis (SVM+PCA) to predict onset of aggression using biosensor data in 20 psychiatric inpatient youth with more severe forms of ASD. Our SVM results demonstrate significant improvement in prediction performance (0.98 average AUC) compared to the LR-based strategy (0.84 average AUC) we published previously. Other key findings include extended performance predicting aggression onset further into the future (i.e., three minutes opposed to one minute) and superior accuracy for moderate- and high-intensity aggression episodes with shorter past observational windows (i.e., one minute opposed to three minutes).

As a natural extension of the present investigation, we are currently working on assessing generalizability of the results we obtained using kernel-based methods in a larger independent sample of ASD youth. We are also exploring the development of a non-homogeneous Poisson process model capable of providing a probabilistic measure of aggression onset at longer future time intervals.

Finally, we are evaluating how well SVM can discriminate aggression to others vs. self-injurious behavior vs. perceived emotion dysregulation without aggression (i.e., multi-class prediction).

The advantages of biologically based tools to identify processes that underlie behavioral dysregulation in youth with ASD as it unfolds during moments of escalation are numerous, including the potential to enable real-time predictive alerts that facilitate preemptive intervention strategies to help prevent or mitigate significant behavior problems. Such capabilities would be especially useful for those unable to provide reliable self-reports on their arousal states that lead to aggression.

ACKNOWLEDGMENTS

This work was supported by grants from the Department of Defense (12507232, 12507733), National Institute of Child Health and Human Development (R01 HD079512), National Institute on Deafness and Other Communication Disorders (P50 DC013027), National Science Foundation (SCH-1622536, IIS-1118061), Nancy Lurie Marks Family Foundation, and Simons Foundation (SFARI 296318, 61803).

REFERENCES

- [1] M. G. Aman, N. N. Singh, A. W. Stewart, and C. J. Field. 1985. The aberrant behavior checklist: a behavior rating scale for the assessment of treatment effects. *American Journal of Mental Deficiency* 89, 5 (1985), 485–91.
- [2] D. Amendah, S. D Grosse, G. Peacock, and D. S. Mandell. 2011. The economic costs of autism: A review. *Autism spectrum disorders* (2011), 1347–1360.
- [3] L. E. Arnold, B. Vitiello, C. McDougle, L. Scahill, B. Shah, N. M. Gonzalez, S. Chuang, M. Davies, J. Hollway, M. G. Aman, P. Cronin, K. Koenig, A. E Kohn, D. J. McMahon, and E. Tierney. 2003. Parent-defined target symptoms respond to risperidone in RUPP autism study: customer approach to clinical trials. *Journal of the American Academy of Child & Adolescent Psychiatry* 42, 12 (2003), 1443–1450.
- [4] J. Baio, L. Wiggins, D. L. Christensen, M. J. Maenner, J. Daniels, Z. Warren, M. Kurzius-Spencer, W. Zahorodny, C. R. Rosenberg, T. White, et al. 2018. Prevalence of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2014. *MMWR Surveillance Summaries* 67, 6 (2018), 1.
- [5] C.-C. Chang and C.-J. Lin. 2011. LIBSVM: A library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)* 2, 3 (2011), 27.
- [6] C. Cortes and V. Vapnik. 1995. Support-vector networks. *Machine learning* 20, 3 (1995), 273–297.
- [7] D. Cumanasoiu, C. Mazefsky, A. Stedman, C. Peura, P. Tian, Y. Guo, S. Ioannidis, D. Erdoğmuş, M. Siegel, and M. S. Goodwin. 2017. 5.12 physiological biomarkers for prediction of imminent aggression in minimally verbal children with autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry* 56, 10 (2017), S256–S257.
- [8] M. S. Goodwin, C. A. Mazefsky, S. Ioannidis, D. Erdoğmuş, and M. Siegel. 2019. Predicting aggression to others in youth with autism using a wearable biosensor. *Autism Research* (2019).
- [9] M. S. Goodwin, O. Özdenizci, C. Cumanasoiu, P. Tian, Y. Guo, A. Stedman, C. Peura, C. Mazefsky, M. Siegel, D. Erdoğmuş, et al. 2018. 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*. ACM, 201–207.
- [10] T. Imbiriba, J. C. M. Bermudez, and C. Richard. 2017. Band Selection for Nonlinear Unmixing of Hyperspectral Images as a Maximal Clique Problem. *IEEE Transactions on Image Processing* 26, 5 (May 2017), 2179–2191. <https://doi.org/10.1109/TIP.2017.2676344>
- [11] T. Imbiriba, J. C. M. Bermudez, C. Richard, and J.-Y. Tourneret. 2015. Band selection in RKHS for fast nonlinear unmixing of hyperspectral images. In *2015 23rd European Signal Processing Conference (EUSIPCO)*. 1651–1655. <https://doi.org/10.1109/EUSIPCO.2015.7362664>
- [12] T. Imbiriba, J. C. M. Bermudez, C. Richard, and J.-Y. Tourneret. 2016. Non-parametric detection of nonlinearly mixed pixels and endmember estimation in hyperspectral images. *IEEE Transactions on Image Processing* 25, 3 (March 2016), 1136–1151. <https://doi.org/10.1109/TIP.2015.2509258>
- [13] T. Imbiriba, J. C. M. Bermudez, J.-Y. Tourneret, and C. Richard. 2014. Detection of nonlinear mixtures using Gaussian processes: Application to hyperspectral imaging. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*. 7949–7953.

- [14] S. M. Kanne and M. O. Mazurek. 2011. Aggression in children and adolescents with ASD: Prevalence and risk factors. *Journal of Autism and Developmental Disorders* 41, 7 (2011), 926–937.
- [15] C. Lord, M. Rutter, P. DiLavore, S. Risi, K. Gotham, and S. Bishop. 2012. Autism diagnostic observation schedule - second edition (ADOS-2). *Los Angeles, CA: Western Psychological Corporation* (2012).
- [16] O. Özdenizci, C. Cumpănăsioiu, C. Mazefsky, M. Siegel, D. Erdoğan, S. Ioannidis, and M. S. Goodwin. 2018. Time-series prediction of proximal aggression onset in minimally-verbal youth with autism spectrum disorder using physiological biosignals. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 5745–5748.
- [17] J. Platt. 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers* 10, 3 (1999), 61–74.
- [18] C. E. Rasmussen and C. K. I. Williams. 2006. *Gaussian Processes for Machine Learning*. The MIT Press.
- [19] I. Steinwart and A. Christmann. 2008. *Support vector machines*. Springer.
- [20] J. A. K. Suykens, T. V. Gestel, J. D. Brabanter, B. D. Moor, and J. Vandewalle. 2002. *Least squares support vector machines*. World scientific.